Alma Mater Studiorum – Università di Bologna

DOTTORATO DI RICERCA IN

GEOFISICA

Ciclo 33

Settore Concorsuale: 02/C1 - ASTRONOMIA, ASTROFISICA, FISICA DELLA TERRA E DEI PIANETI

Settore Scientifico Disciplinare: FIS/06 - FISICA PER IL SISTEMA TERRA E IL MEZZO CIRCUMTERRESTRE

ANALYSIS OF FAR INFRARED (FIR) HIGH SPECTRAL RESOLUTION DATA FOR CLOUD STUDIES

Presentata da: William Cossich Marcial de Farias

Coordinatore Dottorato

Nadia Pinardi

Supervisore

Tiziano Maestri

Esame finale anno 2021

Abstract

The Far-Infrared Outgoing Radiation Understanding and Monitoring (FORUM) mission has recently been selected by the European Satellite Agency (ESA) as 9th Earth Explorer mission. FORUM mission aims at studying the water vapor and clouds by filling the long-standing gap in Far-Infrared (FIR) spectral observations from space. In the framework of the FORUM mission, this thesis analyses FIR measurements to characterize the spectral signatures of radiance in presence of ice clouds. At this purpose, a cloud identification and classification code (named CIC) is implemented. CIC is an innovative machine learning algorithm, based on principal component analysis, able to perform cloud detection and scene multi-class classification. CIC is easily adaptable to different datasets and type of spectral sensors. It is firstly tested against a synthetic dataset comprising simulated measurements of the FORUM mission. Subsequently, CIC is applied to airborne interferometric data and finally it is used for the analysis of measured downwelling radiances collected in very dry conditions on the Antarctic Plateau. Provided the excellent performances of the algorithm, especially in the identification of thin cirrus clouds, CIC is adopted as the classificator in the official ESA FORUM End-to-End simulator (FE2ES). The FE2ES is a complex chain of codes used to simulate the entire FORUM mission from satellite orbit and geometry to level 2 product analysis. An extensive use of CIC is performed on ground-based radiances collected in Antarctica. The dataset is exploited to test and to optimize the CIC algorithm and for the developing of punctual statistic of cloud occurrence in the Antarctic Plateau. Meteorological conditions from this region are also analysed and correlated with the presence of clouds. The studies presented in this work showed the potentiality and the importance of the exploitation of spectral radiance measurements in the FIR for cloud identification and classification.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

In September 2019 the Far-infrared Outgoing Radiation Understanding and Monitoring (FORUM) mission[1] was selected as ESA's ninth Earth Explorer. Earth Explorer missions are devoted to innovative measurement techniques to explore and understand different aspects of the Earth system, addressing questions that have a direct bearing on scientific and societal issues, such as the availability of food, water, energy and resources, public health and climate change (ESA, 2019). FORUM has the objective to evaluate the role of the FIR in shaping the current climate and thus reduce uncertainty in predictions of future climate change by:

- building a highly accurate global dataset of far infrared radiances for validation of the present-day state as captured by Earth system models;

- using these measurements to understand and constrain the processes that control far infrared radiative transfer and hence the Earth's Greenhouse Effect;

- updating the parametrisations of these processes for implementation in radiative transfer codes, and ultimately in Earth system models;

- characterising critical feedback mechanisms.

The main instrumental component of the FORUM mission is an infrared spectrometer devoted to high spectral resolution measurements from 100 to 1600 cm^{-1}, thus including the so called FIR region, spanning 100 to 667 cm^{-1}, which has never been observed from space, spectrally resolved, and in its entirety.

The FIR represents an important fraction of the Earth's Outgoing Long Wave Radiation (OLR), which makes a considerable contribution to the planetary energy balance (Harries et al., 2008). Assuming that the Earth emits radiation as a black-body, viewed from space the planet has a global mean effective emitting temperature of around 255 K

---
[1]https://www.forum-ee9.eu/

(Liou, 2002), which results in a peak of energy occurring at about 500 cm^{-1}, within the FIR part of the spectrum. Moreover, up to 35% of the Earth's energy which escapes to space is at wavenumbers below 500 cm^{-1} (Sinha and Harries, 1995), and climate model simulations suggest that 40–65 % of Earth's OLR is found in the FIR (ESA, 2019).

In this part of the electromagnetic spectrum, the atmospheric emission is driven by the rotational absorption band of the water vapour molecules. This band is characterised by strong absorption lines interspersed by narrow regions (called *dirty micro-windows*), where absorption is less intense. The strong absorption features of water vapour roto-vibrational lines cause atmospheric weighting functions in the FIR to peak in the middle/upper troposphere. It makes the on-line upward emission particularly sensitive to the atmospheric thermodynamic profile and water vapour content of the highest tropospheric levels. Although the spectral properties of the cloud-free atmosphere at FIR are dominated by water vapour, next to 15 μm the CO_2 has also a significant contribution to the atmospheric absorption.

Harries et al. (2008) indicated that the atmospheric absorption in the FIR is so strong that from space the surface becomes obscured, except in the coldest and driest conditions near the poles. These authors made a summary of FIR properties in the Earth's atmosphere, which is described as follows. The pure rotational band of water vapour extends from the microwave region, starting next to 6 cm^{-1} , to the CO_2 ν_2 vibration-rotational band centered at 667 cm^{-1}. The center of this pure rotational band is around 100–150 cm^{-1}, comprising many hundreds of strongly absorbing individual pure rotational lines. The intensities of transitions fall to lower values in the region between 400 and 600 cm^{-1}, where a partially transparent window occurs and begins the lower-frequency edge of the very strong CO_2 band. Other molecular species, like O_3, N_2O, HNO_3, and NH_3, also have spectral features in the FIR, which can be used for retrieval of atmospheric composition. However, in terms of radiative energy transfer, their contribution is insignificant compared with water vapour (Harries et al., 2008). Moreover, the interactions between the water vapour with itself and other atmospheric molecules produce an additional radiance absorption called *water vapour continuum*, defined as all absorption not attributable to a Lorentz (collisional broadening) line, within 25 cm^{-1} of each line center (Clough et al., 1989). Merrelli (2012) points out that the strength of the continuum absorption due to water vapour is highest in the FIR. Harries et al. (2008) indicate that this continuum absorption reduces the outgoing thermal radiation at the Top of Atmosphere (TOA) in clear-sky conditions by about 2.2%, while downwelling longwave radiation at the surface increases by about 5.8%. The *water vapour continuum absorption* is usually modelled through observations (Mlawer et al., 2012; Serio et al., 2008).

Observed from the space, the FIR becomes particularly useful for remote sensing of the atmospheric composition in the upper troposphere and in the stratosphere, where the water vapour absorption is weaker and so does not block absorption lines due to other gases (Carli, 1984). This implies that much of the Earth's OLR comes from the upper troposphere/stratosphere, because the lower troposphere is totally opaque over

most of the planet in the FIR, with no energy from surface escaping to space in this part of the spectrum. However, in very dry and clear conditions, such as those often seen in the Arctic and Antarctic interiors, some FIR channels between 350 and 600 cm^{-1} become more transparent, potentially allowing the retrieval of the surface properties of these regions.

Therefore, the FIR allow us to see the important contribution of the water vapour for the greenhouse effect. This gas can absorb energy from the warm surface below and reemit it to space at a lower temperature in the upper troposphere, evidencing the strong feedback, in which any greenhouse warming due to increasing levels of CO_2 is amplified if more water vapour enters the atmosphere as a result (Harries et al., 2008).

The condensed phases of water, in form of liquid water and ice clouds, also affect the Earth's Radiation Budget (ERB) significantly (Sinha and Harries, 1995). In particular, the presence of ice clouds causes lower emitting temperatures and, hence, a shift towards longer wavelengths (thus towards the FIR) of the black-body emission distribution function peak. For this reason, a detailed study of the ERB should account for global, accurate, all-sky conditions measurements of the exiting radiance, including the FIR.

Many recent studies have shown, due to the large sensitivity of the upward radiance in the FIR part of the spectrum to water vapour and clouds, that the FIR can be used to complement standard remote sensing measurements performed in the Mid InfraRed (MIR) part of the spectrum. Both FIR and MIR are used in order to improve the retrieval of atmospheric water vapour profile, and for cloud detection, classification, and properties derivation (Palchetti et al., 2016; Di Natale et al., 2017; Maestri et al., 2019a).

In this context, some studies are presented in the next section regarding the FIR contribution for a better understanding of the climate feedback and dynamics, and cloud properties, following by the objectives and structure of this thesis.

1.1 FIR and climate feedbacks

As previously indicated, the FIR part of the spectrum is highly sensitive to the distribution of the atmospheric's water vapour, which has a dominant effect on the ERB. The vertical distribution of this gas determines that between the middle and upper troposphere occurs the maximum cooling of the planet. Then, the FIR has a critical role not only in regulating Earth's clear-sky cooling to space but also in influencing atmospheric dynamics (Harries et al., 2008), and, consequently, the climate (Dessler et al., 2008).

Studies, like that developed by Merrelli and Turner (2012) and Sinha and Harries (1995), have shown that the good sensitivity in the FIR to identify small changes in water vapour concentration can be exploited to improve our knowledge on water vapour concentrations in the Upper Troposphere/Lower Stratosphere (UTLS). Additionally, in a sensitivity study, ESA (2019) showed that a perturbation on the water vapour

concentration (an increase of 12%) in different levels of the atmosphere, shifts the peak response, in terms of greenhouse effect [2], from the MIR to the FIR as the water vapour increases with the atmospheric levels. This sensitivity test has also shown that the changes in lower stratosphere water vapour can only be observed in the FIR, and require high radiometric accuracy and spectral resolution to be identified.

Solomon et al. (2010) showed that the stratospheric water vapour is an important driver of decadal global surface climate change. The authors demonstrate that slight variations in water vapour (less than 1 ppm) in the lower stratosphere represent an important source of the decadal variability in the surface temperature. Furthermore, Held and Soden (2000) highlighted the status of the water vapour as the most important greenhouse gas. They showed that a relatively small change in the water vapour concentration is required to match the integrated effect of a doubling in CO_2.

It is important to remember that the water vapour and the CO_2 have a known positive feedback. The atmospheric warming due to increasing levels of CO_2 is amplified if more water vapour enters the atmosphere as a result, i.e., as CO_2 concentrations rise in the atmosphere, the atmospheric and surface temperatures rise, increasing the evaporation from the oceans, and so rising the water vapour concentration in the atmosphere (Peixoto and Oort, 1992). This additional water vapour enhances the original greenhouse warming and is a major contribution to global climate change (IPCC, 2013).

In terms of climate responses/feedbacks, the research from Huang et al. (2010) is a good example of the importance of the FIR to improve our climate knowledge. The authors used OLR spectrum to isolate and quantify the contributions of the temperature, water vapour and clouds, on different atmospheric levels, to the total OLR change in clear and all-sky conditions, assuming a response of doubling CO_2, i.e., comparing the climate simulation using the CO_2 concentration of the pre-industrial conditions (280 ppm) with the double of this initial concentration (560 ppm). Their results showed that a substantial proportion of all feedbacks, and indeed the CO_2 forcing, occurs within the FIR part of the spectrum, although the accuracy in surface temperature and cloud feedbacks is limited.

The clouds have one of the greatest importance as regulators of the ERB. Their presence increase the percentage of solar flux that is reflected back to space, which reduces the solar flux available to the earth-atmosphere system and causes a cooling of the system. On the other hand, the clouds reduce the thermal radiation emitted to space, by absorbing the IR flux from the earth's surface and the atmosphere below the cloud layer, and also by emitting thermal radiation at normally colder cloud-top temperatures. This effect, known as the *IR greenhouse effect*, increases the radiation budget and generally results in a warming of the Earth's atmosphere (Liou, 2002). The

---

[2]The greenhouse effect parameter is defined by Raval and Ramanathan (1989) as the infrared radiation energy trapped by atmospheric gases and clouds, calculated through the subtraction of the surface emission (E_ν) by the outgoing longwave radiation at the TOA (F_ν), monochromatically given by $G_\nu = E_\nu - F_\nu$

balance between these effects is very important to Earth's climate system and varies with the cloud height, microphysical properties, phase, and geometrical thickness, besides the background atmosphere in which the cloud is embedded. Hartmann (1994) indicates, for instance, that a low cloud over the ocean reduces the net radiation, because it increases the albedo without affecting significantly the longwave flux at the TOA, while high and thin clouds, such as the cirrus clouds, can greatly change the longwave fluxes without affecting significantly the solar absorption, causing an increase of the net radiation, which leads to surface warming.

The greenhouse effects of both water and ice clouds have been recognized to be of particular importance in the MIR atmospheric window and in the FIR spectrum regions (Rizzi et al., 2002). However, the knowledge about the cloud fields, their radiative properties, their contribution to the radiation budget and how this may evolve in a changing climate carries yet many uncertainties in climate predictions (ESA, 2019). IPCC (2013) reports that the role of thin cirrus clouds for cloud feedback is yet unknown and remains as a source of possible systematic bias for the Global Climate Models (GCM). In this sense, section 1.2 presents studies regarding the relationship between FIR and clouds.

Finally, a more recent feedback mechanism, which has particular relevance for polar regions and for the FIR, was identified by Feldman et al. (2014). The authors introduced the feedback mechanism regarding the relation ice-emissivity, in which the sea ice melts, in response to an initial surface warming. The melting exposes the less emissive ocean below the ice, which results in an additional surface warming. This identification was made by using simulations with the Community Earth System Model (CESM), in which was used spectrally varying surface emissivity in the FIR and compared it with a blackbody surface emission. Their results show an increase in the decadal average Arctic surface temperatures of up to 2K, concomitantly with decreases of up to 15% in Arctic sea-ice, after only 25 years of model integration.

Then, considering that climate feedbacks are usually derived from climate model simulations, and since the vertical structure of the atmosphere is reflected in the spectral distribution of the outgoing fluxes, studies on the FIR and climate become essential to improve our knowledge, and reduce the uncertainties of the climate predictions.

1.2   FIR and clouds

Clouds regularly occupy at least 50% of the sky on a global scale and are the most important regulators of the ERB (Liou, 2002). As previously explained, they affect the earth's energy balance by reflecting and scattering solar radiation and by absorbing the earth's infrared radiation (Ahrens, 2011).

The net radiation available in the atmosphere, as well as the differential heating of this component of the climate system, is strongly regulated by clouds in terms of their horizontal extent, vertical position, thermodynamic phase, liquid and ice water content,

and particle size distribution (Liou, 2002).

Hartmann (1994) indicates that clouds become opaque to long-wave radiation when the liquid water path exceeds about 20 gm^{-2}. If this value is reached in an altitude where the temperature is essentially uniform, it can be assumed that the cloud surface absorb and emit terrestrial radiation like a black-body. On the other hand, the albedo of clouds continues to increase with additional liquid water content, even after the cloud has become opaque to long-wave radiation. The net radiation effect between the albedo and the *IR greenhouse effect*, however, is yet not completely understood, mainly when it comes to cirrus clouds. Indeed, IPCC (2013) reported that "The role of thin cirrus clouds for cloud feedback is not known and remains a source of possible systematic bias". In this context, Hong et al. (2016) found that the global net radiative effect of ice clouds tends to warming due to a much higher occurrence of thin clouds compared to thick ice clouds.

According to Veglio (2013), the cirrus can be described as cold and optically thin clouds, composed of ice crystals, that reside in the upper parts of the troposphere. This type of cloud covers more than 20% of the Earth surface, and represents about 40% of the clouds in the globe (Liou, 1986; Wylie et al., 1994, 2005).

The importance of cirrus clouds in shaping the Earth's outgoing long wave spectral radiation and the total energy balance has been highlighted by many authors (Kiehl and Trenberth, 1997; Cox et al., 2010). The optical depth of this type of cloud is very important to define its radiative forcing effect. If the cirrus is an optically thin cloud, it transmits the incident solar radiation and absorb the long-wave Earth's surface radiation. As it is a cold cloud, a small amount of long-wave radiation is emitted back to space, which results in a warming of the Earth's atmosphere. By the other hand, when the cirrus cloud is optically thick, the cloud albedo is higher, and the reflection of the solar radiation becomes greater than the long-wave absorption, resulting in a cooling of the Earth's surface (Baran, 2009). Hence, the cirrus clouds play an important role in the infrared atmospheric outgoing radiation and on the energy balance of the atmospheric layers.

Choi and Ho (2006) evaluated the radiative effects of cirrus clouds during six years (2000-2005) over the tropical region (latitude zone of $\pm 25^o$), using data from the Moderate Resolution Imaging Spectroradiometer (MODIS) and the Clouds and the Earth's Radiant Energy System (CERES). They found that the optically thin cirrus (considered in their work as the cirrus cloud with OD<10) has a significant effect on the tropical energy balance, where more than 60% of the ice clouds are thin cirrus cloud. Later, Hong and Liu (2015) showed that globally radiative forcing due to ice clouds introduces a net warming of the earth–atmosphere system, mainly due to the thin and sub-visual cirrus clouds (visible OD<0.3), which contribute to around 40% of the total ice cloud occurrence in the globe.

These studies are very important to help us better understand the impact of the cirrus cloud in the ERB. In this sense, the FIR can also provide a very important contribution

to our knowledge, besides helping to identify the cloud presence and to retrieve cloud properties from spectrally resolved measurements.

Rizzi and Mannozzi (2000) showed that the increase of cloudiness rises the contribution of the FIR to the spectrum in outgoing TOA fluxes, mainly due to the cold top cloud emission. Maestri and Rizzi (2003) contributed to clarify the importance of the FIR in the presence of a cirrus cloud addressing the problem of radiative cooling. They showed that the presence of a cloud layer changes the diabatic heating of all layers below the cloud. Moreover, they observed that the atmospheric concentration of water vapour has an important impact on the cloud's energy balance, since the more the atmosphere below the cirrus is transparent, the more energy is absorbed by the cloud. The authors analysed not only the impact of the cloud in the FIR part of the spectrum, but also highlighted the interaction between water vapour and clouds.

Efforts to constrain the radiative impact of cirrus clouds have primarily focused on the visible and MIR spectral regions, due to the data availability from both aircraft and satellite observations, which have allowed the development and validation of cirrus scattering models (Baran, 2012).

However, many authors have recognized the importance of the FIR to improve the retrieval of the cirrus cloud geometrical (i.e. altitude), microphysical (i.e. effective dimension), and radiative (i.e. optical depth) properties, either from downward or upward radiation measurements (Bantges et al., 1999; Baran, 2005; Cox et al., 2010; Merrelli, 2012; Palchetti et al., 2016; Maestri et al., 2014; Di Natale et al., 2017; Maestri et al., 2019a; Bantges et al., 2020). This type of measurement can contribute significantly to a better understanding of the ERB, and improve the radiation models, hence, the climate models.

Simulations have also been used to show that FIR can contribute to improve the detection of thin cirrus cloud (Maestri et al., 2019b,a; Magurno et al., 2020) and that FIR and MIR measurements can be combined to efficiently retrieve the cloud properties and the atmospheric state (water vapour and temperature profiles) (Palchetti et al., 2016; Di Natale et al., 2017).

The detection of thin cirrus clouds and the derivation of some of their features have been improved by using active satellite sensors (Nazaryan et al., 2008; Sassen et al., 2008; Veglio and Maestri, 2011). Additionally, Magurno et al. (2020) highlighted that the application of a cloud identification methodology directly to punctual infrared hyperspectral radiance measurements, and the subsequent derivation of the main cloud properties from a single spectrum, allow to avoid spatial and temporal colocation and biases due to the use of multiple and heterogeneous instrumentation.

1.3   Objectives and thesis structure

Inserted in the context of the FORUM mission studies devoted to cloud applications, the main goal of this thesis is to exploit FIR radiances, observed and/or simulated, to characterize the signal of this part of the spectrum in the presence of ice clouds, besides developing tools to detect optically thin and sub-visible cirrus clouds, and classifying clouds in general, contributing to retrieve ice-cloud optical, microphysical and geometric properties.

To reach these objectives, the thesis is structured as follows. In chapter 2 is presented a brief description of the evolution of the Radiative Transfer Equation (RTE), and exposed the Radiative Transfer (RT) models applied in this work. Chapter 3 contains the methodology used to simulate the high resolution spectra covering both the FIR and the MIR. Moreover, sensitivity tests are performed to investigate the signal and information that can be derived from the FIR part of the spectrum. Chapter 4 is dedicated to present the Cloud Identification and Classification (CIC) algorithm (Maestri et al., 2019b), which is a machine learning algorithm based on Principal Component Analysis (PCA), that performs cloud detection from high spectral resolution radiances, covering the MIR and the FIR part of the spectrum. Then, in chapter 5 is presented the FORUM end-to-end simulator (FE2ES), a multi-use tool used for parametric, scientific, and industrial studies. The Scene Generator Module (SGM) and the submodule L2M_CIC are inserted in the development of this thesis to evaluate the CIC performances and highlight impacts on the cloud identification to retrieve atmospheric and surface parameters, from high spectral resolution radiances. Chapter 6 shows the results discussed in Magurno et al. (2020). In this manuscript, the CIC algorithm is used for the first time to identify cirrus clouds by using upward radiances measurements. Chapter 7 presents results using the CIC to detect ice and water clouds from ground-based measurements in Antarctica. Finally, in chapter 8 a summary of the main subjects discussed in this thesis is presented.

Chapter 2

Radiative Transfer Modelling

In this chapter is presented briefly the evolution of the RTE and discussed its terms to compute radiances in the FIR and MIR part of the spectrum. Next, a description of the RT models used in the development of this work is made, followed by a presentation of the scattering databases employed in the models to radiances computation.

2.1   Radiative Transfer Equation

The first mathematical and experimental radiative transfer studies to write the RTE were developed by Pierre Bouguer, Johann Heinrich Lambert and August Beer. Bouguer published in 1729 the "Optical Essay on the Dimming of Light", describing your experiment in which the light propagated through a medium decreased on a pattern that could be fit by a logarithmic function. Next, Lambert used the Bouguer's work to put the empirical result into the mathematical form presented in Equation 2.1, where I indicates a light measurement, and K is an attenuation coefficient.

$$I(x) = I(0)e^{-Kx} \tag{2.1}$$

Following, Beer published in 1852 a paper entitled "Determination of the Absorption of Red Light in Colored Liquids", in which he investigated the dependence of light transmission on the concentration various salts in aqueous solution. Beer found the same Lambert's equation to explain the light decrease, however his intepretation includes that the attenuation coefficient k is determined by the concentration of the absorbing material. Then, to give due credit to everyone, this equation is called Beer-Bouguer-Lambert law and is showed in Equation 2.2.

$$I_\lambda(s) = I_\lambda(0)e^{-k_\lambda u} \tag{2.2}$$

where $I_\lambda(0)$ is the incident intensity at $s = 0$ and wavelength λ. k_λ denotes the mass extinction cross-section (in units of area per mass) for radiation of wavelength λ, which is

assumed independent of the distance s in a homogeneous medium. u is the path length, determined by $\int_0^s \rho ds$, with ρ indicating the density of the medium and ds the thickness that the radiance pass through.

Later, Lommel (1889) and Chwolson (1889) introduced the concept of the scattering in the RTE. However, the paper from Schuster (1905) has traditionally been attributed as the first work to highlights the importance of the multiple scattering in the RTE. Furthermore, Schwarzschild (1906) derived a differential equation for the radiance, as a function of depth and direction under the assumption that the sun's atmosphere only absorbed and emitted by blackbody radiation, i.e. the scattering was neglected. In this work, Schwarzschild considered for the first time the hydrostatic equilibrium of the solar atmosphere under radiative equilibrium. The Local Thermodynamic Equilibrium (LTE) was used to determine the temperature distribution in the solar atmosphere. Moreover, he introduced the concept of Optical Depth (OD), defined as an opacity-weighted depth in an absorbing medium (Equation 2.3).

$$\tau_\lambda(s) = \int_0^s k_\lambda \rho ds \tag{2.3}$$

where $\tau_\lambda(s)$ is the optical depth.

Afterward, Schwarzschild (1914) presented the RTE in the form referred as Schwarzschild's equation (Equation 2.4). In this equation, he considered the context of the Kirchhoff's law (Kirchhoff, 1860) and treated the radiances analogous to Planck (Planck and Masius, 1914). The first term in the right-hand side denotes the reduction of the radiant intensity due to absorption, and the second term represents the increase in the radiant intensity arising from blackbody emission of the material.

$$\frac{dI_\lambda}{k_\lambda \rho ds} = -I_\lambda + B_\lambda(T) \tag{2.4}$$

Assuming the OD in its differential form (Equation 2.5), and integrating the thickness ds from 0 to s_1, the integral solution for the Schwarzschild's equation is obtained (Equation 2.6). In Figure 2.1 is shown a scheme of the I_λ passing through an absorbing material, indicating its OD.

$$d\tau_\lambda(s_1, s) = -k_\lambda \rho ds \tag{2.5}$$

$$I_\lambda(s_1) = I_\lambda(0)e^{-\tau_\lambda(s_1,0)} + \int_0^{s_1} B_\lambda(T(s))e^{-\tau_\lambda(s_1,s)} k_\lambda \rho ds \tag{2.6}$$

Schwarzschild (1914) also continued the Schuster's investigation, improving the solution for pure scattering in the RTE. For that, he used a source function (J_λ) in place of the only emission term ($B_\lambda(T)$) of Equation 2.4, as shown in Equation 2.7.
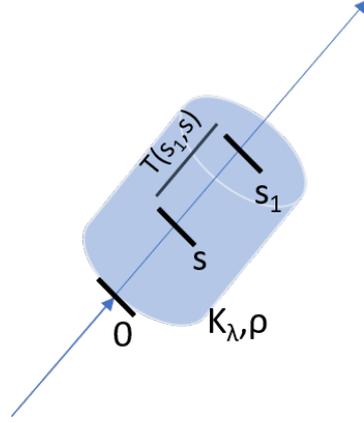
Figure 2.1: Scheme of the I_λ passing through an absorbing material, indicating its OD.

$$\frac{dI_\lambda}{k_\lambda \rho ds} = -I_\lambda + J_\lambda \tag{2.7}$$

Beside Schwarzschild, many others authors contributed for the development of the RT Theory, providing solutions for the RTE, like Louis King, Edward Milne and Richard Gans.

Later, Chandrasekhar (1950) developed numerous solution methods and techniques for the RTE in plane-parallel atmospheres. Equation 2.8 shows the base for the solutions of upward and downward radiances, respectively.

$$I(0, \mu, \phi) = I(\tau_*, \mu, \phi)e^{-\tau_*/\mu} + \int_0^{\tau_*} J(\tau', \mu, \phi)e^{-\tau'/\mu}\frac{d\tau'}{\mu} \tag{2.8a}$$

$$I(\tau_*, -\mu, \phi) = I(0, -\mu, \phi)e^{-\tau_*/\mu} + \int_0^{\tau_*} J(\tau', -\mu, \phi)e^{-(\tau_*-\tau')/\mu}\frac{d\tau'}{\mu} \tag{2.8b}$$

where ϕ is the azimuthal angle; μ is equal to $cos(\theta)$, with θ indicating the inclination to the upward normal, i.e. the zenith angle; τ is the atmospheric optical depth, bounded on two sides ($\tau = 0$ and $\tau = \tau_*$), as shown in Figure 2.2. In Equations 2.8a and 2.8b, the subscript λ is omitted.

The source function (J) is treated basically as the sum of three terms as indicated in Equation 2.9. The first one is the isotropic emission term. The second one accounts for the direct solar irradiance coming from direction (μ_0,ϕ_0), transmitted through the atmosphere with transmittance $exp(-\tau/\mu_0)$, and diffused in direction (μ,ϕ) within the
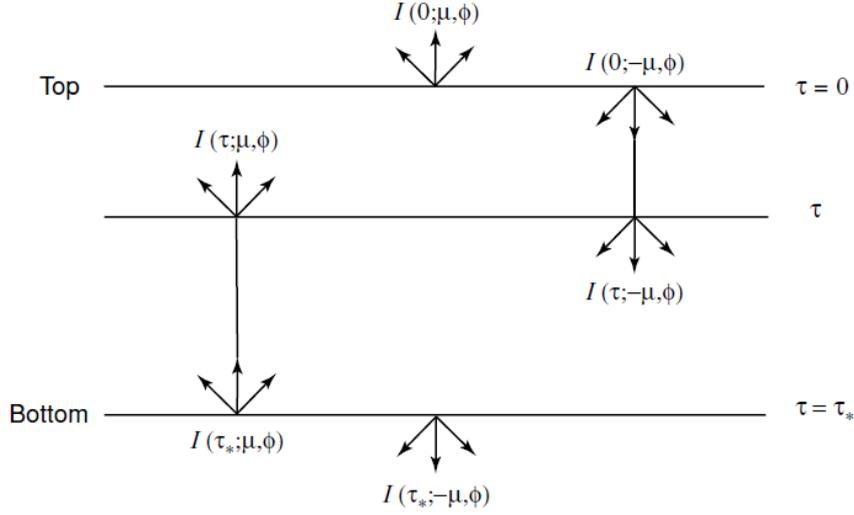
Figure 2.2: Upward (μ) and downward (-μ) intensities at a given level τ and at top (τ=0) and bottom ($\tau = \tau_*$) levels in a finite, plane-parallel atmosphere. From Liou (2002).

volume of air whose the vertical coordinate is τ. The last one corresponds to the scattering term, where the radiances that arrive from the direction (μ',ϕ') interacts with the volume at optical depth τ and is scattered into direction (μ,ϕ).

$$J = J_{emi} + J_{dir} + J_{scat}$$

$$J = (1 - \tilde{\omega})B\left[T(\tau)\right] - \frac{\tilde{\omega}}{4\pi}SP(\Omega, -\Omega_0)e^{-\tau/\mu_0} - \frac{\tilde{\omega}}{4\pi}\int_{4\pi}I(\tau, \Omega')P(\Omega, \Omega')d\Omega' \qquad (2.9)$$

In Equation 2.9, Ω corresponds to the pair of zenithal and azimuthal angles (μ and ϕ, respectively); $\tilde{\omega}$ is the single scattering albedo, which determines the amount of energy diffused in all directions and is computed by the ratio between the scattering and extinction coefficients; S is the solar irradiance at the TOA; $P(\Omega, -\Omega_0)$ and $P(\Omega, \Omega')$ are the phase functions, which determines the angular distribution of scattered energy within the 4π solid angle around the volume of air.

Taking as example the upward radiance, which is calculated by using Equation 2.8a combined with Equation 2.9, the differential change of diffuse radiance emergent from below the layer will be driven by the following processes (Figure 2.3):

1. reduction from the extinction attenuation;

2. increase from the single scattering of the unscattered direct solar flux from the direction (-μ_0,ϕ_0) to (μ,ϕ);
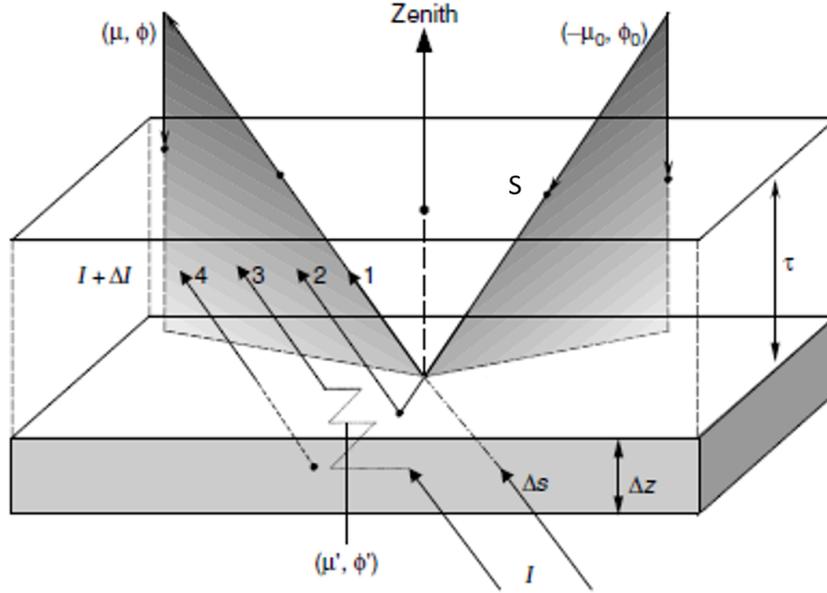
Figure 2.3: Transfer of diffuse radiance from below in plane-parallel layers: (1) attenuation by extinction; (2) single scattering of the unscattered solar flux; (3) multiple scattering; and (4) emission from the layer. All the radiative parameters are defined with reference to a small volume containing a spectrum of molecules and/or particulates. The notations are defined in the text. From Liou (2002).

3. increase from multiple scattering of the diffuse intensity from directions $(-\mu',\phi')$ to (μ,ϕ); and

4. increase from emission within the layer in the direction (μ,ϕ).

According with Liou (2002), the main parameters that will drive the radiances in the RTE are the optical depth, the single-scattering albedo, and the phase function. Moreover, in the thermal infrared, the second term of the source function (Equation 2.9) is suppressed, and it suffices to take the azimuth-independent component. Then, the source function may be expressed as:

$$J = (1 - \tilde{\omega})B\left[T(\tau)\right] + \frac{\tilde{\omega}}{2} \int_{-1}^{1} I(\tau, \mu')P(\mu, \mu')d\mu' \qquad (2.10)$$

The atmospheric gaseous optical depth is given by the sum of each optical depth line (τ_j), which accounts for each wavenumber and atmospheric specie (Equation 2.11). In this case, the absorption coefficient must be computed at wavenumber intervals that are smaller than the line half-width, and it can be expressed in terms of line strength $(S_j(T))$ and line shape $(f_{\nu,j}(p,T))$, as shown in Equation 2.12.

$$\tau = \sum_{j=1}^{N} \tau_j = \int_u \sum_{j=1}^{N} k_{j,\nu}(u) du \tag{2.11}$$

$$k_\nu(p,T) = \sum_{j=1}^{N} S_j(T) f_{\nu,j}(p,T) \tag{2.12}$$

In this work, the atmospheric gaseous optical depth are calculated by using the LBLRTM model, which is based in an algorithm for the accelerated convolution of line shape functions with spectral line data, and includes the continuum absorption (Clough et al., 1981, 1992). More detailed characteristics of the model are presented in Section 2.2.1.

The phase function can be expressed in terms of Legendre Polynomials (P_l), based on the addition theorem for spherical harmonics (Equation 2.13).

$$P(\cos \Theta) = \sum_{l=0}^{N} \tilde{\omega}_l P_l(\cos \Theta) \tag{2.13}$$

where Θ is the scattering angle and $\tilde{\omega}_l$ is the expansion coefficient given by:

$$\tilde{\omega}_l = \frac{2l+1}{2} \int_{-1}^{-1} P(\cos \Theta) P_l(\cos \Theta) d \cos \Theta \tag{2.14}$$

Assuming the azimuth-independent component in the thermal infrared, the phase function is computed by:

$$P(\mu, \mu') = \sum_{l=0}^{N} \tilde{\omega}_l P_l(\mu) P_l(\mu') \tag{2.15}$$

Then, combining Equations 2.8a, 2.10 and 2.15 give us the RTE in the thermal infrared. However, there is no analytical solution for that. In this way, many methods can be applied to solve it numerically (Sun et al., 2017).

2.2 Radiative Transfer Models

In this work, the multiple scattering radiances are computed by using the model RTX (Evans and Stephens, 1991), based on the doubling and adding method, and the LBLDIS (Turner et al., 2003; Turner, 2005) system, which uses the Discrete Ordinates Radiative Transfer (DISORT) model (Stamnes et al., 1988), based on the discrete-ordinate method. Additionally, the LBLRTM model is applied to compute the atmospheric optical depth used as input for the previous multiple scattering RT models. Next, a brief description of each one of these RT models is presented.

2.2.1   LBLRTM

The LBLRTM is an accurate, flexible and efficient line-by-line radiative transfer model, derived from the Fast Atmospheric Signature Code (FASCODE, Clough et al., 1981, 1992). This model is used to compute radiances, in clear sky conditions, and atmospheric optical depth, being extensively validated from the UV to the sub-millimeter frequencies (e.g. Turner et al., 2004; Alvarado et al., 2013; Liu et al., 2015).

The main features of LBLRTM are described in Clough et al. (2005) and are highlighted following. The Voigt line shape is calculated at all atmospheric levels and provides the foundation for the LBLRTM line shape. Modification to the Voigt are implemented accounting for the database updates. Spectroscopic parameters are used including the pressure shift coefficient, the half-width temperature dependence and the coefficient for the self-broadening of water vapour. The model also incorporates the MT_CKD continuum model (Mlawer et al., 2012), which includes self- and foreign-broadened water vapour continua as well as continua for CO_2, O_2, N_2, O_3. The continuum model describes the contribution from each spectral line as the sum of a sub-Lorentzian line shape and a small, broad additional line shape, that provides the needed super-Lorentzian absorption in the intermediate line wings (Harries et al., 2008).

The latest version of the LBLRTM model (v12.9) uses the Atmospheric and Environmental Research (AER) v3.7 spectroscopic database. This database is built from High-resolution transmission molecular absorption (HITRAN) 2012 (Rothman et al., 2013), with exceptions for H_2O, CO_2, CH_2 and O_2.

The LBLRTM's line parameter inputs are obtained from the LNFL program, which utilizes the AER line parameter database to generate the adequate line parameter binary file for LBLRTM, called *TAPE3*. The model utilizes temperature dependent cross sections to treat the absorption due to heavy molecules, such as halocarbons. The input cross-section data for all temperatures are at the lowest pressure available in HITRAN database, and the pressure scaling of the cross-section is then performed by a convolution of the cross-section spectrum with an appropriate Lorentz function. Moreover, the LBLRTM utilizes the Kurucz solar source function (Kurucz, 1994), and an algorithm is implemented for the treatment of the variation of the Planck function within a vertically inhomogeneous layer.

Besides the *TAPE3* file, the LBLRTM uses as input a configuration file called *TAPE5*. This file includes information about the observer, such as the altitude of the observer, the view angle, and the altitude of the boundary (either surface or TOA), and the atmospheric vertical profiles of thermodynamics parameters (temperature and pressure) and of the gaseous concentration at each layer height in which the radiance is calculated.

Additional information about the cross sections of heavy molecules, like the CFC's can be provided for the model through the file named *FSCDXS*, which contains the alias names of the molecules that have their cross sections properties available inside of the *xs* folder.
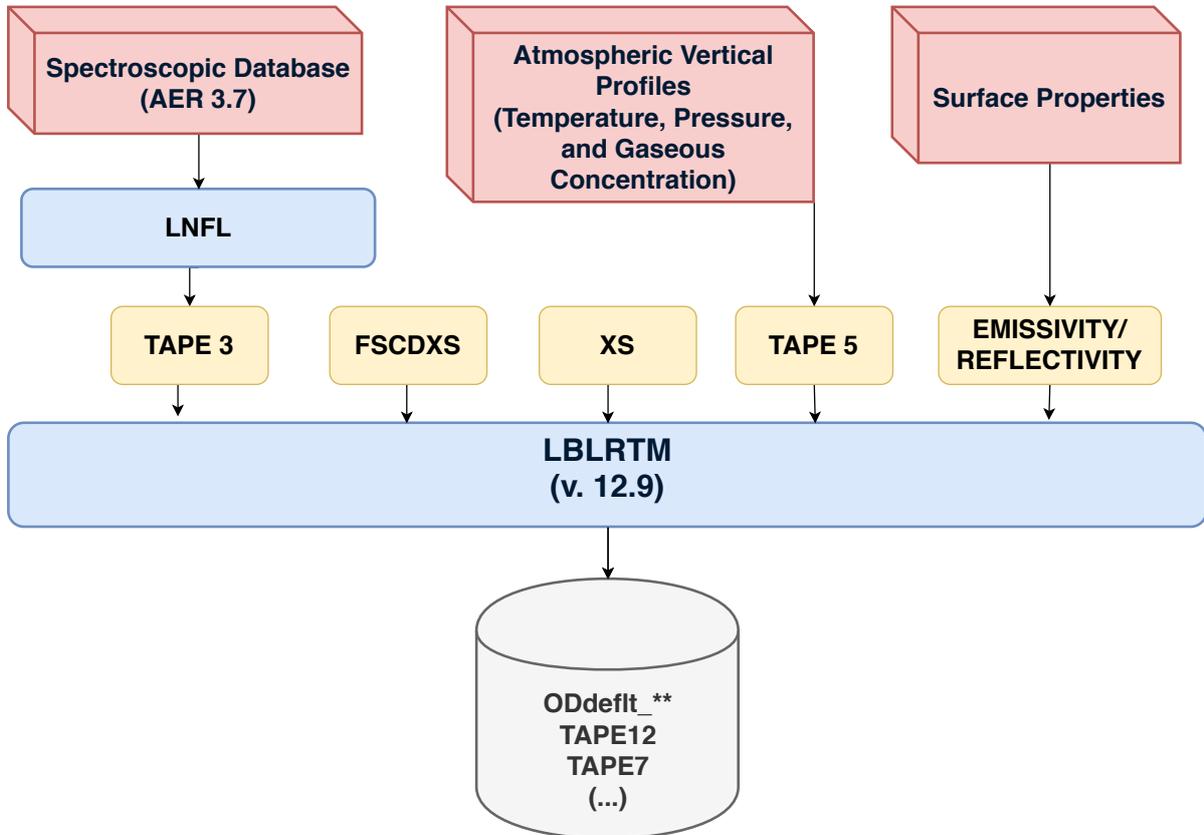
Figure 2.4: LBLRTM's Flowchart, highlighting the input and output files. Red box are used to indicate external databases, blue rectangles indicates computational modules, yellow rectangles are used for input files and the grey cylinder is the output.

Then, all these input files are used to compute the optical depth of each atmospheric layer defined in the *TAPE5* file. However, to compute the radiances in clear sky conditions it is also necessary to include the files containing the spectral emissivity and reflectivity of the boundary (surface or TOA). The calculated radiances from the LBLRTM model are output into the binary file *TAPE12*, while the optical depths are stored in binary files for each atmospheric layer, named *ODdeflt_***, where the **** corresponds to a number referring the atmospheric layer. In Figure 2.4 is shown the LBLRTM's flowchart.

2.2.2   LBLDIS

The LBLDIS is a system developed by David Turner (Turner et al., 2003; Turner, 2005), which combines the LBLRTM and the DISORT models. Its current version (3.0) is able to compute radiances for multiple scattering atmospheres from the UV to the radar

region of the electromagnetic spectrum. As the LBLRTM model has been previously discussed, now a brief description of DISORT is presented.

Laszlo et al. (2016) made a good summary of DISORT, and the main characteristics of the model presented in their work will be outlined following. DISORT has been designed to be an example of good scientific software with clear in-code documentation, as well as to be a scientific tool. It is a discrete ordinate algorithm for monochromatic unpolarized radiative transfer in non-isothermal, vertically inhomogeneous, but horizontally homogeneous media.

The code can treat thermal emission, absorption, and scattering with an arbitrary phase function. The medium may be driven by parallel or isotropic diffuse radiation incident at the top boundary, by internal thermal sources and thermal emission from the boundaries, as well as by bidirectional reflection at the surface.

The discrete ordinate method implemented in DISORT follows the classical approach of transforming the radiative transfer equation into a system of ordinary differential equations by expanding the phase function into a series of Legendre polynomials and the intensity into a Fourier cosine series. The resulting system of equations is then written in matrix form and the solution of the homogeneous equations is formulated as a standard algebraic eigenvalue problem, which is solved by using a robust linear algebra solver. Its solutions are closed-form analytic functions of optical thickness, then the computation time is basically independent of optical thickness, but it does depend linearly on the number of vertical layers (Laszlo et al., 2016).

The LBLDIS uses the DISORT version 2.0, which was compiled using double precision to solve some issues of the numerical precision. The output files *TAPE7* and *ODdeflt_***, from LBLRTM, are used as input of LBLDIS. The file *TAPE7* contains the atmospheric vertical profiles of the temperature, pressure and gaseous mixing ratio to each at atmospheric layer, while the files *ODdeflt_*** correspond to the optical depth of each atmospheric layer. Moreover, two input files are required:

1. a configuration file in which the user specifies information about the zenithal and azimuthal angles, surface emissivity and temperature, beside the cloud layer (cloud optical depth and effective radius)

2. a single scattering property database, wherein a size distribution is computed for a given effective radius and the cloud particle scattering properties are integrated over the size distribution for each wavelength.

Single scattering properties from different databases can be used as auxiliary file to perform a simulation with LBLDIS. A description of the databases used in the cloud simulations is presented in Section 2.3. Finally, in Figure 2.5 is shown the LBLDIS's flowchart.
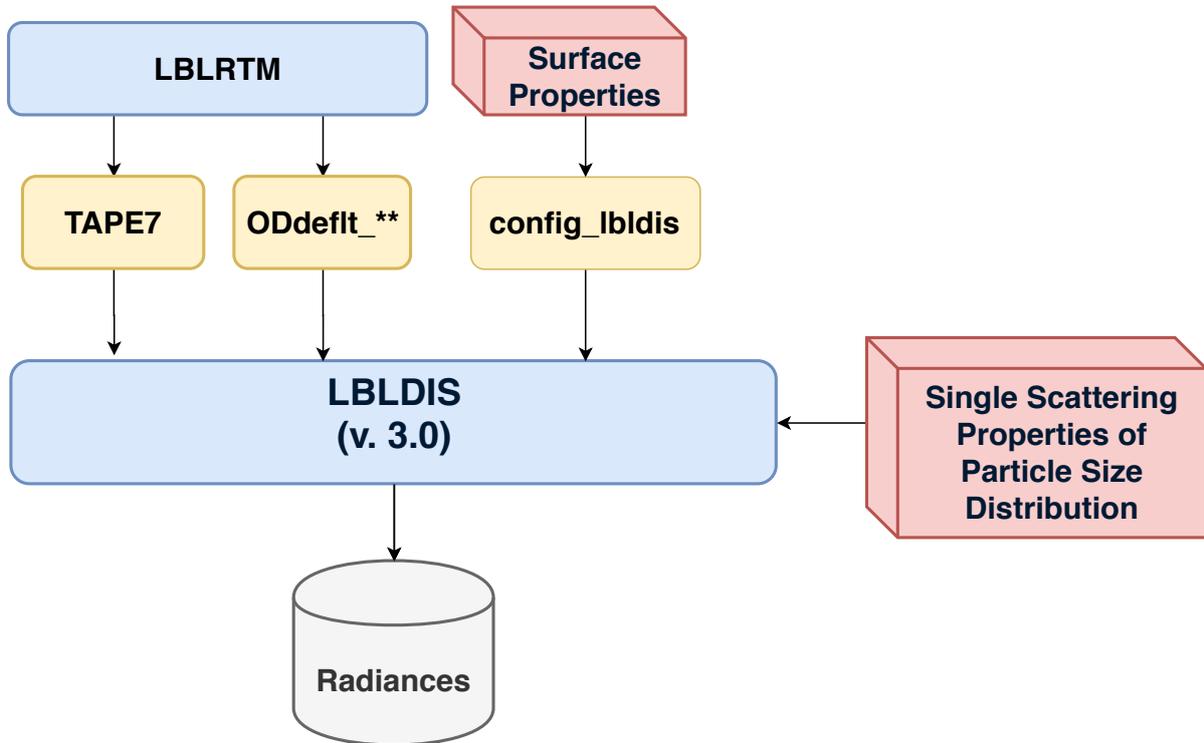
Figure 2.5: LBLDIS's Flowchart, highlighting the input and output files. Red box are used to indicate external databases, blue rectangles indicates computational modules, yellow rectangles are used for input files and the grey cylinder is the output.

2.2.3   RTX

RTX is a modified version of the RT3 model. The RT3 model is a plane-parallel fully-polarized atmospheric RT model that calculates the monochromatic radiation for isotropic media. The model can simulates both solar and thermal sources of radiation. The multiple scattering is calculated based on the doubling/adding approach, considered numerically stable for large optical depths. The original version of the RT3 model was developed by Evans and Stephens (1991), and it was modified allowing the parallelization of the code, named here as RTX.

The key concept behind the doubling and adding method is the approach referred as *interaction principle*, which expresses the linear interaction of radiation with a medium, i.e., the radiation emerging from a layer is related to radiation incident in the boundary and to radiation generated within the layer. This method is applied in the RT3 model dividing non-homogeneous atmosphere into N homogeneous layers, characterised by a single value of single scattering albedo, scattering diagram and optical depth, temperature and concentration of all gases and materials that are optically active. For each layer, the *reflection matrix*, the *transmission matrix* and the *source vector* are computed to

solve the RTE. However, it is necessary first to convert the single-scattering information into a form suitable for applying the interaction principle, by a perform of polarization transformation from the *phase matrix* to the *scattering matrix*. If the layer doesn't scatter, the model uses a subroutine to calculate the reflection and transmission matrices and source vector rather than using initialization and doubling (Evans and Stephens, 1991).

From an initial infinitesimal sub-layer, the doubling method is used to builds up the radiative properties of the finite homogeneous layer performing a number of steps depending on the sub-layers thickness. For each output level an adding subroutine is used to combine the layers above and below the output level. Then an internal radiance subroutine is applied to compute the radiance at the output level from the reflection and transmission matrices, and source vectors for the medium above and below and the radiance incident from the boundaries. The sources of radiation are solar direct beam and thermal emission, and there is assumed to be thermal and/or reflected direct solar radiance from the lower surface. The model incorporates two types of surfaces: Lambertian and Fresnel. The reflection from the lower surface is simply treated as another layer in the medium, with unity transmission and no source term. A more complete explanation of the application of the doubling and adding method in the RT3 model is presented in the reference text of Evans and Stephens (1991).

The input of the model are:

- **fkext**: a binary file containing the gaseous atmospheric optical depth for each atmospheric layer and the surface reflectivity;

- **master.wav**: a binary file containing the spectral information used in the simulation;

- **profilo**: an ASCII file containing the vertical profiles of the temperature, pressure and ice/liquid water content. In this file is also indicated the name file of the auxiliary database of the scattering properties of the cloud particles;

- **Scattering Database**: a binary file containing the scattering properties of the cloud particles (more information about the database used are presented in Section 2.3);

- **Configuration Input File**: a file containing the user's options to execute the model.

In Figure 2.6 is shown the RTX's flowchart.
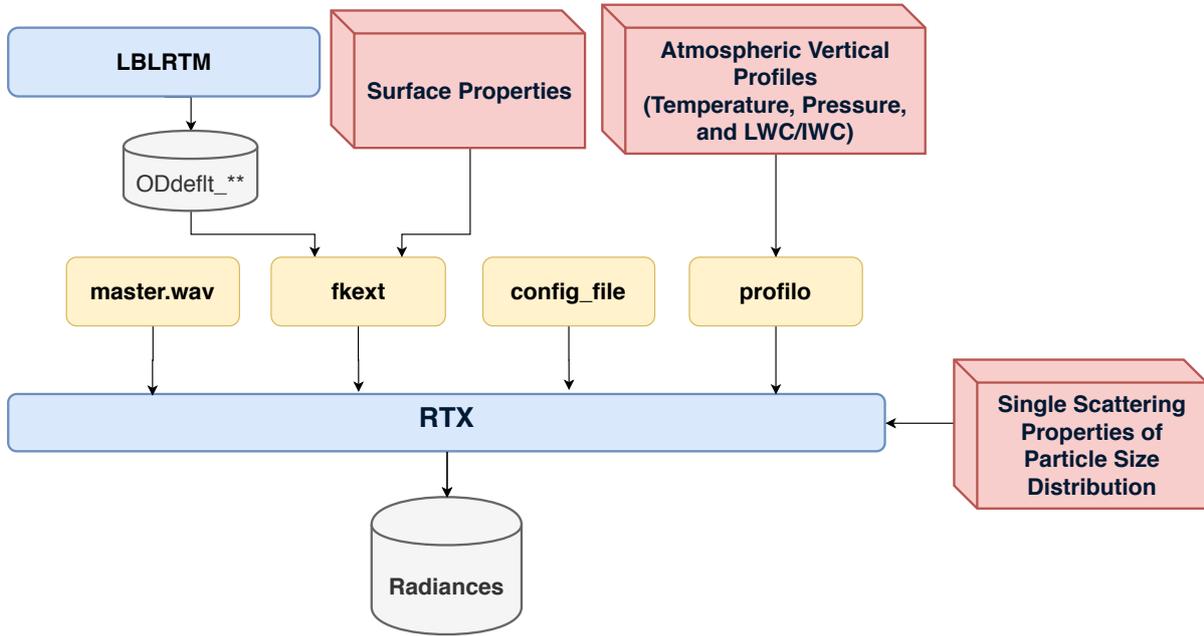
Figure 2.6: RTX's Flowchart, highlighting the input and output files. Red box are used to indicate external databases, blue rectangles indicates computational modules, yellow rectangles are used for input files and the grey cylinder is the output.

2.3   Scattering Database

The ability of the multiple scattering RT models to accurately simulate the cloud radiances relies heavily upon the scattering database used (Fox, 2015).

In this work, the single scattering properties database for single spherical particles of liquid water, ice and mixed-phase, are derived through the *scattnlay* code (Peña and Pal, 2009). The *scattnlay* performs calculations of Mie scattering coefficients and efficiency factors for a single multilayered sphere, combined with standard solutions of the scattering amplitude functions. The efficiency factors calculated are the extinction efficiency, the scattering efficiency and the radiation pressure efficiency, which allow to compute the absorption efficiency, the single scattering albedo and the asymmetry parameter.

To compute the properties of ice spheres, refractive indices from Warren and Brandt (2008) are used, while the refractive indices used for liquid water are obtained from Downing and Williams (1975). The computation for multilayered spheres is an advantage of the *scattnlay* code, and allow us to compute properties of mixed-phase sphere particles, i.e., an ice sphere coated by liquid water using a variable percentage of the coated area. A detailed explanation about the *scattnlay* algorithm is presented in its reference paper (Peña and Pal, 2009), and in Figure 2.7 is presented a simplified flowchart of the *scattnlay* code's input-output. The computation were performed using 361 scattering angles
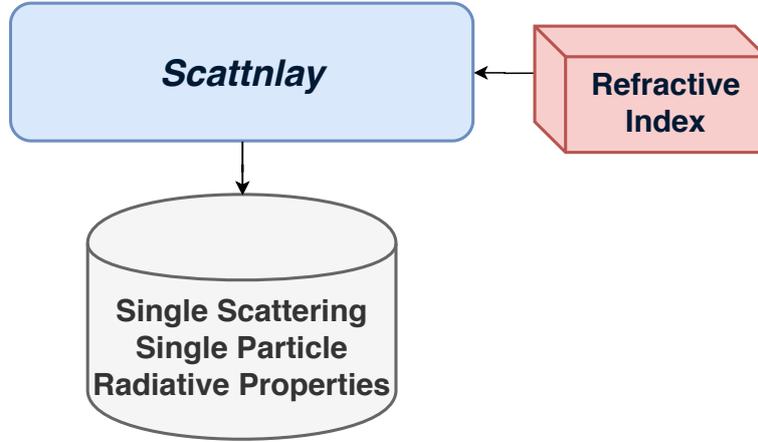
Figure 2.7: Scattnlay code's flowchart, highlighting the input and output files. Red box are used to indicate external databases, blue rectangles indicates computational modules and the grey cylinder is the output.

between 0° and 180°, and particle radius varying between $0.01\mu m$ and $40\mu m$.

For non-spherical ice crystals, the single scattering single particles properties are derived from the Ping Yang database (Yang et al., 2013). This is a full set of single-scattering properties considering three dimensional random orientations for 11 ice crystal habits: compact aggregates composed of eight solid columns (assumed in this work simply as aggregate), hexagonal plates, solid and hollow columns, solid and hollow bullet rosettes, droxtals, prolate spheroids, oblate spheroids, small and large spatial aggregates composed of 5 and 10 plates, respectively (see Figure 2.8). For each ice crystal habit, three surface roughness are considered (smooth, moderately roughened, and severely roughened), and its maximum dimension ranges from $2\mu m$ to $10000\mu m$, in 189 discrete sizes. The database contains the extinction efficiency, the single scattering albedo, the asymmetry parameter, the particle projected area, the particle volume, and six independent nonzero elements of the phase matrix (P11, P12, P22, P33, P43, and P44) with 498 scattering angles between 0° and 180°.

Both spherical and ice crystals single scattering single particles properties are integrated on the Generalized Gamma Particle Size Distribution (PSD) proposed by Hansen (1971) (Equation 2.16).

$$n\left(r\right) = n_0 r^\mu e^{-\lambda r} \tag{2.16}$$

where n_0 is a normalisation factor; r is the radius of the equivalent sphere of single particle, being equal to the particle radius for the *scattnlay* database, and assumed $r = \frac{3V}{4A}$ for the Ping Yang database, where the V is particle volume and A is the particle projected area; μ is the dispersion (or shape) parameter; and λ is the slope parameter. The dispersion parameter is linked to the effective variance (v_{eff}) of the PSD by the
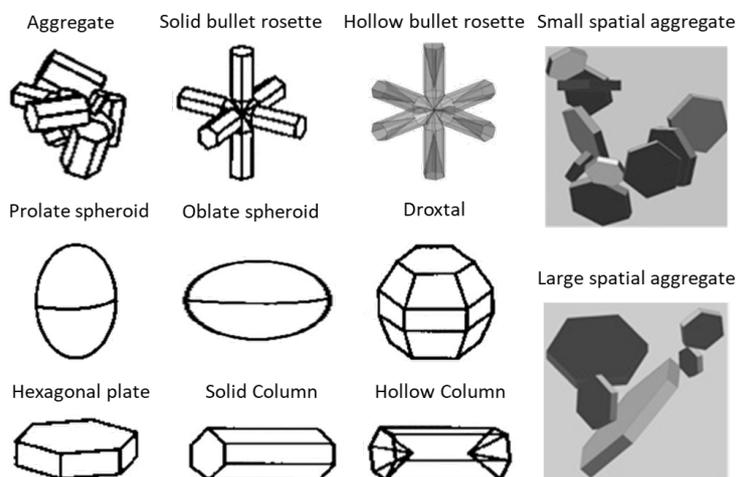
Figure 2.8: Ice crystal habits available from Ping Yang database. Adapted from Yang et al. (2013)

.

relation presented in Equation 2.17, and the slope parameter is given by Equation 2.18.

$$\mu = \frac{1 - 3v_{eff}}{v_{eff}} \tag{2.17}$$

$$\lambda = \frac{1}{r_{eff}v_{eff}} \tag{2.18}$$

where r_{eff} is the effective radius of the PSD.

Finally, the single scattering of PSD properties are computed by Equations 2.19a-2.19g, considering $v_{eff} = 0.1$ in the gamma distribution (Hansen, 1971; Diem, 1948), to get bulk scattering properties of particle size distributions. These properties are used as input in the of the multiple scattering RT models.

$$A = \frac{\int A(r)n(r)dr}{\int n(r)dr} \tag{2.19a}$$

$$Q_{ext} = \frac{\int Q_{ext}(r)A(r)n(r)dr}{\int A(r)n(r)dr} \tag{2.19b}$$

$$Q_{sca} = \frac{\int Q_{sca}(r)A(r)n(r)dr}{\int A(r)n(r)dr} \tag{2.19c}$$

$$\sigma_{ext} = Q_{ext}A = \frac{\int Q_{ext}(r)A(r)n(r)dr}{\int n(r)dr} \tag{2.19d}$$

$$\tilde{\omega} = \frac{\int Q_{ext}(r)\tilde{\omega}(r)A(r)n(r)dr}{\int Q_{ext}(r)A(r)n(r)dr} = \frac{Q_{sca}}{Q_{ext}} \tag{2.19e}$$

$$g = \frac{\int g(r)Q_{ext}(r)\omega_0(r)A(r)n(r)dr}{\int Q_{ext}(r)\tilde{\omega}(r)A(r)n(r)dr} \tag{2.19f}$$

$$P = \frac{\int P(r)Q_{ext}(r)\omega_0(r)A(r)n(r)dr}{\int Q_{ext}(r)\tilde{\omega}(r)A(r)n(r)dr} \tag{2.19g}$$

where A is the projected cross-sectional area, Q_{ext} is the extinction efficiency, Q_{sca} is the scattering efficiency, σ_{ext} is the extinction cross section, $\tilde{\omega}$ is the single scattering albedo, g is the asymmetry parameter, and P is the phase function. The parameters written in function of r, inside of the integral, are related to the single particle with equivalent radius r, while the parameters written without the indicated function r, are related to the PSD with a specific effective radius.

Chapter 3

Simulated Spectra and Sensitivity Tests

The FIR part of the spectrtum is extremely important in cloudy sky conditions, since water and in particular ice particles efficiently absorb and scatter longwave radiation (Rizzi et al., 2002). Refractive indexes are fundamental parameters in the determination of the scattering and absorption properties of the cloud particles. The imaginary part of refractive index determines the absorption coefficient. If there is no absorption, values of the real part are responsible for scattering processing. When absorption is involved, the amount of energy scattered and absorbed depends on both the real and imaginary parts (Liou, 1992).

The imaginary part of the ice's refractive index has a minimum around 410 cm^{-1} (Figure 3.1), which makes the scattering significant. Moreover, the highest absolute difference of this imaginary part between the ice and the liquid water is observed in the range 400-600 cm^{-1}. These characteristics highlight that the absorbing and scattering of ice particles in the FIR can provide additional information to observations in the MIR, regard to cloud identification and their properties.

As there are few high spectral resolution satellite measurements covering the FIR, simulated spectra are used to evaluate the spectrum sensitivity as response of the cloud properties, like the particle phase, shape, and effective radius.

In this chapter the methodology applied to compute the synthetic spectra is presented. This methodology is based on the chain of codes described in Section 2.2. Simulations are performed in order to highlight the sensitivity in the FIR wavenumbers either in clear and cloudy sky conditions. The methodology applied and some of these sensitivity tests were presented in Maestri et al. (2019b), a paper that is part of the development of this thesis.
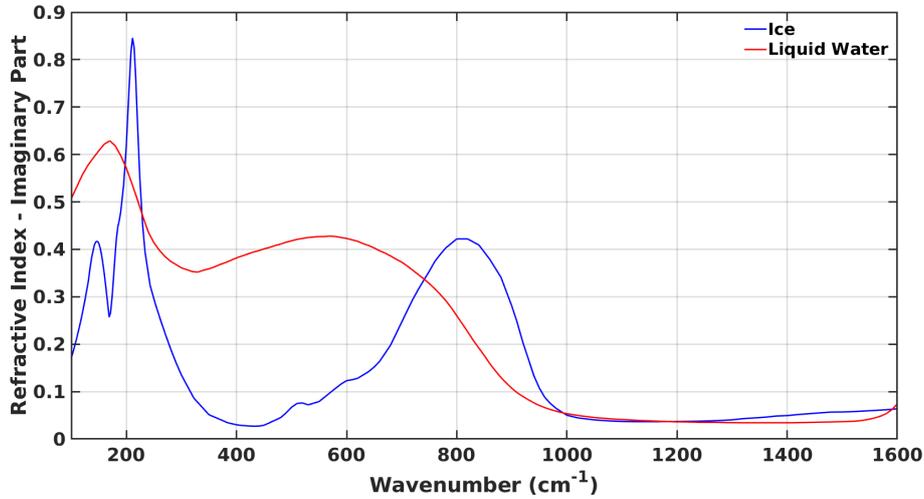
Figure 3.1: Imaginary refractive indices for liquid water (Downing and Williams, 1975) and ice (Warren and Brandt, 2008).

3.1 Simulation methodology: forward model

The computational methodology used to simulate the spectra covering the FIR part of the spectrum by using the multiple scattering models is summarized in Figure 3.2. This structure is used to compute high spectral resolution radiances. Subsequently, the instrument characteristics simulating the FORUM satellite instrument on nadir view is applied.

The line-by-line computations of layer spectral optical depths are performed using the Model LbLRTM v12.7 (Clough et al., 2005), with the AER version 3.6 spectroscopic database, built from HITRAN2012 (Rothman et al., 2013), and including the updated water vapour continuum parameterisation, MT_CKD version 3.2 (Mlawer et al., 2012).

The atmospheric thermodynamic vertical profiles and gas mixing ratios are derived from different sources. The first one is the ECMWF Re-Analysis (ERA) Interim reanalysis (Dee et al., 2011), with a horizontal resolution equal to 0.75 degrees (approximately 80 km) and 60 vertical levels from the surface up to 0.1 hPa. This database, that provides 4 sets of data per day, is used to retrieve profiles of temperature, pressure, specific humidity, ozone mixing ratio, and surface geopotential height from which the geometric surface and atmospheric level heights are computed. Grid points from Tropics, Mid latitudes and Polar regions (Artic and Antarctic) are selected.

The second source of information is the Initial Guess Climatological Database no. 2 (IG2) v5.4 (Remedios et al., 2007) which includes atmospheric profiles for six latitude bands, four seasons, and two times of the day. This database spans from 2007 onwards, with constant vertical resolution of 1 km, from the surface up to 120 km and provides time-averaged data concerning altitude, pressure, temperature and a wide range of gas
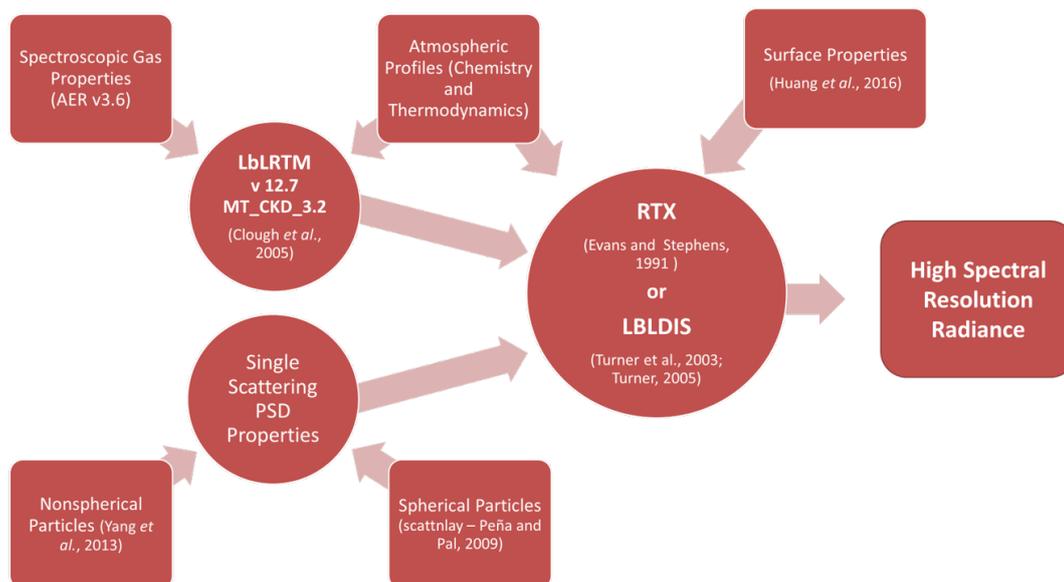
Figure 3.2: Schematic software architecture of the forward models to simulate synthetic spectra covering the FIR part of the spectrum. The circle represent the steps involving the codes, while the boxes are auxiliary datasets, except the output box named "High Spectral Resolution Radiance".

mixing ratio profiles. In this work, 2013's data are used, being that the CO_2 has been linearly corrected to the current concentration level, multiplying by a scale factor to adjust the mixing ratio estimated in 2013 to the concentration level in 2018.

ERA Interim reanalysis data is thus used to characterize the atmospheric profiles of temperature, pressure, water vapour and ozone mixing ratio up to approximately 60 km of height and the IG2 is used to add information in the highest atmospheric layers (up to 80 km) and to provide information concerning minor gases mixing ratios.

Surface emissivity properties are selected in accordance with geolocation by using the global database produced by Huang et al. (2016). The database includes 11 types of spectral emissivity in the infrared region of the spectrum representative of temperate deciduous forest, conifer forest, grass, dry grass, desert, a merged composition of desert and grass, ocean, coarse, medium and fine snow, and ice. An additional rainforest emissivity is also computed, by using the average data from Huang et al. (2016) in the grid points of the Amazon forest.

The cloud microphysical properties of non-spherical ice crystals are generated by integrating the Ping Yang database (Yang et al., 2013), while for liquid water and mixed phase sperical particles the radiative properties are derived through the *scattnlay* code (Peña and Pal, 2009). The description of these properties was presented in Section 2.3. The mixed phase spheres consist of a core of ice surrounded by a coating of liquid water. This shell of liquid is modelled as a coating of 10 or 20% of the radius of the entire

particle.

For each selected atmospheric condition, high spectral resolution optical depths of atmospheric layers are computed by using the LbLRTM model. Its results are passed as inputs to the multiple scattering models (RTX or LBLDIS), and, when in presence of clouds, the gas optical depths are merged with those derived from cloud properties. Then, a multiple scattering model (RTX or LBLDIS) is run to obtain high spectral resolution radiances.

3.1.1   Synthetic Dataset

The synthetic dataset described here, computed by using LBLRTM and RTX models, is used afterward to develop the CIC algorithm, which is discussed in Chapter 4. The spectra are produced using daily ERA Interim reanalysis data from January, April, July, and October 2016, in order to reproduce the seasonal and daily variations of the thermodynamic variables.

In the simulations many cloud properties are varied: cloud top height, geometrical thickness, optical thickness, particle effective dimension, shape, and phase (ice, liquid water, and the two levels of mixed phase). The cloud properties inputs to the radiative transfer code are modified over large ranges of values in order to account for the largest possible variability encountered in nature. Some datasets and recent studies are considered as baseline. The International Satellite Cloud Climatology Project (ISCCP) (Rossow and Schiffer, 1999; Hahn et al., 2001), is an example. Cirrus clouds properties in the Tropics and at Mid Latitudes are mostly based on what was found by Veglio and Maestri (2011), while Antarctic cloud properties are obtained from several sources (i.e., Adhikari et al., 2012; Bromwich et al., 2012; Lachlan-Cope, 2010).

In Table 3.1 the range of variability of some key cloud properties is reported.

After compute the high spectral resolution radiances, the spectra are convoluted with an ideal FORUM like Instrument Line Shape (ILS) that is assumed to be a sinc function. This operation produces unapodised spectra with 0.3 cm^{-1} spectral resolution over a spectral range spanning from 100 to 1600 cm^{-1} representative of FORUM mission noiseless observations.

In a subsequent step, a nominal Noise Equivalent Spectral Radiance (NESR), as the goal reported by ESA (2019), is added to the simulated radiances in order to produce a realistic FORUM observations dataset. The new dataset is computed by adding a gaussian wavenumber-dependent noise to the noiseless radiances. The spectral dependence and amplitude of the noise are derived from the technical specification of the Fourier Transform Spectrometer instrument described in ESA (2019) and reported in Table 3.2. The NESR values reported in the table corresponds to a typical percentage noise of about 1% in the 200-800 cm^{-1} wavenumber interval. The exact value depends on the specific wavenumber and observational conditions accounted for. Below 200 cm^{-1} and above 1400 cm^{-1} the percentage noise can be as high as 15% due to the low values of

Table 3.1: Range of variability of the main cloud properties used in the simulations. The habit type indicates the assumed pristine shapes of the ice crystals. The mixed phase water coating indicate the percentage of liquid water coating with respect to the dimension of the assumed spherical particle.

Cloud property	Range
Top height (km)	1.0–17.0
Geometrical thickness (km)	1.0–5.0
Optical depth at 900 cm^{-1}	0.02–30.0
Habit type	Hexagonal Plate, solid column, solid bullet rosette and aggregate
Mixed-phase water coating	10 % and 20 %
Ice particle eff. dimension (μm)	4–100
Liquid water eff. radius (μm)	3–15
Mixed-phase eff. radius (μm)	3–15

Table 3.2: Random noise used to simulate the Fourier transform spectrometer of the FORUM mission.

Interval (cm^{-1})	Noise $(mW/(m^2 src m^{-1}))$
100-200	1.0
200-800	0.4
800-1600	1.0

the radiance.

A spectral random noise is computed for each spectrum. The Central Limit theorem is used so that the sum of random numbers (r_{tot}) from a uniform distribution ranging from -0.5 to 0.5 (variance is 1/12) is used to generate a gaussian variable (r_{gauss}) with mean 0 and a standard deviation, σ_ν, assumed to be equal to the FORUM noise. The spectral random noise is thus obtained by using the formula presented in Equation 3.1

$$r_{noise} = \sqrt{\frac{12}{r_{tot}}} \cdot \sigma_\nu \cdot r_{gauss} \tag{3.1}$$

A scheme to build the synthetic dataset, from high spectral resolution radiances, is shown in Figure 3.3. A schematic summary of the whole dataset, comprising of 4244 simulated spectra, is presented in Table 3.3 for each latitude belt and for the clear or cloudy class. The dataset is not equally balanced between the types of clouds, to highlight cirrus cloud, specially sub-visible ice clouds (ice clouds with optical depths lower than 0.03).

Figure 3.3: Summary of the steps to build the synthetic dataset, from high spectral resolution radiances.

Table 3.3: Number of clear and cloudy simulated spectra for each latitude belt. In parenthesis the number of liquid water clouds (Liq), of mixed phase clouds (Mixed), of ice clouds (ice) and, among these, the number of sub visible cirrus (svc), consisting in high altitude cirri with optical depths less than 0.03, are shown.

	Clear sky	Clouds [Liq / Mixed / Ice (svc)]
Tropics	704	986 [16 / 61 / 909 (212)]
Mid Latitudes	615	765 [96 / 48 / 621 (173)]
Polar regions	492	532 [00 / 48 / 484 (132)]

3.2 Sensitivity tests on FIR

Sensitivity tests are performed by using simulated spectra. The tests regard to clouds and surface emissivity are made by using some spectra from the synthetic dataset previously described. These tests are presented in Sections 3.2.1 and 3.2.2, respectively. Additionally, some tests varying the cloud properties in specific FIR and MIR wavenumbers are made using simulated spectra computed by using the LBLDIS code.

The sensitivity is evaluated by comparing the spectra's Brightness Temperature (BT) with a reference spectrum. In most tests, the reference spectrum assumed is the clear sky spectrum for the same case that is being evaluated.

For example, in Figure 3.4, the BT sensitivity, with respect to a reference clear sky case, is shown for three different cases in a tropical atmosphere: a sub-visible cirrus cloud (black line), the clear sky case with a 3 K decrease in skin surface temperature (yellow line) and the clear sky case with an increase of 10% along the vertical profile of the water vapour mixing ratio (blue line). The cirrus is assumed to be composed of hexagonal plates with effective dimension of 20 µm, optical depth at 900 cm^{-1} OD=0.03, geometrical thickness 1 km and cloud top 14 km. Results show distinctive spectral features due to the presence of the cirrus cloud in the satellite view. The differences in terms of radiances are higher than the goal noise of the FORUM mission, both in wavenumbers in the FIR and in the MIR, either in the presence of a cirrus cloud or with the increase of the water vapour concentration in the vertical profiles. On the other hand, the radiance signal in the FIR exiting from the surface is masked by the
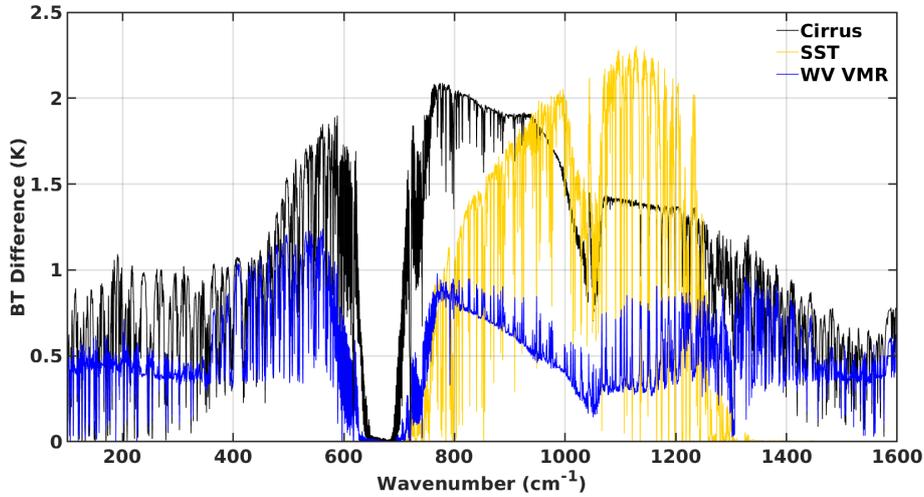
Figure 3.4: BT sensitivity to a sub-visible cirrus cloud (black line), to a 3 K decrease in surface temperature (yellow line) and to an increase of 10% in water vapour mixing ratio at all levels (blue line). The BT differences are obtained in reference to a Tropical clear sky case over the ocean.

strong absorption by the water vapour rotational band (which is almost saturated for wavenumbers below 300 cm^{-1}), in particular for the tropical region.

3.2.1 Cloud sensitivity tests

Sensitivity tests regard to the clouds are performed to evaluate the variations that can be observed either in the FIR and in the MIR, with changes in the cloud particle effective dimension, water phase, and habit.

In Figure 3.5 the large BT sensitivity to cloud particle size distribution effective dimensions is highlighted for channels in the MIR windows and also for wavenumbers between 400 and 600 cm^{-1}. Cloud properties are: OD=1.5, geometrical thickness 1 km, cloud top is at 14 km.

Sensitivity to cloud particle phase is shown in Figure 3.6. In the computations a Polar cloud (OD=7, geometrical thickness 1.5 km, cloud top at 5 km) made of spheres of pure ice, or liquid water, or two diverse mixed phases is assumed. The mixed phase are spheres with an ice core and liquid water coating. The coated part corresponds to 10% or 20% in radius of total sphere radius. The results show that the highest BT sensitivity to phase is found at FIR wavenumbers. Note also how the mixed phased spheres resemble the pure ice phase in the MIR window channel and the pure liquid water in the FIR channels.

In Figure 3.7, the sensitivity to crystal habit is shown. Four different habits (aggregates, hexagonal plates, solid bullet rosettes and solid columns) are assumed in the
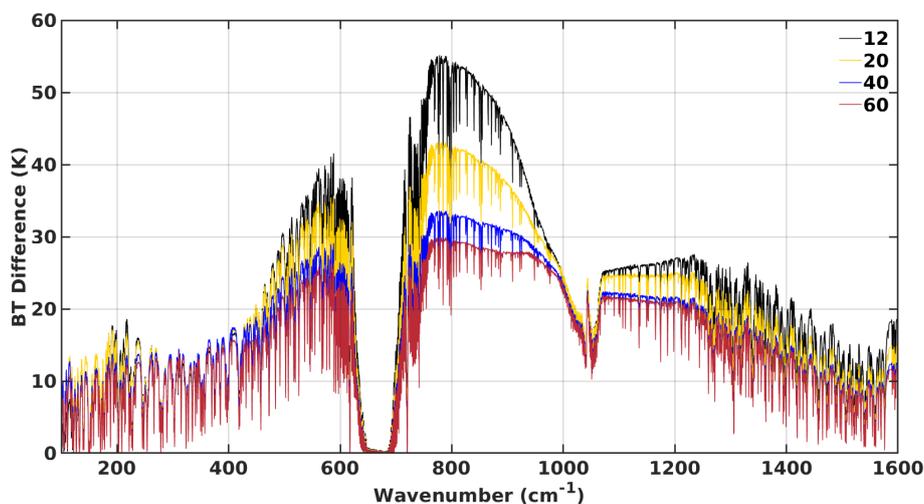
Figure 3.5: BT sensitivity to cirrus cloud particle size distribution effective dimensions. Dimensions values in the legend are in μm. The assumed shape is the hexagonal plate type, optical depth is 1.5, the cloud thickness is 1 km and cloud top is at 14 km. The BT differences are obtained in reference to a Tropical clear sky case over the ocean.
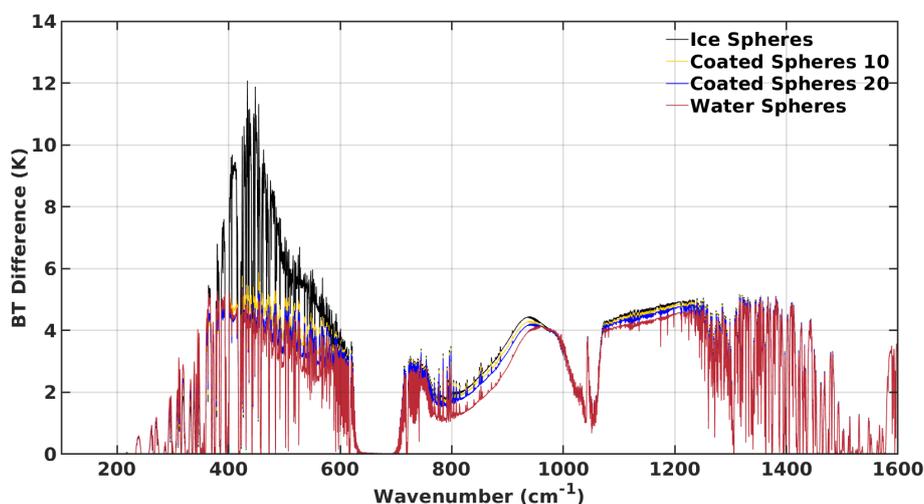
Figure 3.6: BT sensitivity to particles phase. A polar cloud is assumed made of a particle size distribution of spheres of ice (black line), liquid water (red) and two mixed phases. The mixed phase are spheres with an ice core and liquid water coating. The coating is respectively a 10% (yellow line) and a 20% (blue line) in radius of total sphere radius. Cloud optical depth is 7, cloud thickness is 1.5 km and cloud altitude is 5 km. The BT differences are obtained in reference to an Antarctic clear sky case over a snowed surface.
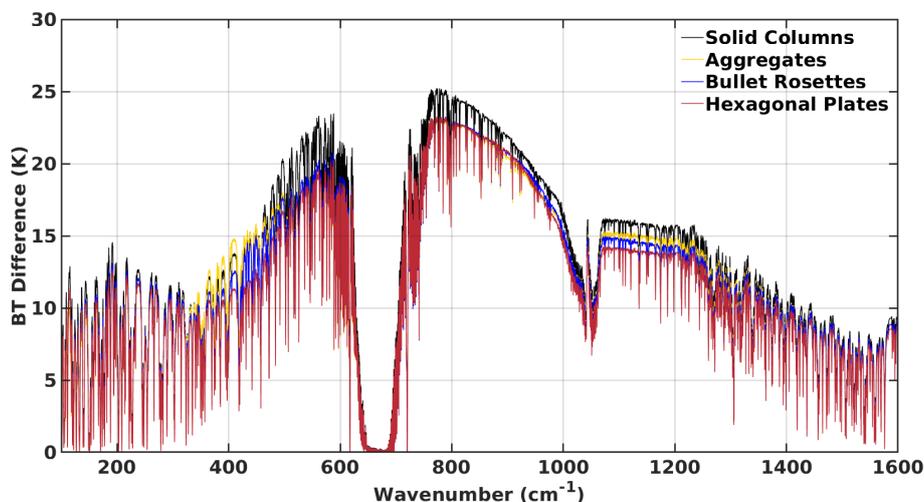
Figure 3.7: BT sensitivity to cirrus cloud crystal habit. The crystal's habits assumed are reported in the legend. The simulated tropical cirrus cloud is 2 km thick, with optical depth 1 and with effective dimension equal to 32 μm. The BT differences are obtained in reference to a Tropical clear sky case over the ocean.

simulation of 4 different tropical cirrus clouds with the same features (OD=1, geometrical thickness is 2 km, cloud top is at 15 km and effective dimension is 32 μm). Habit sensitivity is much larger in the FIR (about 4 K spread among the curves for different shapes) than in the MIR windows (about 2 K). This is mostly due to a minimum in the imaginary part of the ice refractive index at around 410 cm^{-1} that imply a minimum in absorption at FIR wavenumbers and a relative larger importance of scattering processes that are related to crystal shape.

This FIR largest sensitivity is noted to increase a little with the dryness of the atmosphere and thus amplified when moving towards higher latitudes. To show this little higher sensitivity, in Figure 3.8 the absolute BT differences between cirrus clouds in Tropical and Antarctic latitudes are presented. The cloud have the same OD (OD=1), geometrical thickness (2 km) and effective dimension (32 μm), however their top are different due to the tropospheric height, being 15 km in the tropical region and 8 km in Antarctic. The surface emissivity is also different, since the sea water is used for the tropical simulations, and fine snow is used for the Antarctic case. The BT sensitivity in this case is evaluated with respect to a cirrus cloud composed by hexagonal plates, in order to compare just the differences between the various shapes. It is possible to note a larger difference about the shapes around 410 cm^{-1} in Antarctic when compared to the Tropical case.

The lasts sensitivity tests concerning clouds are done varying the ice crystal habit, the particle effective dimension and the cloud optical depth. These analyses are made for the wavenumbers 410 and 900 cm^{-1}, to compare the effect of the cloud' properties
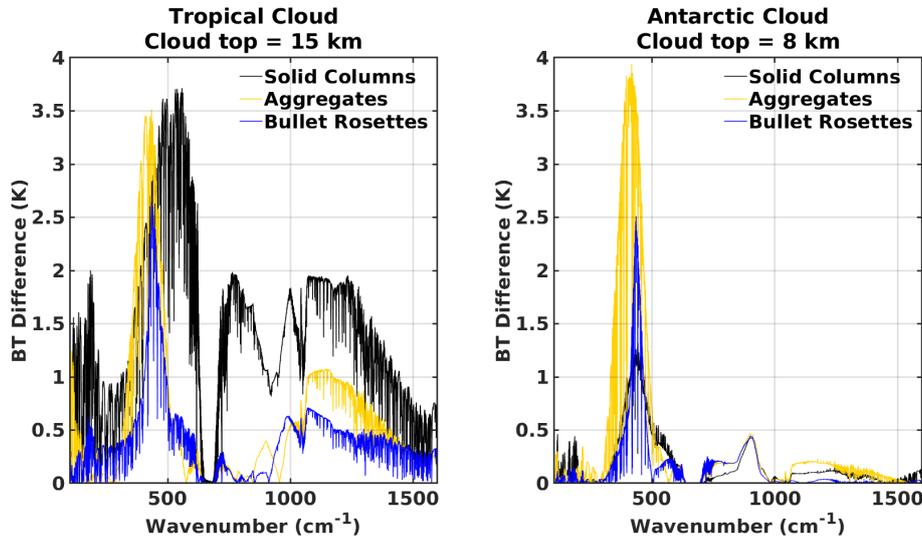
Figure 3.8: BT sensitivity to cirrus cloud crystal habit. The crystal's habits assumed are reported in the legend. The simulated cirrus cloud is 2 km thick, with optical depth 1 and with effective dimension equal to 32 µm. The cloud top is 15 km to the Tropical case and 8 km to the Antarctic. The absolute BT differences are obtained in reference to the same cirrus cloud composed by hexagonal plates. The Tropical case is over the ocean, and the Antarctic case is over a snowed surface.

changes in the FIR and MIR, respectively.

In this case, the simulations are performed by using the LBLDIS with a spectral resolution of 0.01 cm^{-1}. A cirrus cloud 1 km thick, with top at 14 km is assumed in all the computations. The particle effective dimension vary between 20 µm and 100 µm, and four different habits (aggregates, hexagonal plates, solid bullet rosettes and solid columns) are considered. The sensitivity tests are shown in Figure 3.9 for a cloud with optical depth 0.5 (a), 1.3 (b) and 5 (c).

The results show that maximum BT differences in the MIR part of the spectrum, considering the different particle shapes, range within 0.8 K for clouds with OD 0.5, 1.8 K for clouds with OD 1.3, and 2.1 K for clouds with OD 5. On the contrary, BT differences in the FIR span over more than 1 K for small crystals (less than 40 µm) in the thinner optical cloud, up to and 8 K for medium and large crystal sizes in the thicker optical cloud. This suggests that the analysis of combined observations in the FIR and MIR might be sufficient to allow an identification of the cloud properties signal in the spectrum.
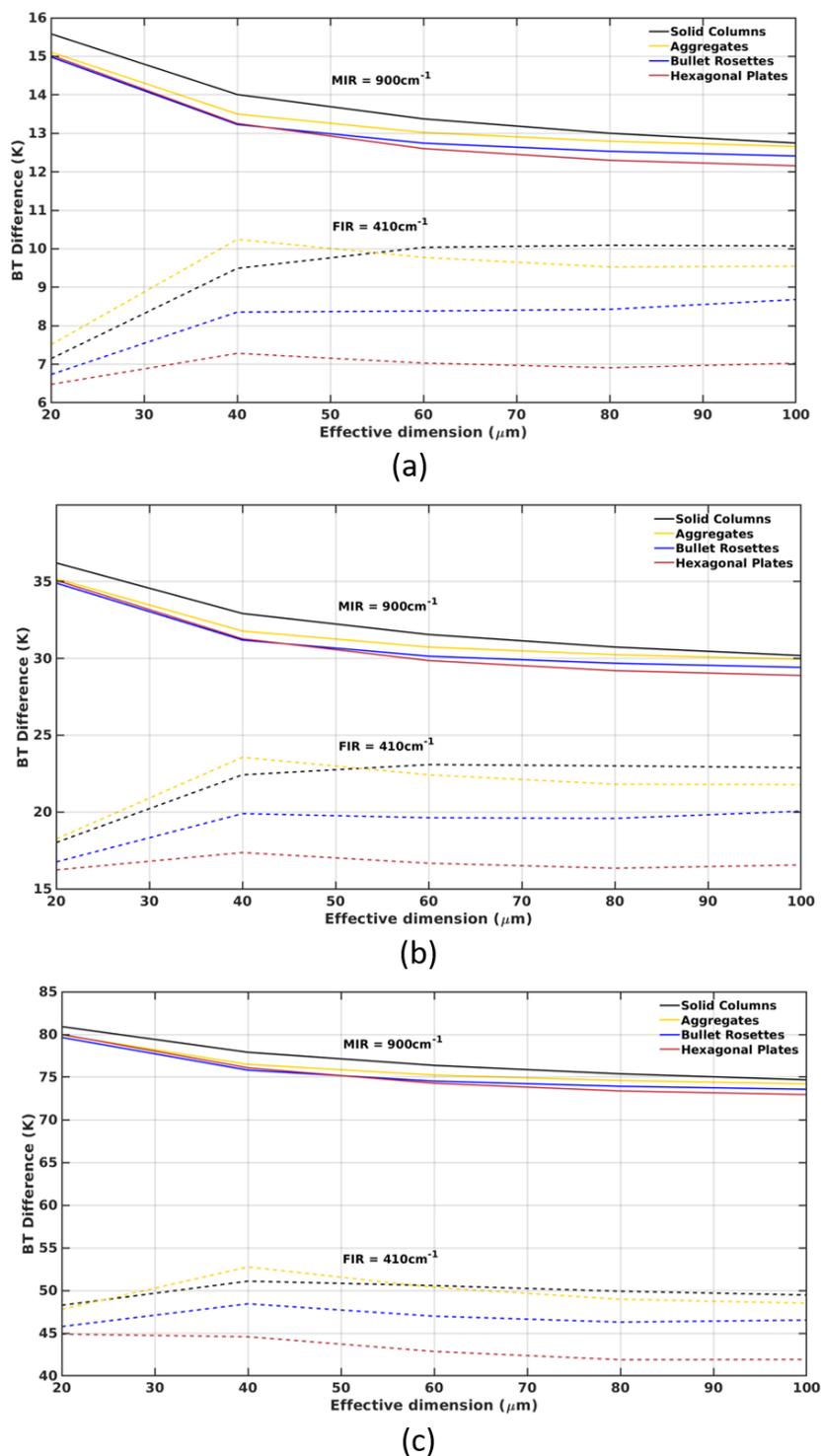
Figure 3.9: BT sensitivity in FIR and MIR wavenumbers to a cirrus cloud, varying the crystal habit and the particle effective dimensions. The crystal's habits assumed are reported in the legend. The simulated tropical cirrus cloud is 1 km thick, with at 14 km, and optical depth equal (a) 0.5, (b) 1.3, and (c) 5. The BT differences are obtained in reference to a Tropical clear sky case over the ocean.
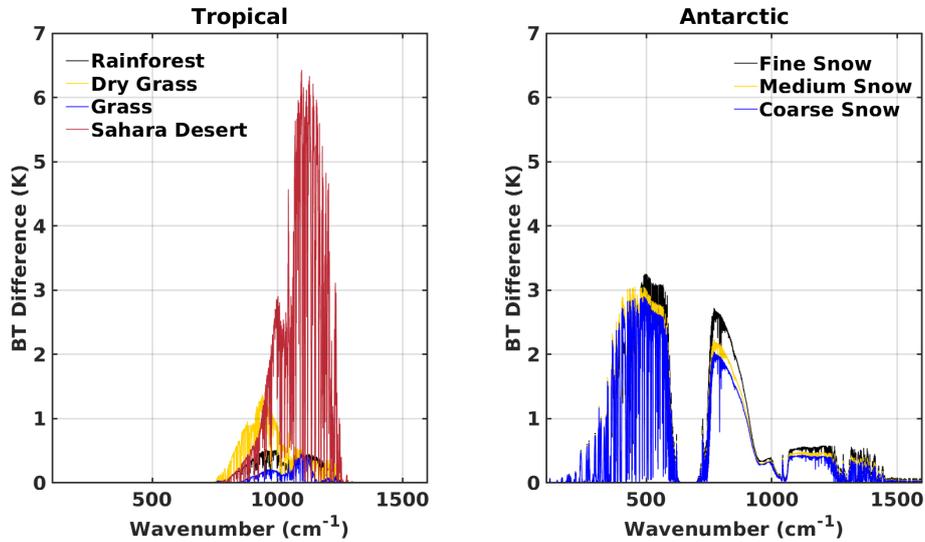
Figure 3.10: BT sensitivity to surface emissivity. The surfaces used are reported in the legend, and the simulations are made in clear sky conditions. The absolute BT differences are obtained in reference to the sea water surface in the tropical case, and in reference to an ice surface in the Antarctic case.

3.2.2   Surface emissivity sensitivity tests

Some surface emissivity sensitivity tests are also performed. These tests highlight the strong absorption by the water vapour rotational band. The absolute BT sensitivity is analysed in two clear sky cases, one for the tropical zone, which is a very moist zone, and one for a drier antarctic zone (Figure 3.10). For each latitudinal zone, only the surface emissivity changes. For the tropical case, the reference surface emissivity is the sea water, while for the Antarctic case the ice is the surface of reference. In the tropical case, the effect of changing surface emissivity is observed just in the MIR, whilst in the Antarctic case, which has less water vapour available, it is possible to see the surface emissivity effects also in the FIR. Therefore, the signal in FIR can be used to improve our knowledge about surface characteristics in some drier conditions. These results are complementary to that shown and discussed in Figure 3.4.

3.3   Summary

Initially, the methodology employed to produce a large dataset to simulate high spectral resolution radiance, like that can be observed in the FORUM mission, is described. The measured FORUM radiance will covers the 100-1600 cm^{-1} spectral bands (thus including the under-explored far infrared part of the spectrum) with a nominal spectral resolution of 0.3 cm^{-1} and a goal noise of 0.4 $mW/(m^2 sr cm^{-1})$ in the 200-800 cm^{-1} interval and 1

$mW/(m^2 sr cm^{-1})$ outside.

Simulations are performed by using multiple surface properties, atmospherics profiles and different cloud features for liquid, mixed phase and ice clouds (including multiple ice habits), and the sensitivity tests shows that distinctive spectral features are observed in the presence of the cirrus cloud.

Both MIR and FIR are sensible to the cloud thermodynamics, geometrical and optical properties. The sensitivity to cloud PSD effective dimensions is highlighted for channels in the MIR window (around 900 cm^{-1}) and for wavenumbers between 400 and 600 cm^{-1}. On the other hand, it is noted that the radiance sensitivity to the ice habit is larger in the FIR than in the MIR. Moreover, a higher radiance sensitivity to the water phase is found at FIR wavenumbers. The FIR radiance sensitivity to the cloud particle water phase and shape is relevant, mainly around 410 cm^{-1} where the imaginary part of the refractive index of ice has a local minimum, making scattering significant. Furthermore, the highest cold cloud emissivity is in the FIR part of the spectrum.

The combined use of spectrally resolved radiances in FIR and MIR allows the identification of multiple processes involving atmosphere, ground and clouds. Then, in the next chapter, the simulations described here are used to develop the CIC algorithm, an innovative tool able to identify and classify clouds, by using high spectral resolution radiances.

Chapter 4

Cloud Identification and Classification - CIC

Machine learning techniques based on principal component analysis, artificial neural networks, or support vector machines, among others, have been extensively applied to identify clouds from high resolution satellite imagers (Mahajan and Fataniya, 2019). Most of these techniques exploit data spanning from infrared to shortwave, which limits their applicability to daytime hours only, as the research developed by Ahmad and Quegan (2012), which uses a PCA based detection algorithm.

However, it is less common to find cloud detection algorithms based on hyperspectral infrared sounders. The existing ones are generally applied by selecting specific spectral channels and/or evaluating brightness temperature thresholds and brightness temperature differences (Serio et al., 2000; McNally and Watts, 2003; Lavanant et al., 2011; Luo et al., 2019; Wang et al., 2019), which can limit their applicability and require elaborate preliminary calibrations. Commonly, they are also employed in conjunction with high spatial resolution imagers (Lavanant et al., 2011; Eresmaa, 2014), which implies a relatively high computational cost due to the necessity to combine two different types of measurements.

Taking the cue from existing cloud detection algorithms, a Cloud Identification and Classification (CIC) method is developed. The CIC is a machine learning algorithm based on PCA, which performs an identification and classification using a single threshold applied to an univariate distribution of a newly defined parameter called Similarity Index (SI) (see Section 4.1.1), which determines the relatedness with a specific class (Training Set (TS)). The algorithm is computationally very fast and requires a limited number of spectra as training sets, which makes it very flexible, efficient, user-friendly and easy to adapt to different types of sensors.

Another CIC's advantage is that it is based only on the spectral analysis of the radiance signal and does not rely on any ancillary information or model output. Therefore, it can potentially be applied to any spectral sensor in the infrared on ground-based,

airborne, or satellite platforms, and it is able to work, with the same performances, on daytime or nighttime.

Next, the algorithm is presented and applied by using the synthetic dataset described in Chapter 3. The cloud detection is performed both using the MIR only or the full spectrum (FIR and MIR) so that the detection performances could allow an evaluation of far-infrared channels information content in realistic conditions (Section 4.2). The algorithm description and this cloud detection performance evaluation were presented in the manuscript developed by Maestri et al. (2019b).

4.1 Algorithm description

CIC is an innovative classification algorithm based on PCA. The methodology relies on a machine learning algorithm that requires the definition of a certain number of TS equal to the number of classes used for the classification. A descriptive example of the identification process of clear and cloudy cases (*cloud detection*) is first provided in order to facilitate the comprehension of the rigorous mathematical treatment that follows this brief introduction. In Figure 4.1 a flowchart of the algorithm is depicted.

The first step requires the definition of a *clear sky training set* ($\mathbf{TS}_{\mathrm{cle}}$), consisting of a number T_{cle} of clear sky spectra, and a *cloudy sky training set* ($\mathbf{TS}_{\mathrm{clo}}$), consisting of T_{clo} cloudy sky spectra. For each TS the Principal Components (PCs) are computed and stored in a matrix. Each spectrum of the test set (one at a time) is then added to the $\mathbf{TS}_{\mathrm{cle}}$ and thus an *extended clear sky training set* ($\mathbf{ETS}_{\mathrm{cle}}$) is defined. $\mathbf{ETS}_{\mathrm{cle}}$ is a group of $T_{cle} + 1$ spectra. Then, the PCs of the $\mathbf{ETS}_{\mathrm{cle}}$ are computed. Supposing that the test set spectrum under consideration is a clear sky spectrum it is expected that the PCs computed for $\mathbf{ETS}_{\mathrm{cle}}$ are very similar to the ones computed for $\mathbf{TS}_{\mathrm{cle}}$ that is to say that the test set element has the same basic features as the elements belonging to the TS (clear in the running example). In this case, it is also expected that the PCs computed for the cloudy sky TS $\mathbf{TS}_{\mathrm{clo}}$ differ from the PCs obtained from the extended cloudy sky training set ($\mathbf{ETS}_{\mathrm{clo}}$) that is obtained by adding the spectrum in consideration (that is clear) to the cloudy sky training set.

The CIC evaluates the variation of the PCs of the TS due to the addition of a new spectrum (from the test set). The association of a spectrum to a specific class is obtained by evaluating the similarity of PCs of the extended training sets to those of the original training sets: small changes in PCs are interpreted as that the spectrum belongs to the class, while large changes suggest that the spectrum belongs to a different class. The variations of the PCs obtained for the Extended Training Set (ETS) with respect to the original ones are evaluated by means of a new parameter called *similarity index*.

The notation for similarity indices is:

$$\mathrm{SI}(i,j), \qquad i \in \{1,2\}, \quad j \in \{1...J\} \tag{4.1}$$
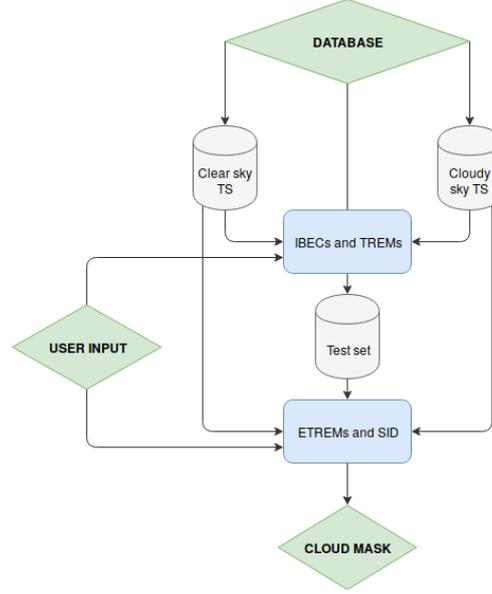
Figure 4.1: Scheme of the dataflow used for the cloud detection process.

where i is the class label, j is the test set spectrum label and J is the number of spectra in the test set to be classified. As an example (that is used in the whole text) it is assumed that the class label is 1 for clear sky spectra and 2 for cloudy sky spectra.

The computation of the similarity indices are now described mathematically. The first step is the definition of the *training set matrices*.

$$\mathbf{TS}_i(\nu, t) \tag{4.2}$$

$$i \in \{1, 2\}, \quad t \in \{1...T_i\}, \quad \nu \in \{1...\nu_{max}\}$$

where t is the spectrum index, T_i is the number of spectra in each training set i, ν is the wavenumber index that spans from 1 to ν_{max} that is the highest wavenumber index.

The second step consists in the computation of the PCs of each TS matrix by evaluating the eigenvectors (eig) of their covariance (cov) matrix:

$$\epsilon\mathbf{TS}_i(\nu, p) = \text{eig}(\text{cov}(\mathbf{TS}_i(\nu, t))) \tag{4.3}$$

$$i \in \{1, 2\}, \; t \in \{1...T_i\}, \nu \in \{1...\nu_{max}\}, \; p \in \{1...P\}$$

where $\epsilon T S_i$ is the *training set eigenvector matrix*, p indicates the pth principal component, and $P = \max(T_i, \nu_{max})$ is the total number of principal components. Each row of this matrix contains normalised eigenvectors:

$$\sum_{\nu=1}^{\nu_{max}} \epsilon \mathbf{TS}_i(\nu, p)^2 = 1 \tag{4.4}$$

Given J spectra in the test set, a number of J new matrices is defined for each class. These matrices are simply the concatenation of the *training set matrices* with each single spectrum (j) from the test set, and are called *extended training set matrices*. Let

$$\mathbf{TestSet}(\nu, j), \quad 1 < j < J$$

be the matrix containing all the spectra from the test set. The *extended training set matrices* are defined as follows:

$$\mathbf{ETS}_{i,j}(\nu, t') = [\mathbf{TS}_i(\nu, t) \| \text{row}_j(\mathbf{TestSet}(\nu, j))] \tag{4.5}$$

$$t' \in \{1...T_i + 1\}$$

where the notation $\|$ indicates matrix concatenation. Note that $\text{row}_j(\mathbf{TestSet}(\nu, j))$ is a single test set spectrum in a one-dimensional array.

CIC evaluates the principal components of the *extended training set* $\epsilon ETS_{i,j}$ as follows:

$$\epsilon \mathbf{ETS}_{i,j}(\nu, p) = \text{eig}(\text{cov}(\mathbf{ETS}_{i,j}(\nu, t))) \tag{4.6}$$

The Training Set Eigenvector Matrix (ϵTS) and Extended Training Set Eigenvector Matrix (ϵETS) are used to compute the similarity indices (SI) for each test set spectrum (j) and for each class (i):

$$\text{SI}(i, j) =$$
$$1 - \frac{1}{2P_o} \sum_{p=1}^{P_o} \sum_{\nu=1}^{\nu_{tot}} |\epsilon \mathbf{ETS}_{i,j}(\nu, p)^2 - \epsilon \mathbf{TS}_{i,j}(\nu, p)^2| \tag{4.7}$$

where ν_{tot} is the number of features (channels) used for PCA and P_o is the number of principal components that are associated to the physical signal (real variability) characterizing the spectrum.

The set of optimal principal components (P_o) characterizing the signal constitutes the Information-BEaring principal Components (IBECs). The P_o elements are extracted

by minimising the factor Indicator function (IND) defined by Malinowski (2002) and Turner et al. (2006):

$$\text{IND}(p) = \frac{\text{RE}(p)}{(P-p)^2} \tag{4.8}$$

where $\text{RE}(p)$ is defined, in Turner et al. (2006), as the *real error*

$$\text{RE}(p) = \sqrt{\frac{\sum_{i=p+1}^{P} \lambda_i}{T_i(P-p)}} \tag{4.9}$$

where λ_i is the ith eigenvalue of the covariance of some data matrix and T_i is the number of spectra in the training set i.

The natural number P_0, obtained through this minimisation process, is the number of eigenvectors associated with the physical signal corresponding to the number of IBECs. In CIC, P_o is computed separately for both TS matrices ($i = 1, 2$).

Once P_o is determined the similarity index can be calculated using Equation 4.7. For consistency, the same value of P_o is used when the SI computation is applied to the two training sets; the minimum value for P_o is utilized.

Interpreting the eigenvectors as directions in the multi-dimensional space, the SI is an estimate of how much the PCs of the TS *rotate* after a new spectrum is added to the set. For this reason, similarity indices do not depend on *eigenvalues* but on *eigenvectors*: all the PCs describing the physical signal are accounted for with the same weight in Equation 4.7, since all of them might be important for classification.

Similarity indices defined in this way are normalised. In fact, since the absolute value of the difference between the square loadings of two eigenvectors is at most equal to 2, the sum of P_o differences can reach the maximum value of $2P_o$. And being an absolute value, it turns out that:

$$0 \leq \sum_{p=1}^{P_o} \sum_{\nu=1}^{\nu_{tot}} |\epsilon\mathbf{ETS}_{i,j}(\nu,p)^2 - \epsilon\mathbf{TS}_i(\nu,p)^2| \leq 2P_o \tag{4.10}$$

and therefore:

$$0 \leq \text{SI}(i,j) \leq 1 \tag{4.11}$$

With reference to Equation 4.7, the largest value of the similarity index (1) is obtained for identical ϵTS and ϵETS matrices meaning that the analysed test set spectrum is not adding any diverse physical information to the training set spectra. An SI close to zero means that the two matrices are described by very different PCA and the test set element is bringing additional information to the original training set.

A graphical example of the similarity index is provided in Figure 4.2. The plot shows the SI computed for cloudy elements of the tropical test set only in the left hand side
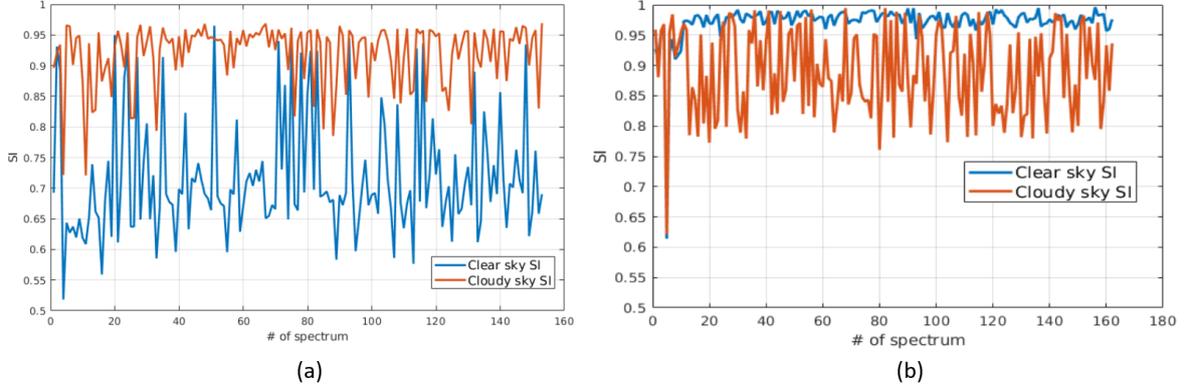
Figure 4.2: SI computed using (a) tropical cloudy spectra only or (b) clear sky spectra only. In orange the SI is computed using the cloudy sky training set and in blue the clear sky training set. The tropical case is accounted for. The full spectrum is used in the SI computations. Details on the training sets are provided below in the text.

panel and for clear sky elements only in the right hand side panel. For cloudy sky test set cases, when the SI is computed using the cloudy TS (orange line) the SI is very close to one while the SI values are mostly lower when the clear sky training set is used. Thus, the inequality $\mathrm{SI}(clo, j) > \mathrm{SI}(cle, j)$ holds for most cloudy sky spectra j. The situation is reversed when clear sky elements of the test set are used (blue line), showing that highest SI values are generally obtained when using the clear sky training set. CIC exploits results from both the comparisons: SI computed using the clear sky and the cloudy sky training sets.

4.1.1   Classification

Elementary approach

The CIC classification requires that each test set element is used for the computation of both similarity indices (one for each training set of the two classes). Once the SIs for the ETS (the one containing cloudy spectra and the one for clear ones) are computed a comparison is performed. Continuing with the running example, when

$$\mathrm{SI}(\mathrm{cle}, j) > \mathrm{SI}(\mathrm{clo}, j),$$

the spectrum j is expected to be clear sky (cle). And of course for:

$$\mathrm{SI}(\mathrm{clo}, j) > \mathrm{SI}(\mathrm{cle}, j)$$

the spectrum j is expected to be cloudy (clo).

These two conditions may be unified in a compact index that is defined as the Similarity Index Difference (SID):

$$SID(j) = SI(clo, j) - SI(cle, j) \qquad (4.12)$$

and thus if

$$\text{if } SID(j) > 0, \quad \text{then } j \in \{\text{cloudy spectra}\} \qquad (4.13)$$
$$\text{if } SID(j) < 0, \quad \text{then } j \in \{\text{clear spectra}\} \qquad (4.14)$$

Spectra classification based only on the SID sign is defined as the *elementary approach*: SID acts as a binary classification parameter. Each spectrum is analysed sequentially and independently from the other elements of the test set under consideration. This elementary approach has the main advantage of being very simple and straightforward, and the disadvantage of being sensitive to the composition of the training sets. In fact, results might be affected (and biased) if one of the training sets is not well populated by spectra that are representatives of the variability within the class.

An example of the elementary classification is given in Figure 4.3 where the SID distribution for clear and cloudy sky spectra is shown. In this example, the cloudy sky and clear sky training sets are not well characterised (some unbalance is observed) since even if all the cloudy spectra show positive SIDs (as expected), a large number of clear sky spectra has also positive SIDs and thus, those spectra, are potentially misclassified. It is therefore shown that the elementary approach requires an accurate definition of the training sets to work properly.

The selection of the spectra contained in each training set is thus crucial (as in every classification algorithm). In fact, the selected elements need to provide the main features that characterize the class and also its variability in order to correctly identify and classify the most of the analized spectra.

Distributional approach

Better results can be obtained when an optimization of the classificator is performed a-priori by using a methodology called *distributional approach*.

When applied to a set of spectra, a perfect classifier would ideally generate a bimodal SID distribution. In this case, if the elements of each class are correctly classified the bimodal distribution of the SID is made of two separate groups. This is not happening in reality and the overlap depends on many factors among which the spectra used to define the TSs.

At this purpose the CIC is applied separately to each one of the elements of the TSs. These elements are spectra of known class and thus allow to evaluate the SID distribution
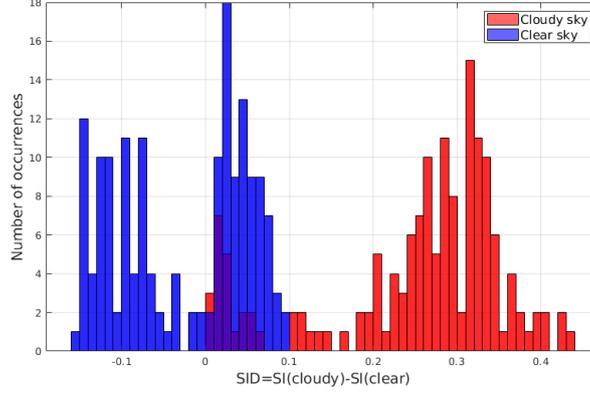
Figure 4.3: SID occurrence distribution for a subset of tropical spectra. The SIDs are computed based on specific (not optimised) training sets. Test set clear sky spectra are in blue, while cloudy sky spectra are in red. Cirrus clouds show very small, positive values of the SID, comparable with values obtained for some clear sky spectra. Class membership of test set spectra is known a-priori since the dataset is synthetic.

for all the elements. Analyzing the SID distributions, an optimal SID delimiter can be defined so that the classification of the TSs elements is maximized.

Once that this delimiter is defined, the classification can be applied by using a new index called *Corrected Similarity Index Difference (CSID)*, that is simply a shifted SID:

$$\text{CSID}(j) = \text{SID}(j) - shift_{opt} \tag{4.15}$$

where $shift_{opt}$ is the delimiter defined by an optimal value of the parameter *shift* that maximises a function that can potentially forecast the performance of the algorithm (a *performance forecasting function*, PFF). This function is called Consistency Index (CoI) and is given by:

$$\text{CoI}(\text{shift}) = 1 - \max(\frac{\text{FP}_{\text{cle}}}{T_{cle}}, \frac{\text{FP}_{\text{clo}}}{T_{clo}}) \tag{4.16}$$

where the clear and cloudy false positives (FP) are counted as:

$$\text{FP}_{\text{cle}}(\text{shift}) = \text{card}(\{t \in T_{cle} \mid \text{SID}(j) - \text{shift} > 0\}) \tag{4.17}$$

$$\text{FP}_{\text{clo}}(\text{shift}) = \text{card}(\{t \in T_{clo} \mid \text{SID}(j) - \text{shift} < 0\}) \tag{4.18}$$

In the above formula, the «card» operator denotes set cardinality, while T_{cle} and T_{clo} are the number of clear and cloudy TS elements respectively. The consistency index

measures the representativeness of the training sets and in particular it computes how many TS elements would be classified correctly if they were part of the test set. The consistency index is large (close to 1) only if both clear and cloudy false positives (FP in the equations above) are rare, that is when the training sets are composed of elements that accurately represent the variability within the specific class.

The new CSID parameter operates as the SID parameter: if it is positive, the spectrum is considered cloudy, and if its negative, the spectrum is considered clear. In formulas:

$$\text{if CSID}(j) > 0, \quad \text{then } j \in \{\text{cloudy spectra}\} \tag{4.19}$$

$$\text{if CSID}(j) < 0, \quad \text{then } j \in \{\text{clear spectra}\} \tag{4.20}$$

The use of the distributional approach significantly improves the performance of the algorithm as it will be shown in the results (Sections 4.2, and Chapters 6 and 7).

Unclassified spectra

Each element of the test set is classified in accordance with the sign of the classifier (SID or CSID). For practical purposes, it can be useful to define some thresholds or limits that determine a set of «unclassified» elements characterised by values of the classifier belonging to a limited interval around 0 (that is the ideal point separating the two modes of the distribution). The simplest way to set this interval is to let the user to define two parameters: Θ_1 (positive) and Θ_2 (negative), representing the inner limits of a confidence interval. Any spectrum whose classifier is such that falls within the interval $[\Theta_2, \Theta_1]$ is considered «unclassified». The Θ_2 and Θ_1 parameters should be defined in accordance with the experimental conditions and sensor performances and their quantification is beyond the scope of this work. For this reason and since we rely on a synthetic test set all the classifications performed will be binary that is each element is classified either clear or cloudy.

Additionally, if the CIC is used not only as a cloud mask (i.e. to define if a spectrum is derived from a clear or cloudy scene) but also to classify a scene according to the type of cloud, the classifier is kept binary, however, it compares each pair of the classification according to the TS available. For example, in the FE2ES (see Chapter 5), the CIC is used to classify a cloudy scene according with type cloud: liquid water, ice, or cirrus cloud (ice clouds with optical depth less than 2). Then, the CIC compares an input spectrum with two classes (TS) at a time, to establish which of the two is more similar to the input. The algorithm executes six comparisons: 1) clear sky and cirrus clouds; 2) clear sky and water clouds; 3) clear sky and ice clouds; 4) cirrus clouds and water clouds; 5) cirrus clouds and ice clouds; 6) water clouds and ice clouds. If there is a prevalent class between all the comparisons, the scene is classified as that class, however, if there is no prevalent class, the scene is unclassified. In Figure 4.4 is provided an example of
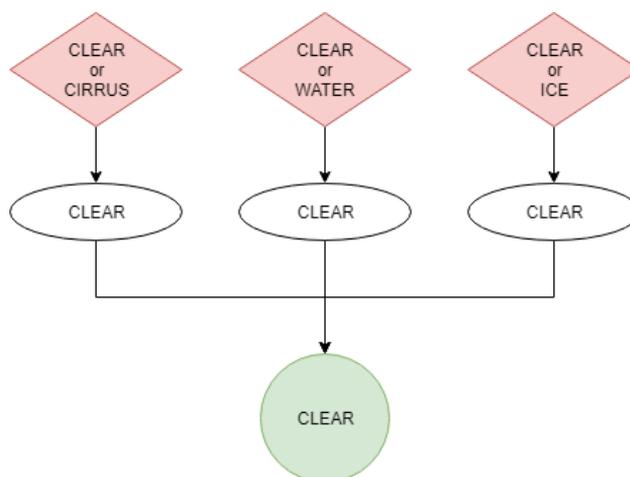
Figure 4.4: Scheme of a scene classification, from CIC used in the FE2ES.

the classification used in the FE2ES. In this case the SI of the clear sky is always higher than other classes. More details of the CIC algorithm used in the FE2ES is presented in Section 5.2.

4.2   CIC - First Results

4.2.1   Scores

There is no unique metric to define the performances of a cloud classification algorithm since their assessment is linked to the goal of the study. For this reason, the set of parameters (i.e. scores) used to evaluate the algorithm performance in this section are somewhat arbitrary. As a general rule, any metric should measure a better performance if more spectra are classified correctly.

In this regard, an element of the test set that undergoes a classification process falls into one of the following categories:

1. *True Positive. TP*: the element is a class i member, and is correctly classified.

2. *False Positive, FP*: the element is not a class i member, but it is classified as belonging to class i.

3. *False Negative, FN*: the element is a class i member, but is not correctly classified as such.

4. *True Negative, TN*: the element is not a class i member, and it is not classified as belonging to class i.

	Actual Class	
	Class X	**No Class X**
Class X	TP True Positive	FP False Positive
No Class X	FN False Negative	TN True Negative

Figure 4.5: Confusion Matrix (or contingency table), indicating the categories of TP, TN, FP, and FN.

The quality of the classification can be defined using two or more of these sets, through scalar attributes widely applied from the called *confusion matrix* or *contingency table* (Wilks, 2011), shown in Figure 4.5.

Between the various attributes that can be used, the first results of the CIC performance are evaluated by using the *precision* of the classification (also known as the *positive predictive value*):

$$\text{precision}(i) = \frac{\text{TP}(i)}{\text{TP}(i) + \text{FP}(i)} \tag{4.21}$$

where $\text{TP}(i)$ is the number of class i true positives and $\text{FP}(i)$ is the number of class i false positives. This score ranges between 0 and 1, and it is maximised for $\text{FP}(i) = 0$, occurring when all the spectra classified by the algorithm as class i are actually belonging to class i. The maximum *precision* imply, also, in a minimum *false alarm* of the algorithm classification.

Then, a detection performance (DP) is defined as the minimum value out of the two *precision* scores relative to the classes under examination:

$$\text{DP} = \min(\text{precision}(\text{class 1}), \text{precision}(\text{class 2})) \tag{4.22}$$

where:

$$0 \leq \text{DP} \leq 1 \tag{4.23}$$

In the definition of the detection performance the minimum *precision* value is chosen because the classification performance is considered high only if this scalar attribute is high for both classes.

4.2.2   First Results

In this section the performance of the CIC algorithm is evaluated for multiple atmospheric conditions. The additional information content of the FIR part of the spectrum
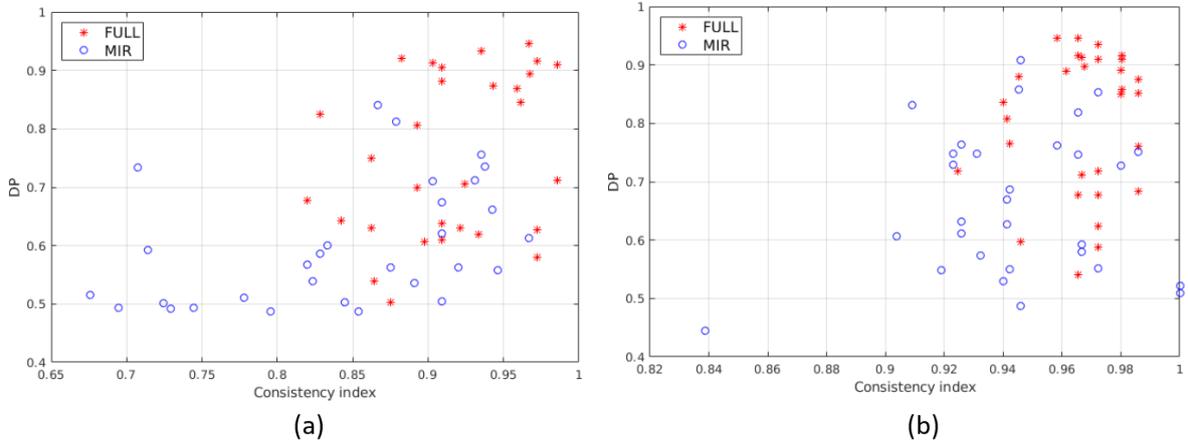
Figure 4.6: Scatter plots relating the CoI and DP for multiple tropical training set couples (clear and cloudy). CIC results are shown for both (a) the elementary and (b) the distributional approach. Blue circles represent classification results obtained using features in the MIR only, while red stars represent classification results exploiting the full spectrum.

is also studied. Classifications are performed both by using the MIR channels only or by using channels spanning over the full FORUM spectrum (FIR+MIR).

The first test is performed for the tropical case by using the 1690 spectra computed for this region (see Table 3.3). The CIC is run using 129 channel in the MIR only, and other 129 channels in the FIR, to consider the full spectrum. The selected channels range in the $[371.1 - 1300]$ cm^{-1} interval, with a fixed sampling except for a small interval at around the 667 cm^{-1} ν_2 vibrational CO_2 band, that is not used.

For this test, 60 couples of training sets (clear and cloudy), with 100 elements/spectra in total for each couple, are used to perform the classification to the simulations for the tropical case. For each one of the 60 training sets, the CIC is applied, and the CoI and the Detection Performance (DP) are computed and plotted as scatter plots(Figure 4.6). This allows to relate the final classification scores (evaluated by DP) with the composition of the training sets (whose characterisation is associated to the CoI).

The results show that the scores (indicated by DP) are generally larger when the full spectrum is exploited. In fact, it is computed that the average values of the DP move from 0.60 to 0.79 for the elementary approach (Figure 4.6a) and from 0.67 to 0.86 for the distributional approach (Figure 4.6b) when using features from the full FORUM spectra instead that those from the MIR only.

These preliminary computations suggest the following:

1. better results are obtained for the distributional approach (right panel) with respect to the elementary one (left panel)

2. a correlation between the CoI and the DP exists.

The points above are taken into account in order to maximise the performances of the CIC algorithm when applied to a test set. In particular, it is suggested that the distributional approach is preferred over the elementary one and that an optimal training set can be arranged before the classification is actually performed since CoI is computed for training set elements.

Thus, a strategy for the application of CIC to the synthetic dataset is defined and the evaluation of the information content in the FIR is planned:

- A Reference Training Set (RETS) is defined using an optimisation methodology based on the CoI values. The RETS is used as a reference to perform test set classification with a variable number of features.

- The performance of the algorithm is studied when using features from the MIR only or when exploiting also an increasing number of FIR channels up to cover the full FORUM spectrum.

REference Training Set (RETS)

In this paragraph, a strategy for the definition of an optimised training set, $RETS$, (intended as a set of clear and a set of cloudy spectra) is outlined. The optimisation applies when the training set elements are randomly chosen from a large subset of the whole dataset. Results could be different if training set elements are manually selected in order to cover the natural variability of the cloud and clear sky spectra encountered for each latitudinal belt and season.

The strategy is based on the correlation between CoI and the CIC DP. As previously mentioned, the CoI measures how well the training set elements would be classified if they were part of the test set.

For simplicity, only the CIC applications to the tropical data set are reported, but the same process is applied to Mid Latitudes and Polar cases. From the Tropical dataset, 315 spectra were randomly selected as test set, and other 854 are exploited to define multiple training sets: 100 elements at a time for each training set couple (clear and cloudy). This strategy is followed in order to have multiple training sets with different combinations of spectra. Of course the optimisation process is particularly important when the spectra that need to be classified are many. The operation is also important in reducing computational cost (see Section 4.2.3). In fact, the ϵETS matrices and the SID distribution of the optimised training set can be saved in a file, reducing the running time of both the elementary and the distributional approach.

The training set elements (100 per each training set couple) are randomly chosen from a set of 854 simulations in clear and cloudy sky, but with a constraint on the number

of clear and cloudy components. Three different Training set Number Configurations (TraNCs) are used:

$$T_{cle}(1) = 70, \ T_{clo}(1) = 30 \tag{4.24}$$

$$T_{cle}(2) = 50, \ T_{clo}(2) = 50 \tag{4.25}$$

$$T_{cle}(3) = 30, \ T_{clo}(3) = 70 \tag{4.26}$$

Twenty different training sets are constructed for each TraNC. Both an elementary and a distributional-approach-based classification are provided for each training set.

CIC is run on the test set elements and a binary classification (each element is classified as clear or cloudy) is performed by exploiting the full spectrum. The DP scores are computed. The results are presented in Figure 4.7, where the DP is plotted as a function of the CoI.

On average, the most accurate performances are obtained for TraNCs in which T_{cle} is larger than T_{clo}. The worst results are obtained for T_{clo} greater than T_{cle}. The average DP values for the three groups of TraNCs are reported in the plots of the figure.

A correlation between the DP and CoI is observed in Figure 4.7, where, on average, large values of DP are obtained for large values of CoI. This figure also shows that, on average, the DP computed using the distributional approach is larger than for results obtained using an elementary approach.

This correlation is significant even when CIC computes these two parameters totally independently, i.e., when the elementary approach is followed. For this reason the CoI can be used as a *performance forecasting parameter*, that is, a parameter estimating the quality of the classification. Nevertheless a quantification of the performances cannot be provided a-priori of the application of the algorithm since it also depends on the spectra of the test set to be analysed.

If a positive correlation between the DP and the CoI is assumed to exist, then the best performances are expected for training sets with the highest CoI's. For this reason, the RETS is defined as the training set with the highest CoI among those considered. In our case, the best performing training set is composed of 70 clear sky and 30 cloudy sky spectra.

The above configuration is selected as the tropical RETS to be used in the analysis shown in the next paragraphs. Similarly RETS for the mid and the polar latitudes are constructed. The RETS for clear sky consists of a set of spectra able to catch the seasonal (and thermal) variability reported in the synthetic dataset which is created to reproduce global conditions for all the four seasons.

The cloud ODs accounted for in the cloudy RETS for Tropics range from 0.05 to 21.8, while the test set ones range from 0.02 to 23; the RETS's ice cloud top heights range from 9 to 16 km and in the test set they span from 4 to 16 km. A summary of the clouds OD used in the RETS, for the Tropical case, is reported in Table 4.1. The Table shows that a large number of elements consists of optically thin clouds. This choice was
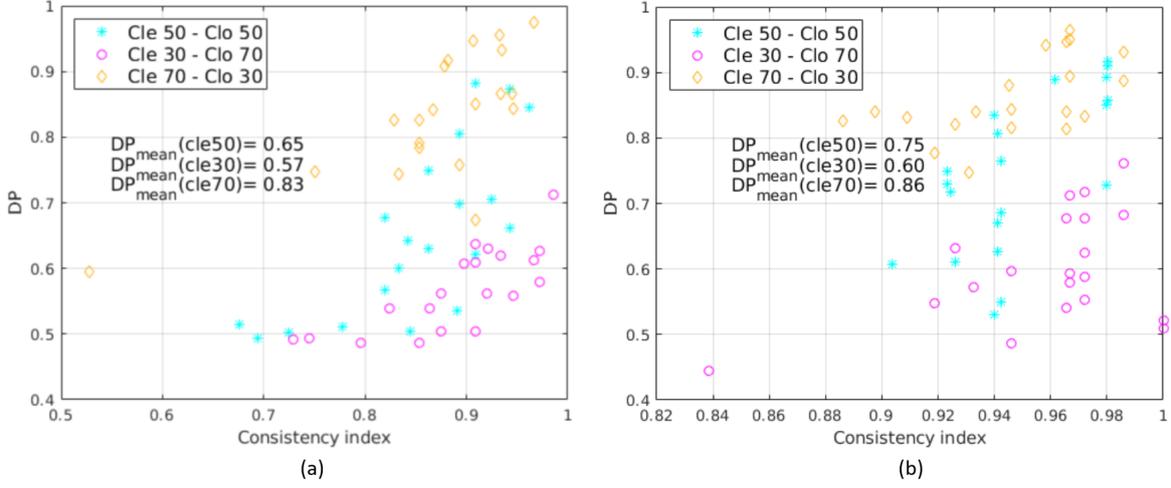
Figure 4.7: Scatter plot of DP and CoI for multiple training sets (grouped in 3 TraNC reported in the legend). Results are shown when using both (a) the elementary approach and (b) the distributional approach. For each group an average value of the DP is computed and reported in the plot.

performed in order to challenge the CIC capabilities to detect clouds in very difficult conditions.

Evaluation of FIR contribution to cloud identification

Multiple classifications, using a variable number of features (i.e., FORUM channels), and accounting both for the full spectrum and the MIR only are performed. The classifications account for a fixed number of MIR channels, while the number of FIR channels is changing to assess if the FIR part of the spectrum is capable to bring additional information content that significantly improves the algorithm's performances.

In order to speed up calculation and to avoid channels with a low Signal to Noise Ratio (SNR), the chosen MIR wavenumbers only range from 667 to 1300 cm^{-1}, while the FIR ranges from 100 to 640 cm^{-1}. Thus the full spectrum spans over the 100-1300 cm^{-1} spectral range with the exception of a limited wavenumber interval in the CO_2 band.

Channels are selected by using a constant sampling with no optimisation criterium applied. The number of selected channels in the FIR is defined by the following formula:

$$N_{feat} = \text{floor}(8 \cdot 2^{\frac{n}{2}}) + 1, \quad n \in 1...10 \tag{4.27}$$

In this way, N_{feat} spans over two orders of magnitude.

In Table 4.2 the number of features used in the FIR is reported. The upper (starting) channel in the FIR is at 639.9 cm^{-1} and the other FIR channels are sampled toward

Table 4.1: Number of cloudy spectra, as a function of cloud OD interval, for elements belonging to the cloudy set of the RETS and to the test set, Tropical case. In the second column, in parenthesis, the percentage value with respect to the total number of elements in the test set for the same OD interval is presented.

Training and test set cloud ODs		
Range	**RETS**	**Test set**
OD ≤ 0.1	8 (6.3%)	126
0.1 < OD ≤ 0.5	3 (2.3%)	131
0.5 < OD ≤ 1	1 (5.3%)	19
1 < OD ≤ 3	14 (4.5%)	308
3 < OD ≤ 10	0 (0%)	21
OD > 10	4 (6.8%)	59
Total	30 (4.5%)	664

smaller wavenumbers every 2.1 cm^{-1}. Thus the data reported in the table should be interpreted as follows: 12 channels means that 12 channels between 639.9 and 616.8 cm^{-1} are accounted for, and so on with the other larger number of channels up to cover the full FIR part of the spectrum.

In Figure 4.8, the results obtained for 11 different classifications are shown in terms of detection performance. DP is plotted as a function of the number of FIR features used in the classification applied to the tropical case. At the value 0, of the x-axis of the figure, the MIR part of the spectrum only is accounted for (256 channels in this case), while in the other 10 cases the FIR part of the spectrum also is exploited with an increasing number of channels indicated by the x-axis values.

Results show that performance gradually improves for increasing number of FIR channels. In particular, there is a slight decrease in performance after adding the channels closest to the CO_2 ν_2 band centre, which is gradually offset by improvements obtained when channels in the $[238.8 - 545.4]$ cm^{-1} wavenumber range are added. The DP slightly decreases if channels in the $[102.3 - 238.8]$ cm^{-1} range are included probably due to a reduced radiance sensitivity to surface and cloud features at those wavenumbers.

In the classifications, both the *distributional approach* (black line in Figure 4.8) and the *elementary approach* (red line) are followed. For the Tropical case, the elementary and distributional approach provide DPs as high as 0.9. Note that both methodologies take advantage of the optimised selection of the RETS and that the information content critical for DP improvements derives from channels spanning the $[238.8 - 545.4]$ cm^{-1} range.

In Figure 4.9 the cloud detection of the Tropical test set is plotted to function of the CSID value and cloud optical depth. Two classifications are performed: one using the MIR only (Figure 4.9a) and another using the full spectrum (Figure 4.9b). The number of MIR channels used in the process is kept the same (128) for both configurations. The

Table 4.2: Lowest wavenumber as a function of the number of FIR channels. The sampling rate is constant and set equal to 2.1 cm^{-1}. The highest FIR wavenumber is equal to 639.9 cm^{-1}.

number of FIR channels	Lowest channel (cm^{-1})
12	616.8
17	606.3
23	593.7
33	572.7
46	545.4
65	505.5
91	450.9
129	371.1
182	259.8
257	102.3
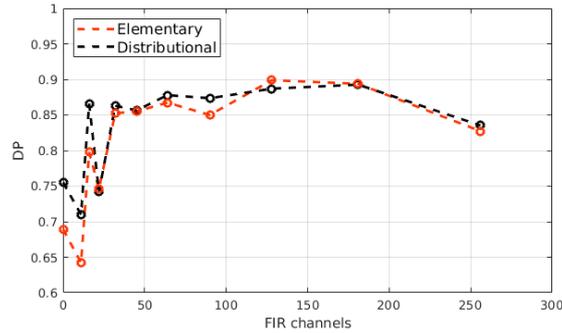
Figure 4.8: CIC cloud detection performance as a function of the number of FIR features (channels) for the Tropical case. 256 MIR channels are used. The black and red dots indicate the distributional and elementary approach respectively.

considered FIR channels, when using the full spectrum, are 129 and span from 371.1 to 639.9 cm^{-1}. For visual purposes the plotted clear sky test set elements are associated to OD=10^{-8}, while all the other OD values are for cloudy sky. Results shown in Figure 4.9 are obtained using the RETS.

True positives (correct classification) for clear and cloudy spectra are respectively orange circles and red asterisks. False positives are blue circle for clear and green asterisks for clouds. The left plot of the figure shows that the number of misclassified cloudy spectra grows for decreasing optical depth when using the MIR only. Clear sky case are all well classified in this configuration. If CIC is run exploiting the full spectrum (right panel of the figure) the overall detection performance is enhanced even if a clear sky case is misclassified (green asterisk). Nevertheless, most cirrus clouds are now correctly
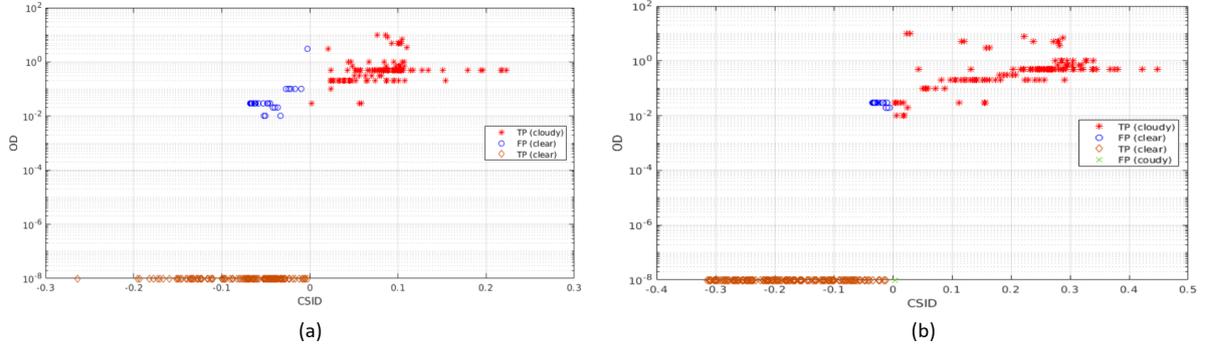
Figure 4.9: Classification of the Tropical test set as a function of the optical depth of the elements (a value of 10^{-8} is used to plot the clear sky cases). Data are plotted as a function of the CSID and cloud optical depth (OD). Left panel: CIC is run using 128 MIR channels. Right panel: CIC is run using the 128 MIR and 129 FIR channels. For both cases the *distributional* approach is assumed. Color codes are reported in the legend.

classified, with the exception of only few cases with optical depth less than 0.5. Note that the DP value is the minimum between the hit rate computed for cloudy sky cases and clear sky cases.

Mid Latitudes and Polar regions

The CIC code is applied to the Mid Latitudes and Polar datasets to test the algorithm performances in different atmospheric conditions. The results obtained using the MIR only and using the full spectrum are again compared. Since detection performances are dependent on the analysed datasets (test sets) the results cannot be interpreted in absolute sense, but in reference to the configuration parameters.

Mid-latitude Winter (MIDWIN), Polar Winter (POLWIN) and Polar Summer (POLSUM) cases are presented. The TraNCs used are generated with the same methodology outlined for the Tropical case thus they are composed of 70 clear sky spectra and 30 cloudy sky spectra randomly chosen from a large subset of the full dataset. Spectra used to define the training sets are not inserted in the test set. In Table 4.3 a summary of the total number of clear and cloudy spectra for each case study is provided. In the table the number of clear and cloudy simulations used for defining the training sets and for the test sets is also reported.

The capability to extract information content from the FIR is evaluated by applying the same procedure as before.

Classification results obtained using 60 different training sets are shown in Figure 4.10 for the 3 considered cases (MIDWIN, POLWIN and POLSUM respectively from the a to c panel). Only results using the distributional approach are presented since

Table 4.3: MIDWIN, POLWIN, and POLSUM data set. For clear and cloudy condition the table columns report the total number of simulations (spectra), the number of spectra used to define the training sets and the number of cases used as test sets.

	MIDWIN	POLWIN	POLSUM
Clear	305	244	248
Train. Set clear	205	164	168
Test S. clear	100	80	80
Cloudy	405	284	248
Train. Set cloudy	177	174	201
Test S. cloudy	228	110	47

CIC, on average, produces higher scores when run in this configuration.

The classification scores increases by adding FIR channels to the MIR ones: average DP values are reported in the plots. However, the correlation between CoI and DP is less significant for the present cases with respect to what was found for the simulations in the Tropics. Also the DPs obtained for the MIDWIN, POLWIN, and POLSUM cases are lower than the ones obtained for the Tropical case. This result might be caused by the mean larger temperature differences between high tropical cirrus clouds and the surface with respect to what found for Mid Latitude and Polar cases. It should be also considered that the Polar regions (especially in winter season) present surface features (ice and snow on the ground) that have radiative properties similar to ice clouds and thus making the clear/cloud identification extremely challenging in particular when analysing a test set containing a large number of thin cirrus clouds as the one under consideration.

The POLSUM case (Figure 4.10c) show an evident correlation between the DP and CoI when the full spectrum is accounted for. Moreover, in this case, DP values are on average larger than 0.7 when CoI are larger than 0.85. Therefore a RETS is chosen for this case to be used in testing the ability of CIC to correctly classify clouds with different optical depths.

In Figure 4.11 all the cloudy cases present in the POLSUM test set are classified using features in the MIR only or from the full FORUM spectrum. Results of the classification are plotted as a function of the CSID and of the cloud optical depth. It is noted that the majority of the cloudy cases (also for optically thick clouds) are missed by CIC when relying on MIR channels only (Figure 4.11a). The scores improve significantly when the full spectrum is accounted for (Figure 4.11b). Nevertheless, optically thin clouds (mostly sub-visible cirrus clouds whose $OD < 0.03$) are still misclassified. The misclassified cases do not show any relation with the type of polar surfaces accounted for in the simulations (fine snow, medium snow, coarse snow or ice).
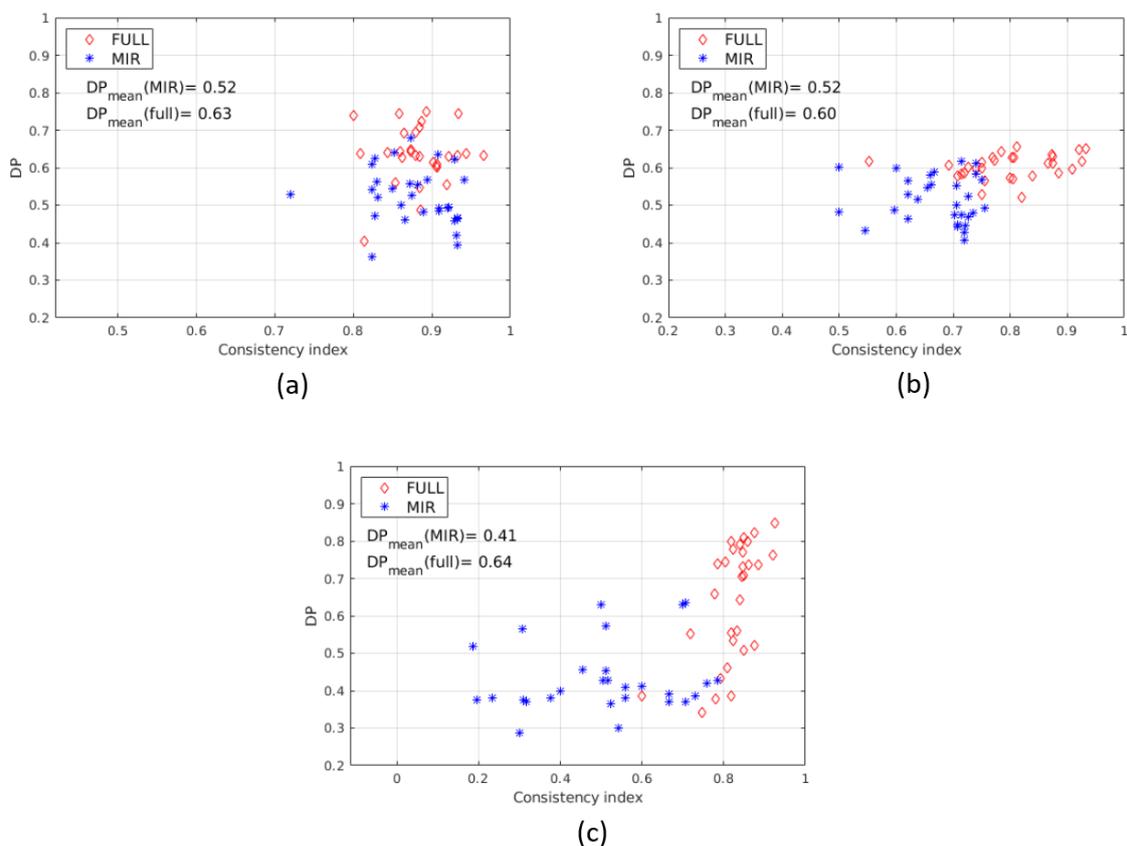
Figure 4.10: Clear/cloud classification DP using the distributional approach for (a) the MIDWIN case, (b) the POLWIN case and (c) POLSUM. Data are plotted as function of CoI and 60 TraNC are used for each case.

4.2.3   Computational time

In this section, a study of the computational time required to calculate the CoI is performed. The CoI computational time is a good indicator of the speed of the core algorithm, since its computation requires the SI computation of all training set elements and the additional optimisation needed for the distributional approach. Moreover, the CIC classification subroutine is the only core routine whose duration depends on the number of features used for the classification. The current version of CIC algorithm is implemented in Matlab® programming language.

When the elementary approach is selected the CoI computation is not necessary. For this reason, the sensitivity study made in this section can also be interpreted as a study of the distributional approach computational cost and results represent an upper limit of the computational time for a classification process in any configuration.
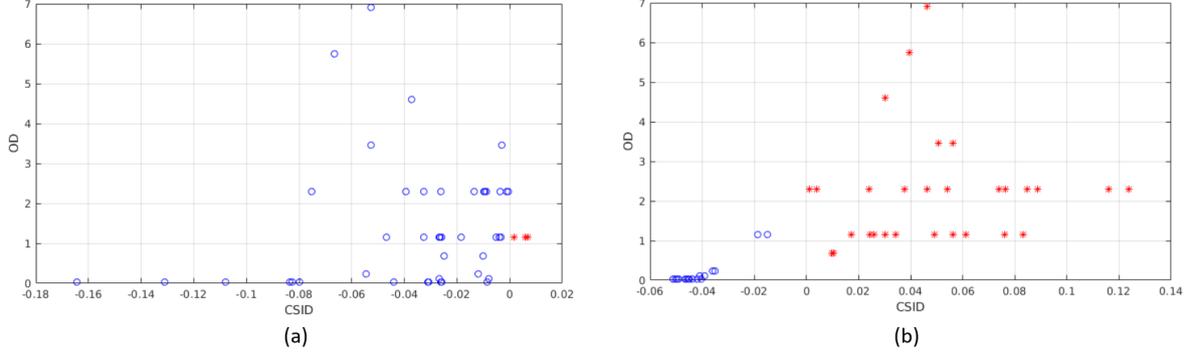
Figure 4.11: Correctly classified (red asterisks) and misclassified (blue circles) cloudy sky spectra from the POLSUM test set. Data are plotted as a function of the CSID and cloud optical depth (OD) and the distributional approach is considered. In (a) only the MIR features are accounted for while in (b) the full spectrum is exploited in the classification.

In Figure 4.12a, the results concerning the computational time of CoI as a function of the number of features (N_{feat}) and of the number of training set elements (here indicated with $T_{tot} = T_{clear} + T_{cloudy}$) are reported. Times are referred to computations performed on a machine with processor Intel i5 (4 cores) and 4GB RAM. It is shown that the computational time increases non-linearly with N_{feat}.

The algorithms for solving linear algebra problems cannot have a linear time complexity (computational complexity describing the time to run the algorithm). A lower bound is given by Raz (2003), where a time complexity lower bound of $\mathcal{O}(n^2 log(n))$ is found for square matrix multiplication. It is demonstrated that the same time complexity bound applies to most other linear algebra problems (Demmel et al., 2007), including eigenvector computation, as performed by CIC.

The CoI computation consists of three main subroutines: factor indicator function computation and its minimisation (see Eq. 4.8), the SI computation (see Eq. 4.7), and $shift_{opt}$ determination (Eq. 4.18). Among the three, the computational cost of the SI is the highest, since the other two only perform a limited and fixed number of cycles involving simple arithmetic operations. This subroutine computes the similarity indices for all T_{tot} training set elements. SI computation requires the costly calculation of the ϵTS matrices (see Eq. 4.3). ϵTS matrices need covariance computation, whose time complexity is $\Omega(N_{feat}T_{tot}^2)$, and eigenvector computation, whose time complexity is $\Omega(T_{tot}^3)$. Therefore, SI computation has an approximate time complexity of $\Omega(N_{feat}^2 T_{tot}^2 + N_{feat}T_{tot}^3)$. The results of the computation of the SI parameter as a function of T_{tot} and for different numbers of N_{feat} are reported in Figure 4.12b.

The running time of the remaining two routines (the one that computes the indicator function minimisation and the one that finds the $shift_{opt}$ parameter) is very limited and less than 0.3 seconds for any of the configurations accounted for N_{feat} and T_{tot}.
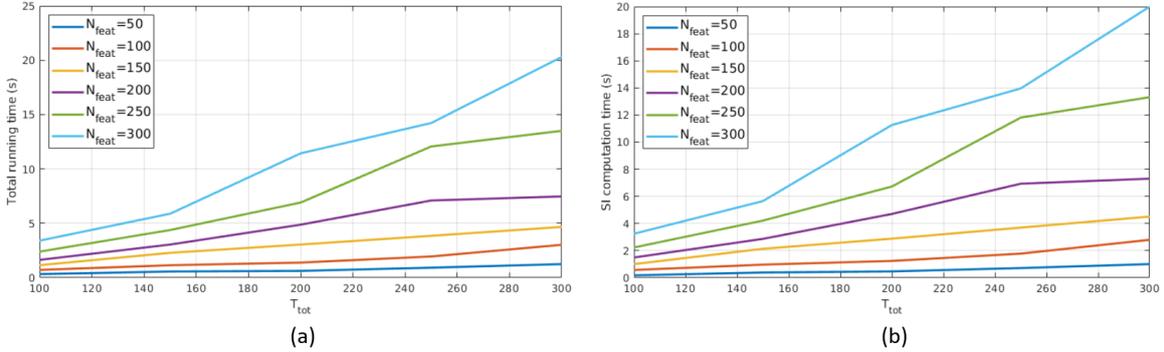
Figure 4.12: (a) CoI coputational time and (b) SI computational time as functions of training set size (T_{tot}) and of the number of features used (N_{feat}). A non-linear relationship between the variables is observed.

Note that the Fourier Transform Spectrometer of the FORUM mission will perform one measurement about every 14 seconds (ESA, 2019), and thus the CIC algorithm could be run operatively on mission data.

4.3 Summary

A new methodology for cloud spectra detection and classification (named CIC: Cloud Identification and Classification) is presented. CIC is a very fast machine learning algorithm based on principle component analysis that depends on a limited number of user defined free parameters. The algorithm exploits spectral variation of the radiance signal, working with the same performance on nighttime or daytime data. Moreover, no a-priori and ancillary information are required.

The algorithm makes use of a training set - TS composed of two or more groups of spectra (each group is a class). The training set elements should represent the observed variability within the classes and thus should include a sufficient number of spectra capable to characterise the radiative features encountered in the area of study. At the same time, they should be 'sensitive' to the addition of new elements with spectral characteristics that are not present in the training set groups. Typically a total number of 100 spectra is sufficient to well characterise the clear sky and cloudy sky training set groups for each latitudinal belt (Tropics, Mid Latitudes, poles) and season. Mathematically the algorithm evaluates the similarity of each test set spectrum with each class and thus provides a classification. CIC, firstly, computes the eigenvectors of the covariance matrix of each TS class - ϵTS. Secondly a test set element is analysed by adding it to each training group (the clear and cloudy sky, for example). The new eigenvectors of the extended covariance matrix - ϵETS (formed by the training set elements of one class

plus the test set element) are computed. An index of similarity is derived for the test set spectrum with respect to the two groups of the training set. It is assumed that a clear sky spectrum, added to the set of spectra defining the clear sky training set, does not significantly modify the group's principal components while some significant modifications will be detected if the clear sky spectrum is added to the cloudy sky training set. Viceversa is for a cloudy sky spectrum of the test set. Similarity indexes are thus defined to quantify the modification of the principal components of the training set when a new element (of the test set) is added. Based on these indexes the element is associated to one of the two considered classes.

CIC can be run by adopting two approaches: the *elementary* and the *distributional* one. The first one is more intuitive and straightforward: the classification for each test set spectrum is done by comparing the similarity indexes computed with the two classes of the training set. The second one requires an additional a-priori optimisation process, at very low computational extra cost (Section 4.2.3). The optimisation is based on the definition of the Consistency Index (CoI) that is related to the detection performance of the algorithm applied to the training set itself. Therefore, if the training set represents natural variability sufficiently well, the CoI serves as a performance forecaster. When optimisation is applied higher scores are obtained, as measured by the increased detection performance (DP, see Section 4.2.1) parameter.

The CIC is initially tested against a large synthetic dataset computed to simulate high spectral resolution radiance from satellite, specifically as possibly observed by the 9th Earth Explorer mission - FORUM mission.

The dataset is divided in subsets in accordance with the latitudinal belt, and season, and the CIC is applied by accounting for different configurations. Results show that the CoI can be used to optimise the training set and that, statistically, the distributional approach is better performing than the elementary one. The code is also used to assess the additional information content derived from the analysis of the far infrared part of the spectrum with respect to the mid infrared only. In this regards it is shown that the overall detection performances increase when the radiance spectra in the far infrared are accounted for. In particular, the radiance exiting the 238-545 cm^{-1} is improving the cirrus detection performances in almost all the atmospheric conditions (latitudinal belt and season). Very thin cirrus clouds (i.e. sub visible cirrus) are better detected when exploiting the full FORUM spectrum than the mid infrared part of the spectrum only. It is shown that, in tropical regions, the detection performances exploiting the full spectrum can be very high (higher than 0.9 for the present dataset that is very challenging for the large presence of sub visible cirri) when the appropriate training set is selected. It is finally noted that clear/cloud spectra identification performances decrease when moving from Tropics to Poles mostly due to the decreased sensitivity of cloudy spectra because of the colder atmospheric and surface temperatures and the increased similarities in the surface and cloud radiative properties.

The CIC functionalities are illustrated for cloud detection application in presence of

high spectral resolution far and mid infrared radiance observations. The results highlight that the combined use of FIR and MIR provides a better classification of the simulated spectra.

Nevertheless, the same algorithm can, in principle, be implemented to work with different kind of data (i.e low spectral resolution data) and also to perform sub-classifications, such as cloud phase identification.

Chapter 5

FORUM End-to-End Simulator

The FE2ES is a tool to predict and assess the Level-1 and Level-2 performances of the FORUM mission. Its architecture is shown in Figure 5.1. This architecture is based on established and well proven end-to-end architectures representative of passive optical remote sensing techniques and instruments (ESA, 2019).

The simulator models the elements relevant to the Level-1 and Level-2 mission performance. Each one of its modules can be run sequentially, continuously (i.e. end-to-end), or individually. The FE2ES starts with an input of the orbit time and geolocation, by using the Geometry Module (GM), followed by the computation of the TOA radiances within the SGM. The TOA radiances are sampled on a spectral resolution much finer than the FORUM mission instruments: the FORUM sounding instrument (FSI) and the FORUM embedded imager (FEI). Spatial and spectral complexities are required to analyse geometric and radiometric uncertainties in the Observing System Simulator (OSS) module for the sounder and imager. After simulating the instruments and the Level-1 processor, the Level-2 products are generated in the Level-2 Module (L2M). The L2M simulates the retrieval of atmospheric, surface and cloud properties by using the radiances which arrives from the OSS modules. The submodule L2M_CIC is used to classify the scene as clear or cloudy. If the scene is identified as cloudy, the L2M_CIC submodule determines the cloud parameters used as initial guess to the inversion submodule (L2M_I). Finally, the Performance Assessment Module (PAM) is used to compare the outputs of the Level-2 processor with the initial inputs (ESA, 2019).

Within the scope of the project "End to End (E2E) Simulator for FORUM – Earth Explorer 9 Fast Track Candidate Mission" (Reference for State of Work: EOP-8MP/2017-12-2053), and its extension in the Change Request N°1 (CR1/CCN2), the research group from University of Bologna has developed the SGM and L2M_CIC. This activity is inserted in the thesis' context since the modules are used to perform sensitivity tests and evaluate the contribution of the FIR to identify a cloud scene.

In this context, the SGM and L2M_CIC are described following, in Sections 5.1 and 5.2, highlighting its functionalities in the framework of the FE2ES. A summarized
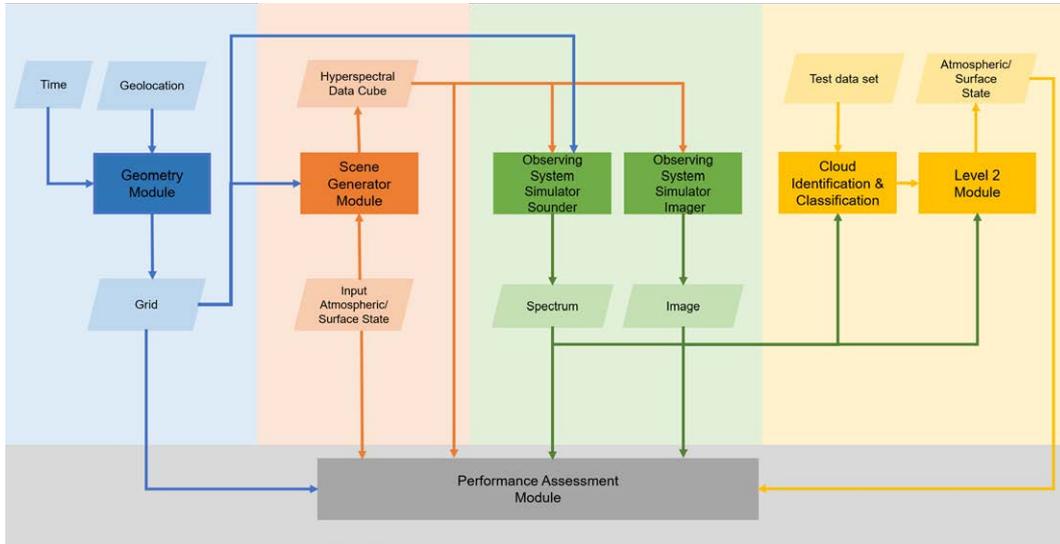
Figure 5.1: High level architecture of the FE2ES. From ESA (2019).
.

description of these and other FE2ES modules is presented in ESA (2019). Then, in Section 5.3, a CIC's performance is presented, and an example of the impact of the cloud identification to retrieve the surface temperature is provided.

5.1   Scene Generator Module

The general goal of the SGM is to define and implement mission representative inputs scenes to be used as inputs for the production of realistic high spectral resolution TOA radiances. The objective is pursued by using the RT models LBLRTM and DISORT, exploiting surface emissivity, atmospheric profiles and cloud properties databases. The observed scenario might be affected by the presence of clouds of any type: ice, liquid phase and mixed phase.

The chain of codes in the FE2ES must be consistent with the previous modules, and also with the modules that come following in the simulator. In this sense, the computation structure of a scene produced by using the SGM is different of that used to build the synthetic dataset presented in Section 3.1, because in the SGM everything must to be coherent and automatic.

The SGM is managed by a bash shell script, which launches their sub-modules: SGM1_profiles; SGM2_LBLRTM; SGM3a_Rad_Interp; SGM3b_LBLDIS and SGM4_Results. The overall SGM's architecture is reported in Figure 5.2.

The SGM's shell script starts launching the SGM1_profiles, an executable Matlab® function which reads the input geolocation grid and time information, from the GM, and
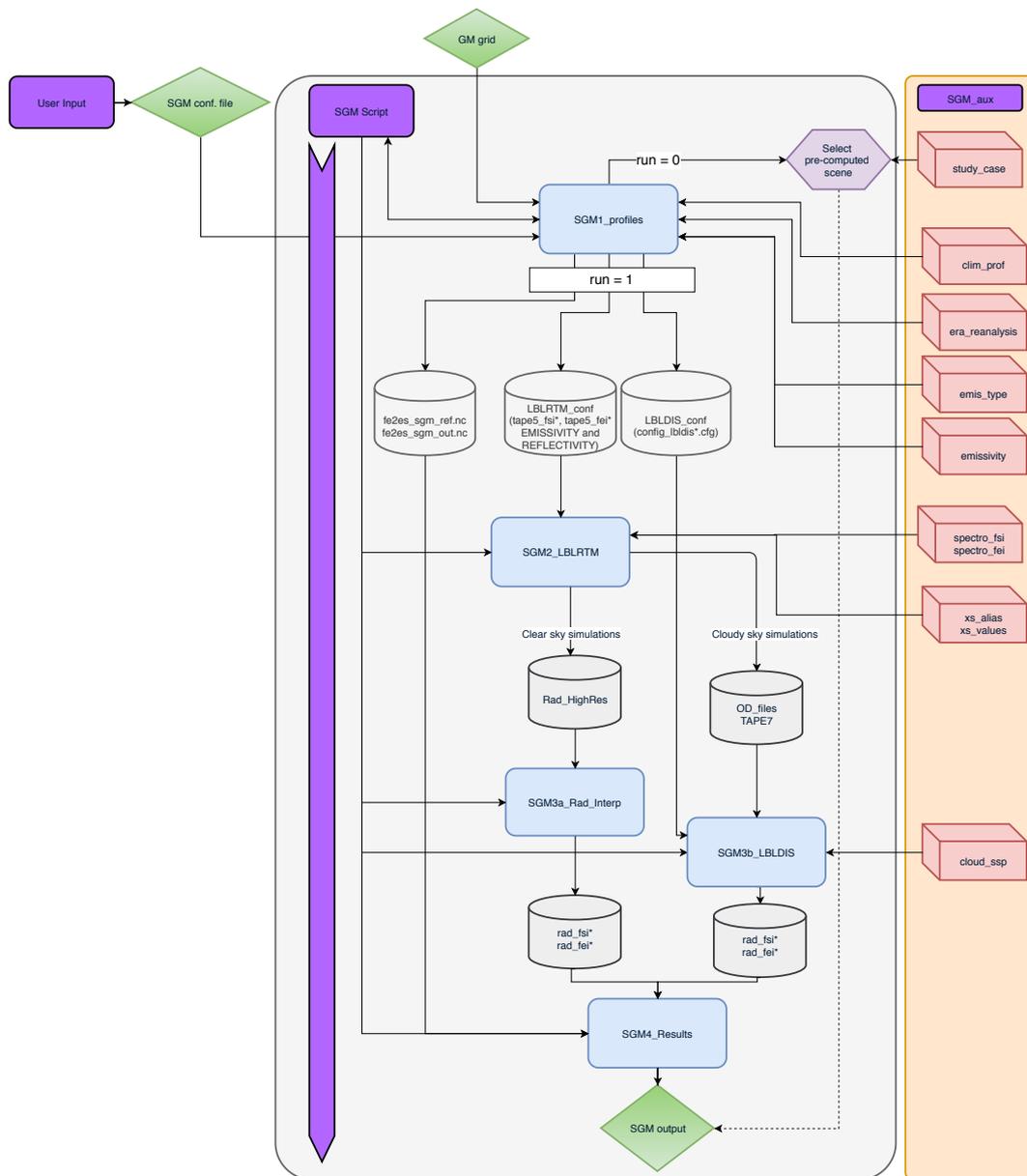
Figure 5.2: SGM's flowchart for clear or cloudy scenes. Light blue box are submodules (SGM1, SGM2, SGM3a, SGM3b, SGM4). Auxiliary files are represented by red boxes; SGM's inputs and outputs are in green; internal data flow and files are indicated by light grey cylinders.

the user's input information, from a configuration file. This function is used also to create the configuration files for the LBLRTM and LBLDIS, besides building the scene from the atmospheric profiles, surface properties (spectral emissivity and temperature) and cloud properties (cloud phase, cloud ice particle shape, cloud top height, cloud geometrical thickness, cloud optical depth at 900 cm^{-1} and cloud effective radius).

From the SGM's configuration file, it is possible setup the scene which will be simulated. The configuration parameters are:

- **Atmospheric Gaseous Composition**: The SGM uses a maximum of 12 gases (H_2O, CO_2, O_3, N_2O, CO, CH_4, O_2, NO, SO_2, NO_2, NH_3 and HNO_3). The H_2O and O_3 are derived from the ERA 5 reanalysis (Hersbach et al., 2020) between the surface and 60 km, merged with the IG2 database (Remedios et al., 2007) between 60 and 80 km. The other gases are all taken from climatological IG2 database, however, the CO_2 has been linearly corrected to the currently concentration level, multiplying by a scale factor to adjust the mixing ratio estimated in 2013 by IG2, to the concentration level in 2018.

- **Surface Inhomogeneity**: The surface's scene can be defined as homogeneous, i.e. the surface emissivity and temperature are kept the same in all gridded points from the GM, or heterogeneous, which is characterized by two surface types with different emissivity and temperature. The second type of surface is defined as a circle or a sector inside of the area observed by the FORUM instruments. In Figure 5.3 are shown examples of the surface inhomogeneity, with a detailed description in the caption.

- **Surface Emissivity**: According to the level of surface inhomogeneity, it can be defined up to 2 types of surface, hence, up to 2 surface emissivities spectrally resolved. The surface emissivity can be defined as constant, by using a constant value between 0 and 1 in all wavenumbers of the spectrum that will be simulated, or can be derived from the global database produced by Huang et al. (2016). This second case is subdivided into three options: the first one is to impose a specific surface emissivity, by using one of the 11 spectral surface emissivity available from Huang et al. (2016); the second one is to use the geolocation and time to get the average surface emissivity derived from MODIS measures between 2008 and 2014 (Huang et al., 2016); the last one is to use a surface category for the geolocation and time, determined by using a minimization algorithm which compute the least standard deviation of the difference between the average MODIS data and the Huang et al. (2016) surface type database. In Figure 5.5 are shown the global surface categories for January and July.

- **Surface Temperature**: As the surface emissivity, it is possible to define up to 2 values of surface temperature, according to the surface inhomogeneity. The surface

temperature can be imposed by using any positive value, in K, or derived from the ERA 5 reanalysis, using the scene geolocation and time.

■ **Surface height**: It is possible to use the surface height from ERA 5 reanalysis for the scene geolocation and time, or over impose it on a level higher than that one.

■ **Cloud Inhomogeneity**: The SGM allows simulate radiances in cloud sky conditions, with the cloud covering all the gridded points of the scene (i.e. overcast), as well as the cloud covering partially a region of the scene. When a cloud cover a part of the scene, it can be defined as a circle or a sector inside of the area observed by the FORUM instruments, like is made for the surface inhomogeneity (see Figure 5.3). The parameters used to characterize a cloud are described in the next item. There is also an option to setup all the cloud parameters automatically, by using ERA 5 variables and applying some simplification and parametrizations, that are explained in Appendix 1.

■ **Cloud Properties**: Clouds are defined in the SGM by the following parameters:

☐ Number of cloud layers/type - the SGM allows to include up to 2 types of cloud in the same vertical profile, e.g., a water cloud in low levels followed by a cirrus cloud in high levels of the atmosphere, as shown in Figure 5.4;

☐ Cloud phase - the cloud can be composed by spherical liquid water or mixed-phase particles, being that the mixed phase consist of a core of ice surrounded by 10% coating of liquid water, or by non-spherical ice crystals;

☐ Cloud particle shape - for non-spherical ice crystals, there are two shapes available: aggregate and hexagonal plates;

☐ Cloud top height

☐ Cloud geometrical thickness

☐ Cloud optical depth at 900 cm^{-1}

☐ Cloud effective radius

■ **Spectral range and resolution**: The spectral range and resolution of the radiance that are used for the FSI.

The first step of the sub-module SGM1_profiles is analysing the user's input information, verifying its limits and the range of applicability. If any setup is not agreeing with the scene that can be computed, the system stops and a error message is displayed.

Next, the inhomogeneous structures are defined and mapped on the FEI and FSI grid provided by the GM. Each grid point is associated to a radiance computation. The number of radiance computations is in accordance with the inhomogeneity level defined by the user. For example, for a scene like that of Figure 5.3d, with two different surface
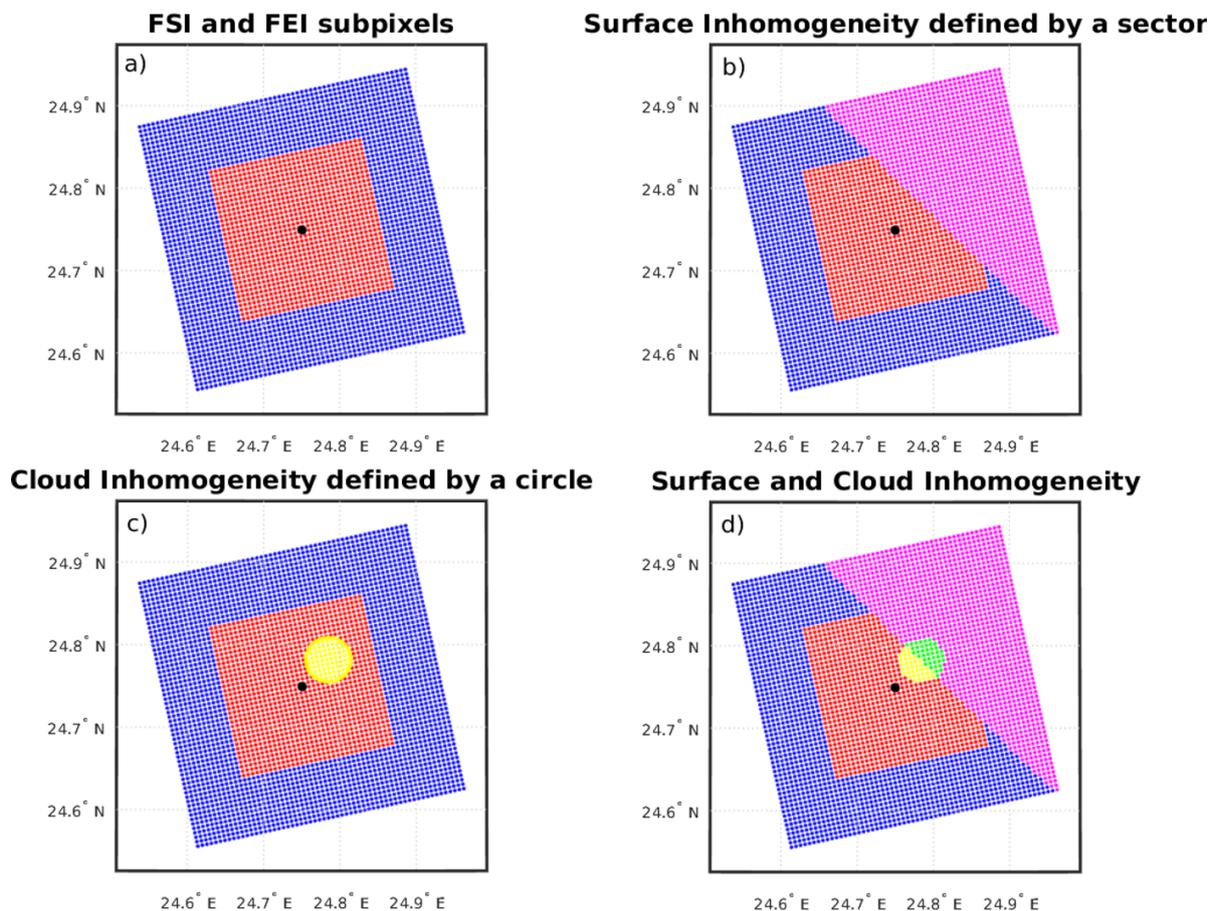
Figure 5.3: (a)Example of a gridded scene, where the red points indicate the satellite subpixels both from FSI and FEI, and blue points indicate only FEI subpixels. The black asterisk indicates the Sub-Satellite Point, which is the reference center point of the satellite observation. The scene inhomegeneity can be characterized by: (b) two different surfaces, being that the second surface is represented by a sector (magenta points); (c) a cloud circle, indicated by the yellow points; (d) two different surfaces, as in (b), with a circle cloud, as in (c), covering partially each one surface type. The green points indicate the area of the second surface type which is covered by the cloud. This image represents the maximum inhomogeneity level of the SGM.
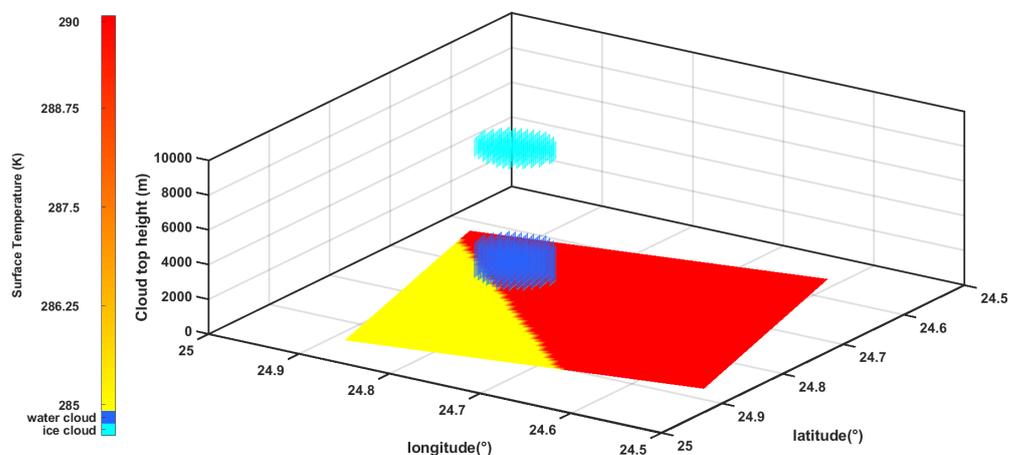
Figure 5.4: Example of an inhomegeneity scene characterized by two different surfaces (two different values of surface temperature), with two layers clouds covering the same area, being the highest cloud compose by ice crystals, while the lowest is a water cloud. This inhomegeneity scene is the same of that presented in Figure 5.3d, with a 3-Dimensional view, to highlight the 2-layers clouds.

types and a partially cloud covered grid, the LBLRTM model will be executed 3 times: twice to calculate radiances in clear sky conditions, one for each type of surface, and one to compute the gaseous optical depths for each atmospheric layer, that will be merged with the cloud radiative properties in LBLDIS. Then, LBLDIS will be executed twice considering a cloud over each type of surface.

Afterward, the sub-module defines the atmospheric vertical profiles, by using the nearest ERA 5 grid point regard to the center point of the scene that will be simulated. The ERA 5 Reanalysis account for atmospheric vertical profiles, between the surface and 60 km, of the temperature, pressure, water vapour and ozone mixing ratios, besides information on the surface temperature, altitude and cloud parameters, restricted to the SGM's area of study (0°-30°E; 0°-85°N). The profiles of the other gases accounted for in the computations (CO_2, N_2O, CO, CH_4, O_2, NO, SO_2, NO_2, NH_3 and HNO_3) are all taken from climatological IG2 v5.4 database, obtained according to the latitudinal belt, day period and season.

The profiles from ERA data are interpolated at fixed altitude levels to obtain homogeneous files with a predefined and controlled number of altitude levels for the profiles used in the computations. Data concerning temperature and ozone are linearly interpolated, while pressure and water vapour are interpolated with a logarithmic function.

The IG2 database contain also climatological vertical profiles of pressure, temperature, water vapour and ozone mixing ratios. These data are merged with the ERA 5
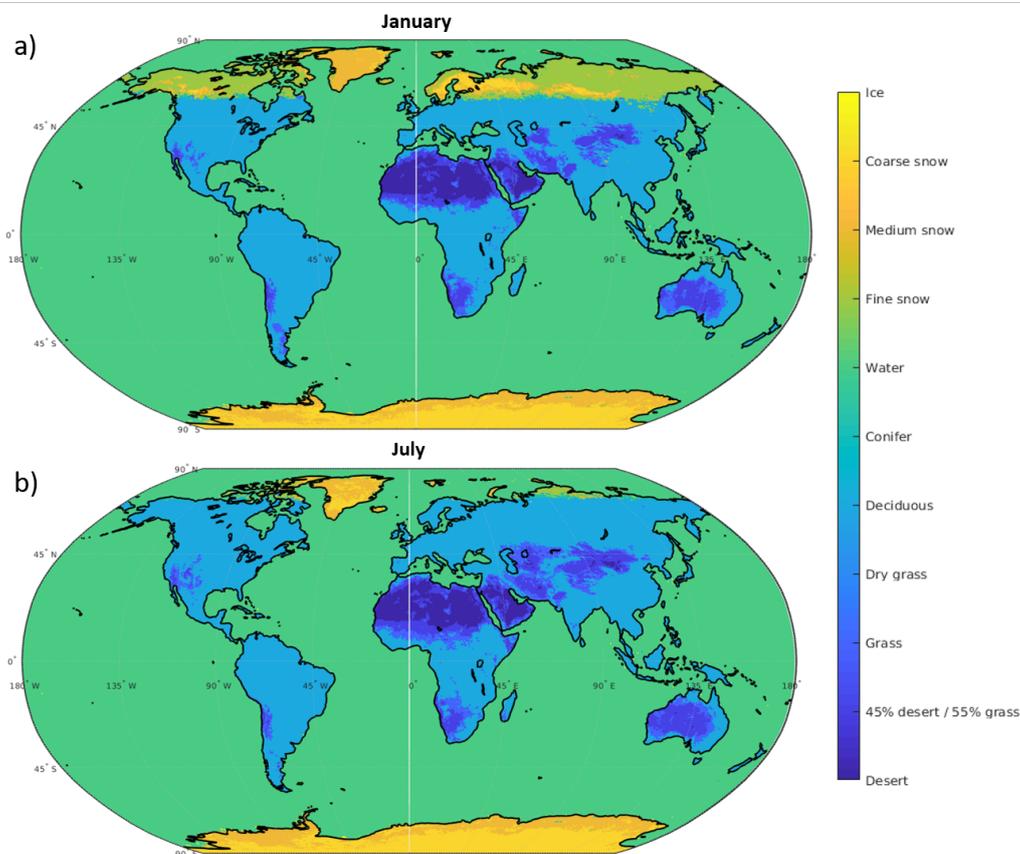
Figure 5.5: Global surface categories according to 11 surface types from Huang et al. (2016) in (a) January and (b) July.

variables between 60 km and 80 km, which is the maximum altitude level used in the SGM. The merging between the different source data is made smoothing the gap between them, i.e., the lastly levels of the ERA data are interpolated with the first data of the IG2, minimizing the distance between the measurements.

Following, the SGM1_profiles writes the configuration input files of the LBLRTM model, for calculating the radiances, in the clear sky case, or the atmospheric layers' optical depths, in the cloudy sky case. The number of files produced is dependent on the scene inhomogeneity level.

The next step is writing the emissivity/reflectivity files, according to the SGM's configuration file. The reflectivity is calculated by assuming the LTE and the Kirchhoff's law, besides assuming a zero transmittance surface ($Reflectivity = 1 - Emissivity$). In case of cloudy scenes, the SGM1_profiles writes also the input configuration file for LBLDIS, by using the cloud parameters defined in the SGM's configuration file.

Finally, the first SGM sub-module prepare the output file containing the grid of

the scene, mapping each high spectral resolution radiance with the corresponding FEI and FSI subpixels. However, the radiances are only added at the end of the SGM execution, by the sub-module SGM4_Results. Furthermore, the output file contains all data provided by the GM, besides the atmospheric profiles, surface information, and cloud properties.

Thereafter, the SGM's bash shell script launches the second sub-module (SGM2_LBLRTM), which is actually the LBLRTM RT code. This version of the model implemented in the SGM uses the auxiliary database of spectroscopic gases properties AER v3.6, derived from the HITRAN database (plus updates), and is consistent with the state of the art parameterization of the water vapour continuum (MT_CKD 3.2) and others gases continua. The model is used to compute radiances in clear sky conditions. In this case, its output file is a binary file, that is converted in an ASCII file by using the sub-module SGM3a_Rad_Interp.

In case of cloud sky conditions, the LBLRTM model is used to compute the atmospheric layers' optical depths, which are used as input by next SGM's sub-module - SGM3b_LBLDIS. This sub-module is basically the LBLDIS. The model uses as input the configuration file written by the SGM1_profiles, the atmospheric layers' optical depth, computed by the LBLRTM model, besides the auxiliary cloud database, derived from the Ping Yang database (Yang et al., 2013) for non-spherical particles, and from *scattnlay* (Peña and Pal, 2009) for spherical particles.

In the last SGM's sub-module (SGM4_Results), the radiances computed by the RT models are merged in the SGM's output files, which are passed to the modules OSS and PAM.

Since the radiances computations are the most time consuming of the SGM, and hence of the FE2ES, the SGM also includes some pre-computed cases which allow to skip computations of the RT models and to pass quickly the results to the subsequent modules.

Currently, there are 6 cloudy and 6 clear scenes performed by using the SGM, that can, then, be repeated. The gridded area of the clear sky scenes is the same as the cloud sky cases, which allow us to evaluate the impact of the cloud in those scenes. All the scenes are homogeneous and the cloud cases represented are: cirrus on the desert; cirrus on the ocean; marine stratus; continental stratocumulus; ice cloud on snow ground; and Po valley cumulonimbus. In Table 5.1 are summarized the surface and cloud characteristics of each one of six cloud study cases. For the clear sky cases, the cloud parameters are unconsidered.

There are also 3 MODIS based scenes, being 2 mostly covered by high clouds (one over the land and one over the sea), and 1 clear sky scene. These scenes were created to try to achieve more realistic scenes as possible.

To compute the spectrally resolved radiances of these scenes a chain of bash shell scripts and Matlab® codes were adapted to reproduce the same SGM's structure, considering, however, all the inhomogeneities related to the scenes. In this sense, different

Table 5.1: Surface and Cloud parameters used in the six cloud SGM study cases

	Cirrus on desert	Cirrus on ocean	Marine stratus	Continental stratocu- mulus	Ice clouds on snow ground	Po valley cumu- lonimbus
Latitude	24.75°N	33.75°N	39.75°N	50.25°N	68.25°N	45.00°N
Longitude	24.75°E	18.75°E	6.75°E	6.75°E	18.75°E	11.25°E
Surface Type	Desert	Sea Water	Sea Water	Deciduous	Fine Snow	Deciduous
Surface Temperature	326.3 K	300.2 K	298.8 K	274 K	253.9 K	296.5 K
Cloud Phase	Ice	Ice	Water	Ice	Ice	Ice
Cloud Habit (for ice clouds)	Aggregate	Aggregate	–	Aggregate	Aggregate	Aggregate
Cloud Top Height	12 km	15 km	1.25 km	10 km	8 km	11 km
Cloud Geometrical Thickness	2 km	1 km	0.5 km	2 km	2 km	6 km
Cloud Optical Depth at 900 cm^{-1}	1	0.3	20	6	3	300
Cloud Particle Effective Dimension	10 µm	6 µm	10 µm	30 µm	18 µm	80 µm

databases were used to reproduce the atmospheric conditions, and the surface and clouds properties.

The surface emissivity was obtained by using the GlobCover 2009 product Bontemps et al. (2011), which is a global land cover map constructed from the global Medium Resolution Imaging Spectrometer Instrument (MERIS) Fine Resolution mosaics for the year 2009, with about 300m of spatial resolution. This map, however, counts 22 land cover classes defined with the United Nations Land Cover Classification System (LCCS). Then, each one of 22 surface types is associated with one of the 11 surface types from Huang et al. (2016) database, which give us the surface emissivity.

The surface temperature and the cloud parameters, i.e., the cloud water phase (ice or liquid water), the cloud particle effective radius, the cloud optical depth and the cloud top height, were derived from the MODIS Cloud Product (Platnick et al., 2015), which has a spatial resolution of 1km at nadir. The cloud geometrical thickness was imposed

to occupy just one layer of the vertical profiles of SGM. This assumption is also used to optimize the time of the radiances computation. Aggregate columns were assumed as the particle ice shape.

The MODIS Atmospheric Profile product (Borbas et al., 2016) was used to extract the temperature and water vapour profiles, the surface height, besides used to calculate the pressure vertical profile. This MODIS product has a spatial resolution of approximately 5 km at nadir, and the vertical profiles arrive at around 30 km of altitude. Then, they are completed with data from the IG2 database, up to 80 km. Moreover, the MODIS Atmospheric Profile product is available only for the clear sky scenes. Thus, for cloud sky scenes, were used the vertical profiles from the ERA Interim reanalysis. The ERA Interim was used, instead of the ERA 5, because in the FE2ES phase A project, the SGM used this reanalysis as the auxiliary database, and in its update in the phase B1 of the project, the MODIS based scenes were not updated.

All these products, from MODIS and MERIS sensors, have a different grid and spatial resolution. Thus, to create the scenes, initially a Matlab® code is used to homogenize the information in its grid. An example of the adjustment of the grid from the MODIS products and GlobCover2009, to the GM grid is shown in Figure 5.6.

After constructing the grid with all cloud, atmospheric, and surface information, this first Matlab® code is used to identify the grid points which contains the same parameters and determine the level of inhomogeneity of the scene (Figure 5.7). This step allows reducing the number of radiances computations. Taking the Figure 5.7 as an example, the FSI grid has 34x34 pixels, which would imply to compute 1156 high spectral resolution radiances, however, the maximum of the simulation necessary for a scene was 415, because there are pixels with the same characteristics of cloud, surface, and atmospheric vertical profiles.

The last step of this first code is to write the configuration/input files for LBLRTM and LBLDIS, that are used to compute the radiances according to the inhomogeneity level, and to produce the SGM output files, as is made in sub-module SGM1_profiles.

Next, a shell script is used to launch the LBLRTM and the LBLDIS, which will calculate the radiances for the grid points of the scene. For the clear sky grid points, the radiances are calculated directly by the LBLRTM model, and the output files produced are binary files. Thus, it is necessary to execute a Matlab® code to read the binary files and write the output as an ASCII file, how is made by the sub-module SGM3a_Rad_Interp. For the cloud sky grid points, the outputs from the LBLDIS are directly in ASCII. In Figure 5.8 are shown the radiances computed for the MODIS scenes based study cases in clear sky condition (Figure 5.8c and d, for FEI and FSI, respectively), and for the case with high clouds over the land (Figure 5.8a and b). It is possible to note the higher spread of the radiances in cloud conditions when compared to the clear sky case, either for FEI or for FSI.

There are other two Matlab® codes to ingest the radiances previously computed in the final SGM output files, and ordering the results in the correct way.
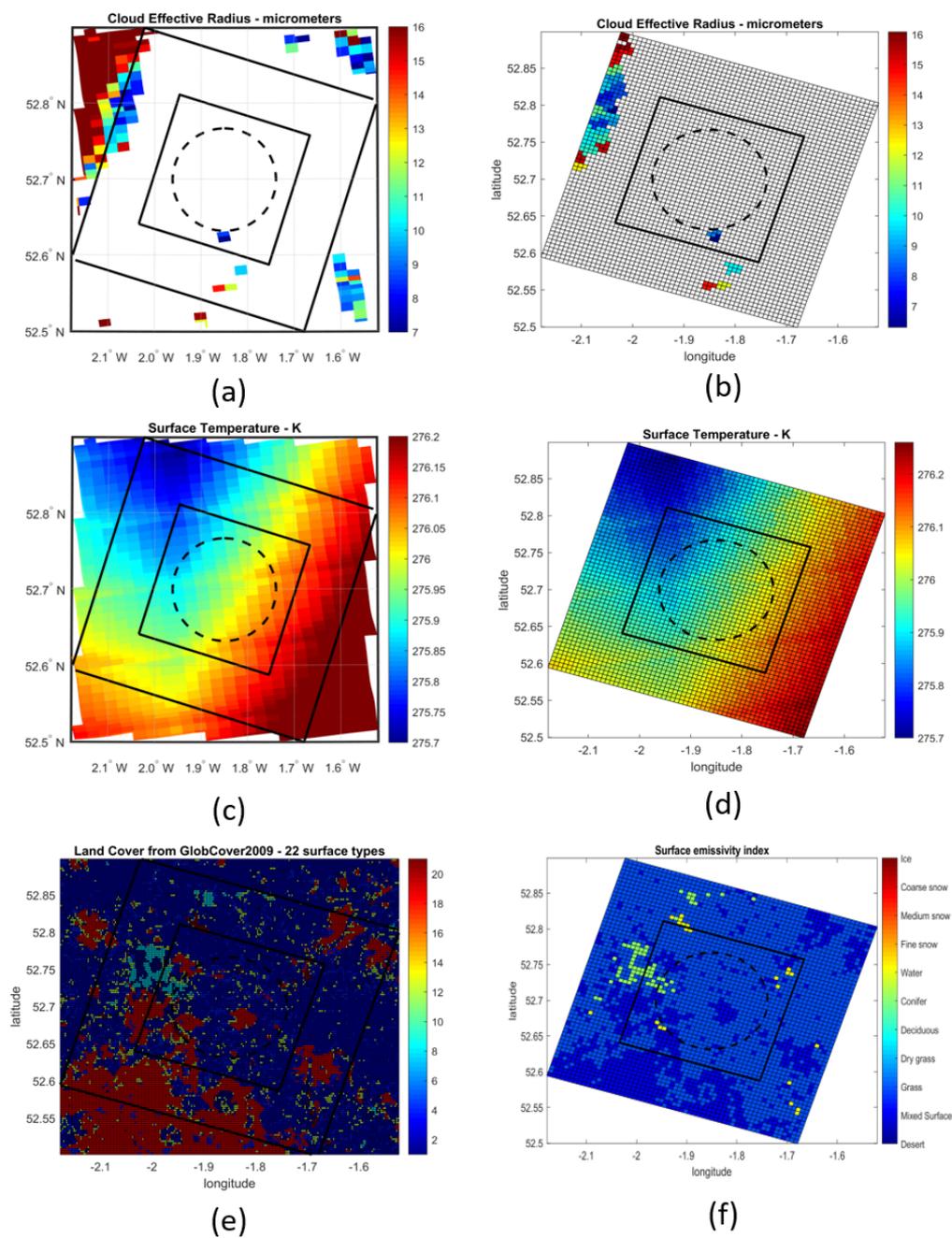
Figure 5.6: Example of the translation of the MODIS products (a and c) and GlobCover2009 (e) to the GM grid (b,d, and f), to be used as input for the SGM MODIS scene based.
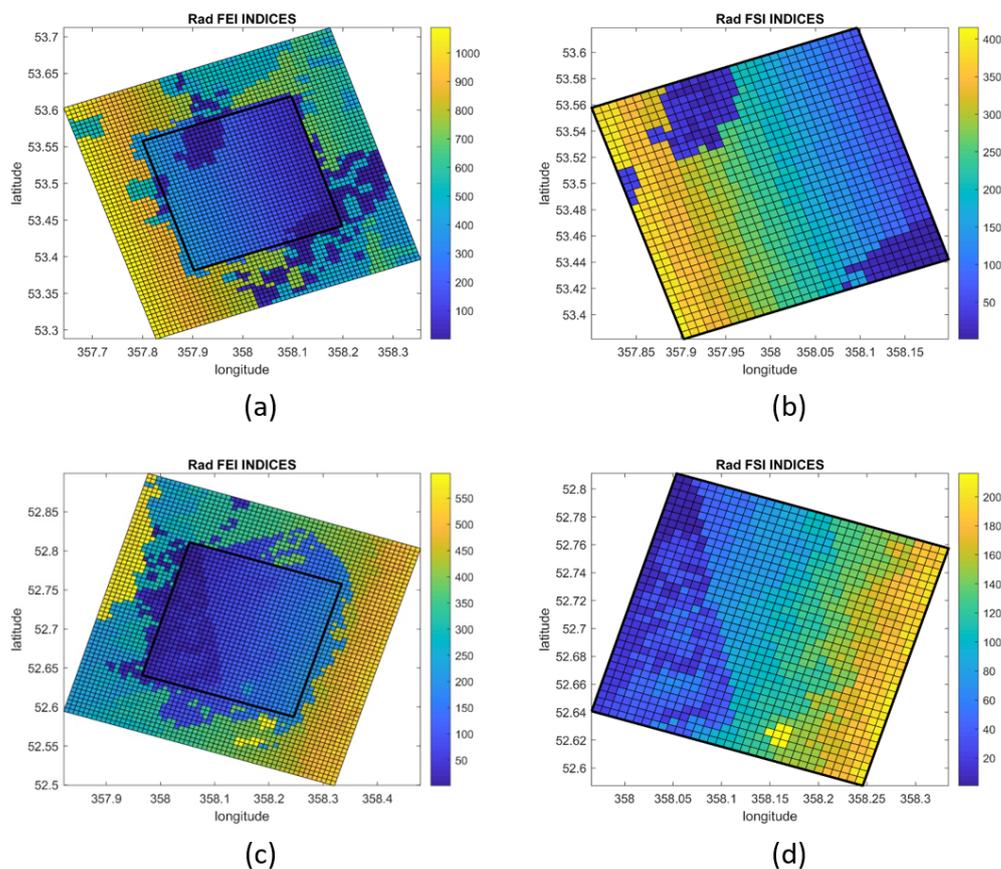
Figure 5.7: FEI (a) and FSI (b) radiances indices which point out the inhomogeneity level of the MODIS high cloud over the land scene. FEI (c) and FSI (d) radiances indices which highlight the inhomogeneity level of the MODIS clear sky scene. When the index is the same for different grid points, it means that the grid points have the same radiance, either for FEI or for FSI.

All these pre-computed scenes are also used to evaluate the performance of the CIC algorithm, that is presented in Section 5.3.

5.2 L2M - CIC

The first results of the CIC were developed within the frame of the SCIEF ("Development of Italian Competences for the FORUM Experiment") project of the Italian Space Agency (ASI), and published in the manuscript from Maestri et al. (2019b). Because of its capability, the CIC algorithm was adapted and inserted in the structure of the FE2ES, in the scope of the submodule L2M_CIC.
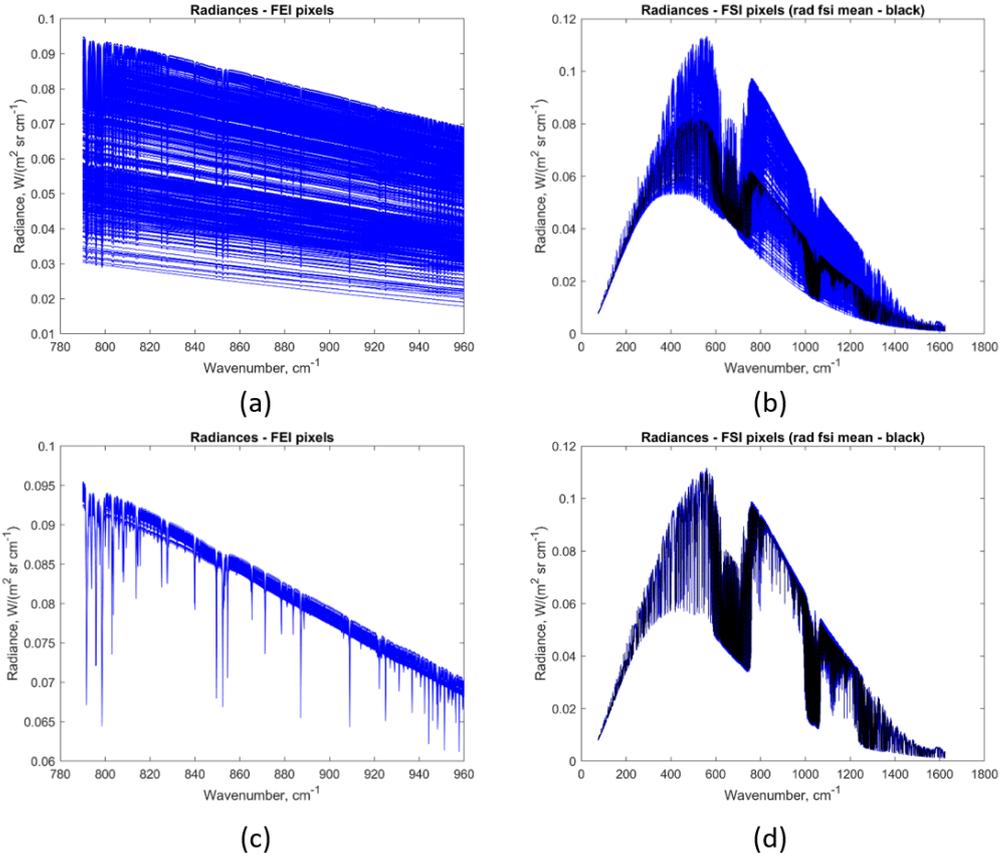
Figure 5.8: FEI (a and c) and FSI (b and d) high spectral resolution radiances for the case of high clouds over the land (a and b), and for the case of clear sky (c and d). The black line in the FSI radiances is the radiance mean inside of the FSI field of view.

As the goal of this chapter is to highlight the application of the CIC algorithm in the FE2ES, just the L2M_CIC is described here. A detailed description of all L2M is presented in ESA (2019).

L2M_CIC consists of four subroutines written in Matlab® and managed by a bash script. The four subroutines are: CIC_coast, CIC_id, CIC_cf and CIC_inguess.

The CIC_coast subroutine is used to determine the sea fraction of the FSI Field of View (FoV). A Global Self-consistent, Hierarchical, High-resolution Geography Database (GSHHG) is used as auxiliary data in this subroutine. This database consists of geolocated polygons that represent the global coastlines. From the FoV of the instrument, determined by the GM, the sea fraction is computed, being equal 1 if all FoV is on water surface.

The subroutine CIC_id is properly the CIC algorithm adapted for the FE2ES. The

algorithm is initially used to discriminate if a radiance, simulated by using the previous modules GM, SGM and OSS, refers to a clear or cloudy sky scene. Next, in case of cloud sky, the cloud is classified in terms of phase and opacity, like liquid water, ice or cirrus cloud. This classification is very important to determine which retrieval will be performed in the L2M. For the case in which a clear sky is identified, the vertical profiles of water vapour and temperature, and the surface temperature are retrieved simultaneously, by using the optimal estimation technique (Rodgers, 2000). Moreover, when the precipitable water vapour is below a certain threshold, the surface emissivity is also retrieved. On the other hand, in case of cloud sky, the cloud parameters are retrieved, i.e., cloud top height, cloud effective radius, cloud optical depth and ice/liquid cloud water path.

As introduced in Section 4.1.1, in the FE2ES the CIC classifies a scene as one of the four classes: clear sky, liquid water, ice or cirrus cloud. One of these classes must be prevalent in comparison with the others for a scene to be classified (see Figure 4.4). If there is no prevalent class, the scene is unclassified.

To perform the classification, training sets are build by using the modules GM, SGM and OSS, considering not only the classes, but also:

- five latitudinal belts in the SGM area: arctic (between 70°N and 85°N), highlatitude (between 55°N and 70°N), midlatitude (between 35°N and 50°N), tropical (between 15°N and 35°N), and equatorial (between 0°N and 15°N);

- two seasons: winter and summer;

- three types of instruments, that can be set-up in the OSS module[1].

In Figure 5.9 is shown a scheme indicating the steps for selection of the training set spectra to perform the cloud identification and classification using the CIC algorithm. In the map of the figure are indicated the geolocations of the TSs, related to each latitudinal belt. The points were selected to cover a wide atmospheric and cloud variability, totalising 1170 spectra both clear and cloudy. All spectra are processed by using the OSS module for the three instruments available.

After performs the scene classification, the subroutine CIC_cf is executed from the L2M_CIC. This subroutine is used to evaluate the homogeneity of the scene. The homogeneity is evaluated by analysing the distribution of the BT observed from the imager instrument, within the FSI FoV. The BT's frequency distribution is computed. If all the FEI pixels fall within a limited range of BT (established here as $\Delta BT = 2K$), the scene is considered homogeneous. If the scene is classified as clear sky, it is assumed uniformly

---

[1]In the preparatory activities of the FORUM mission (Phase A/B1), two technical baseline concepts of the instrument were developed by two industrial consortia led by the Airbus Defence and Space Limited (ADS), and led by Thales Alenia Space UK (TAS), respectively. Moreover, an ideal convolution with a sinc function is included to accounting a third ILS in the FE2ES (SINC)
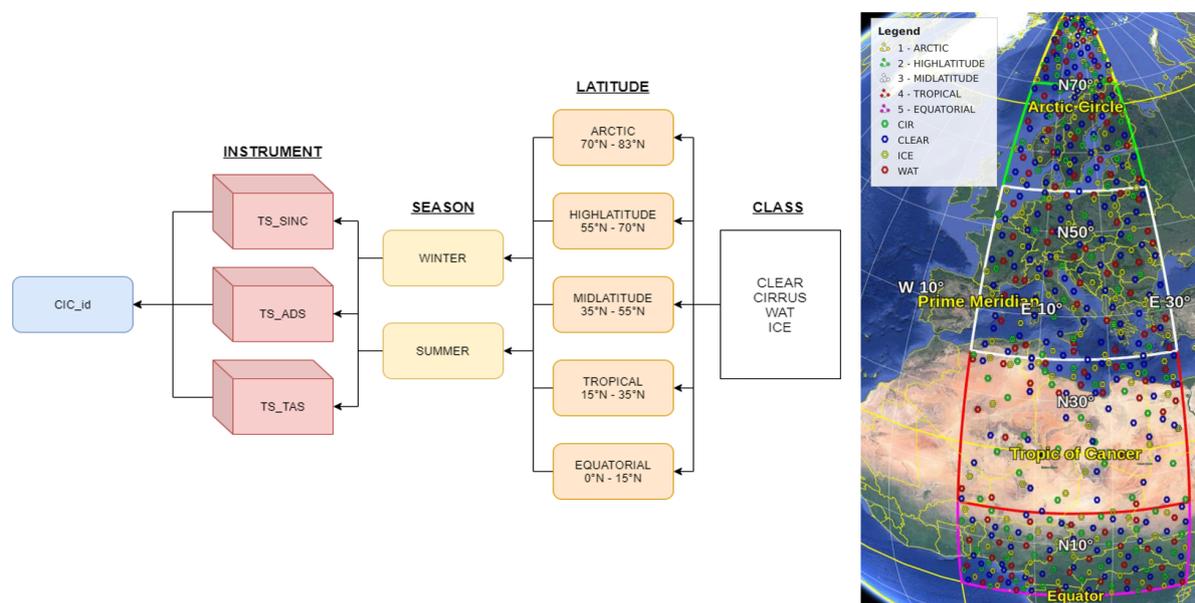
Figure 5.9: Selection scheme for the training set spectra and location on a map. Clear, cirrus, water cloud, and ice cloud scenes are marked with blue, green, red, and yellow symbols, respectively. Solid lines marks the arctic (yellow), highlatitude (green), midlatitude (white), tropical (red), and equatorial (pink) regions.

clear, as well as a cloudy scene is assumed uniformly cloudy when it is indicated to be homogeneous.

In case of differences larger than the selected ΔBT, the inhomogeneities in the scene are evaluated. The histogram distribution of the BT within the FSI FoV is computed at 2 K steps, and a polynomial function is used to fit the BT histogram. The order of the polynomial function can vary between 2 and 10. This order is determined by minimizing the norm of the fitting residuals. Then, maxima and minima of the distribution are found, and the inhomogeneity levels are defined as the histogram elements between the minima. For each group of data between two consecutive local minima, the corresponding FSI FoV fraction is computed. An example of this process is shown in Figure 5.10, which presents the distribution of the MODIS-based scene regards to high clouds over the land. Four levels of inhomogeneity are found, corresponding to fractions of 9%, 60%, 14%, and 17%, which are indicated in the figure by the colors grey, blue, pink, and red, respectively.

The last subroutine executed in the L2M_CIC is the CIC_inguess. This routine only runs if the CIC_id defines the scene as cloudy. The goal of the subroutine is to give a first estimate of the cloud properties for the inversion submodule of the L2M. Climatological databases from ESA CCI project (Stengel et al., 2017) and ERA 5 reanalysis are used to determine the initial guess of the cloud top height, cloud geometric thickness, cloud optical depth, and cloud effective radius.
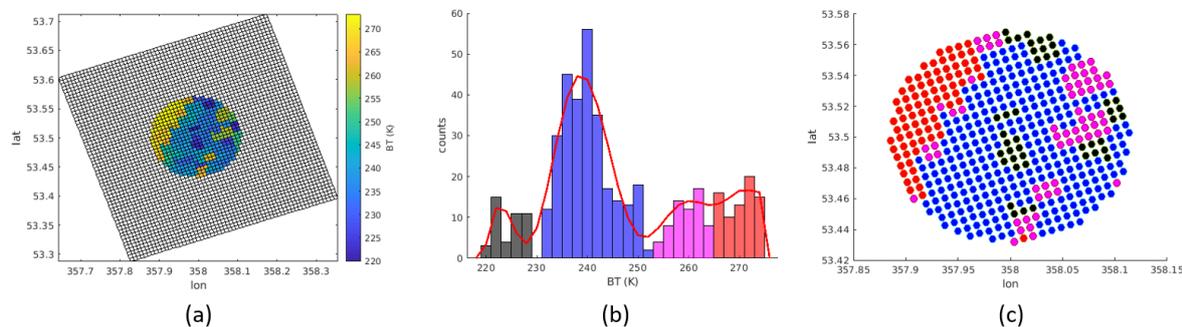
Figure 5.10: BT distribution of FEI pixels within the FSI FoV for an inhomogeneous MODIS-based scene (high clouds over the land). Different colors identify different inhomogeneity levels, both (b) in the histogram and (c) in the geolocated perspective.

The L2M_CIC flowchart is presented in Figure 5.11. Next, some classification from the CIC algorithm in the context of the FE2ES are presented, together with an example of the impact of the cloud identification to perform the retrieval of the surface temperature.

5.3 CIC Applications on the FE2ES

To evaluate the CIC performances in the FE2ES, the classification of the 6 clear and cloudy pre-computed scenes of the SGM, besides the 3 MODIS based scenes, are carried out. It is important to highlight that inside of the FE2ES, the CIC classifies the scenes by using the elementary approach.

In Table 5.2 the results of the classification are summarized, indicating the good performance of the CIC to classify the scenes from the SGM study cases. Just the MODIS based scene regard to high clouds over the land, using the sinc function as ILS, is misclassified. The results also highlight the relevance of the instrument and the preparation of the TS for the correct classification of the algorithm. The TS must consider the wide range of the atmospheric, surface, and cloud conditions.

The FE2ES can be used to assess the CIC's ability to detect clouds in other specific study cases. For instance, evaluating the contribution of the FIR part of the spectrum to identify and classify an inhomogeneous cloud scene, with a cloud covering partially the FSI FoV.

Within this framework, inhomogeneous scenes are simulated by using the SGM, with the same cloud and surface characteristics of the study case "Cirrus on desert" (see Table 5.1). However, the cloud fraction of the FSI FoV varies in these simulations. The cloud is defined as a sector of the FSI FoV, thus the cloud area does not vary linearly.

After computing the radiances, by using the SGM and OSS modules, the CIC's ability
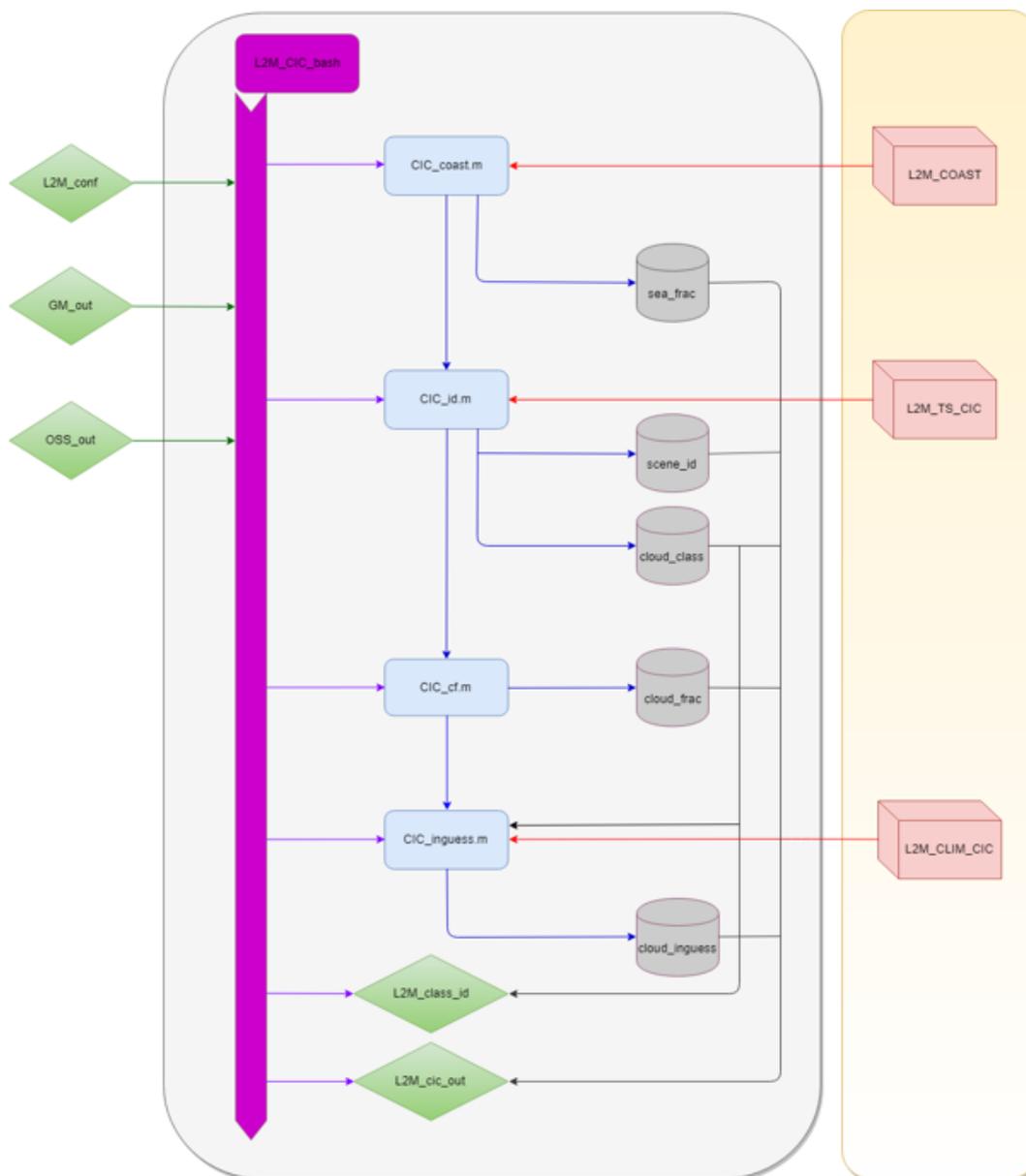
Figure 5.11: L2M_CIC's flowchart. Light blue box are submodules (CIC_coast, CIC_id, CIC_cf, and CIC_inguess). Auxiliary files are represented by red boxes; inputs and outputs are in green; grey cylinders represent the internal data flow.

Table 5.2: Performance of the cloud classification for the 15 study cases of the SGM, according with the instruments. "V" indicates a successful detection, and "X" unsuccessful.

STUDY CASES	ADS	TAS	SINC
Case 11 - Clear sky on desert	V	V	V
Case 12 - Cirrus on desert	V	V	V
Case 21 - Clear sky on ocean	V	V	V
Case 22 - Cirrus on ocean	V	V	V
Case 31 - Clear sky on ocean	V	V	V
Case 32 - Marine stratus	V	V	V
Case 41 - Clear sky on land	V	V	V
Case 42 - Continental stratocumulus	V	V	V
Case 51 - Clear sky on snow ground	V	V	V
Case 52 - Ice clouds on snow ground	V(cirrus)	V(cirrus)	V(cirrus)
Case 61 - Clear sky on land	V	V	V
Case 62 - Po valley Cumulonimbus	V	V	V
Case 7 - MODIS High cloud on land	V	V	X(clear)
Case 8 - MODIS Clear sky on land	V	V	V
Case 9 - MODIS High cloud on ocean	V	V	V

Table 5.3: CIC's classification using only MIR and MIR+FIR channels, according with cloud cover

Cloud covered area in the FSI FoV	MIR	MIR + FIR
1%	clear	clear
3%	clear	clear
6%	clear	cirrus
10%	clear	cirrus
14%	cirrus	cirrus
19%	cirrus	cirrus
24%	cirrus	cirrus
30%	cirrus	cirrus

is evaluated using only the MIR part of the spectrum, and using MIR and FIR channels. In Table 5.3 is shown the CIC classification, according with the cloud cover. The use of FIR channels improves the detection of the simulated cirrus cloud, partially covering the FSI FoV. This result highlights the synergy between FIR and MIR, which allows cloud detection at lower fractional limits than using only the MIR part of the spectrum.

An example of the scene classification impact is shown in Figure 5.12. This figure
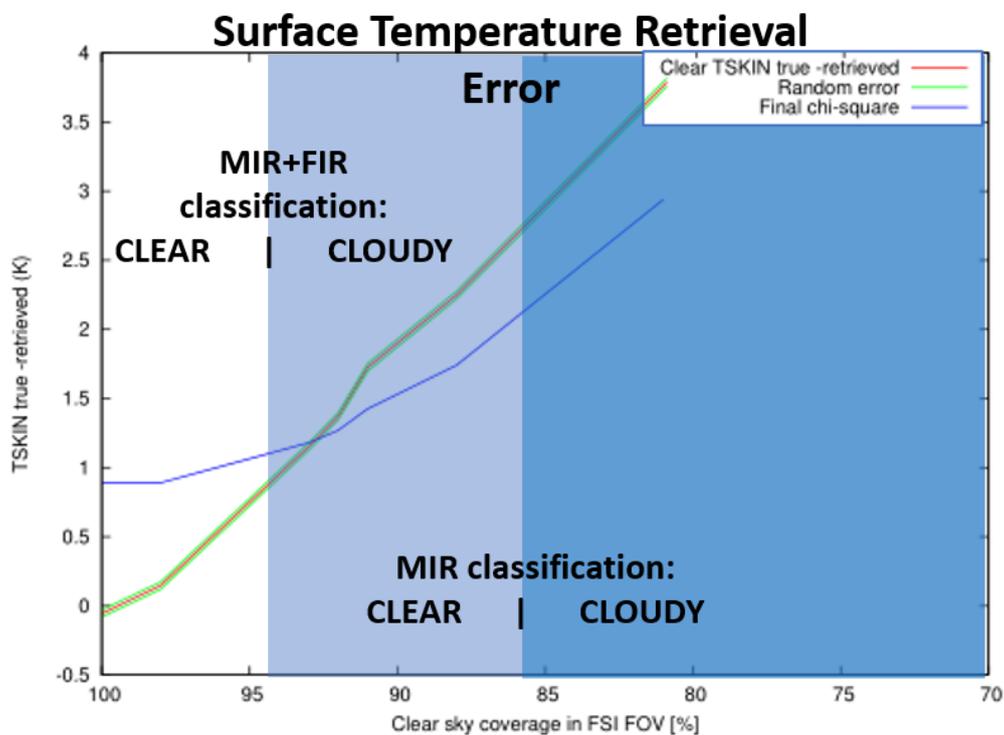
Figure 5.12: Retrieval error of the surface temperature, as a function of the cloud cover. The light blue area indicates the percentage of clear sky area within the FSI FoV, in which a cloud is identified by the CIC algorithm using FIR and MIR wavenumbers. The dark blue area indicates the percentage of clear sky area within the FSI FoV, in which a cloud is identified by the CIC algorithm using either only MIR wavenumbers or FIR and MIR wavenumbers. Figure courtesy of Luca Sgheri.

presents the retrieval error of the surface temperature, as a function of the cloud cover. When the scene is identified as a cloud scene, with the cirrus cloud covering 6% of the FSI FoV, the retrieved surface temperature error is less than 1 K. This error rises as the area covered by the cloud increases. When the cloud is identified using just MIR channels, i.e., the cloud fraction in the FSI FoV is 14%, the retrieved surface temperature error is about 2.6 K. Therefore, the use of the FIR part of the spectrum contributes to identifying the cloud scene, providing also a reduction of the error on retrieval of the atmospheric and surface parameters.

5.4   Summary

In this chapter is made a brief introduction to the FE2ES, which is a tool to predict and assess the Level-1 and Level-2 performances of the FORUM mission. The FE2ES is used to demonstrate the consistency and coherence of the observation concept (including the instrument), the whole processing chain up to Level-2, and the impact of errors in the Level-1 and 2 performance (ESA, 2019). This tool is used to parametric, scientific, and industrial studies.

The simulator can be also used to analyse:

- ■ Specific atmospheric conditions for scientific analysis;

- ■ Sensor performances in specific observational and atmospheric conditions;

- ■ Field campaign data and experimental case studies.

The structure of the SGM and L2M_CIC are presented.

The SGM is used to perform simulations according to the FORUM mission requirements. Inhomogeneities on surface and clouds can be performed to produce synthetic scenes. Cloud parameters can also be set-up automatically, by using ERA 5 reanalysis data, and some parametrization (see Appendix 1). Moreover, the module includes pre-computed study cases based on MODIS data.

The L2M_CIC is a submodule used to identify and classify a scene, by using the CIC algorithm. The submodule contains subroutines to determine the sea fraction in the FSI FoV, and the inhomogeneity level of the scene. Moreover, when a cloud scene is identified, an initial guess of the cloud parameters is provided for the L2M's inversion submodule.

The scenes from the SGM's study cases are used to evaluate the CIC's performance inside of the FE2ES. The results highlight the good CIC's performance, either for cloud detection or for the classification. Furthermore, some simulations are used to understand and quantify the contribution of the FIR part of the spectrum in for the cloud identification, and atmospheric and surface retrievals. The test performed highlight the FIR contribution, since the detection of a cloud covering a lower area of the FSI FoV, reduces the error on the surface temperature retrieved.

In the next chapters are presented the application of the CIC algorithm by using interferometric data in the FIR and MIR collected by airborne campaigns in 2015 and 2019, and to ground based data collected by the REFIR-PAD interferometer since 2012 from the Dome-Concordia station on the Antarctic Plateau.

Chapter 6

Cirrus Cloud Identification from Airborne Far-Infrared and Mid-Infrared Spectra

Currently, very few instruments can observe spectrally resolved radiances in the FIR part of the spectrum and none of them is flying on satellite. Those that do exist include the Radiation Explorer in the Far InfraRed: Prototype for Applications and Development (REFIR-PAD) (Bianchini et al., 2006), the Far-Infrared Radiation Mobile Observation System (FIRMOS) developed, built, tested, and successfully deployed in a field experiment during the FORUM Phase A studies (ESA, 2019), the Far-Infrared Spectroscopy of the Troposphere (FIRST) (Mlynczak et al., 2006), and the Tropospheric Airborne Fourier Transform Spectrometer (TAFTS) (Canas et al., 1997). Some experiments and field campaigns using these instruments have been recently summarized in Palchetti et al. (2020).

The majority of spectrally resolved radiance datasets spanning the FIR consist of downwelling spectra collected from ground-based sites. Therefore, airborne data such as those obtained from TAFTS during the Cirrus Coupled Cloud-Radiation Experiment (CIRCCREX) (Pickering et al., 2015) and in the PiknMix-F field campaign (Met Office, 2019), onboard the Facility for Airborne Atmospheric Measurements (FAAM) are extremely important, since they allow the investigation of radiative properties of the atmosphere and of cirrus clouds from above, with a viewing geometry that is similar to that of a nadir viewing satellite mission such as FORUM. In the context of the FORUM preparatory studies, these kinds of measurements, apart from constituting a unique dataset, provide the opportunity to test innovative algorithms, evaluate their performances and set-up the future operational and data analysis methodologies.

The FAAM flights realized high spectral resolution measurements in the FIR, by using TAFTS, and in the MIR, by using the Airborne Research Interferometer Evaluation System (ARIES) (Wilson et al., 1999). The synergistic exploitation of the two co-located

measurements allows the application of the CIC algorithm, to identify a scene using measured upwelling radiances.

Then, this chapter stands out the results presented in the manuscript from Magurno et al. (2020), searching to evaluate the performance of the CIC algorithm when applied to real upwelling radiance data and to test the additional information content that the FIR part of the spectrum can provide, as suggested by Maestri et al. (2019b).

Initially, the FAAM flights are described along with the relevant instrumentation, measurements, and details of the selection of data to be classified by the identification and classification algorithm (Section 6.1). Next, the methodology to compute the synthetic dataset (Section 6.2) is described, followed by the description of the training and test set used by the CIC algorithm (Section 6.3). Afterwards, the results and discussion are presented (Section 6.4), and the main conclusions of this chapter are summarised in Section 6.5.

6.1 Instruments and Data Sample

6.1.1 TAFTS

TAFTS is a Martin-Puplett polarising interferometer designed to optimise sensitivity in the spectral range 80-600 cm^{-1} and has been deployed in numerous field campaigns on board aircraft and on the ground (e.g. Cox et al., 2010; Green et al., 2012; Fox, 2015; Bellisario et al., 2017). This four-port system has two input ports, associated with the zenith and nadir views and two output ports. The inputs are superposed using a polarising grid, positioned in the optical chain, on the input side of the polarising beam-splitter. One input port in reflection from the combining polariser, is associated with the zenith view, and the other, in transmission, is the nadir view. The two output ports are separated using a polarising analyser. Pairs of detectors at each output measure the interferograms associated with each of the two input ports: these modulated signals are in anti-phase and the resultant spectral signal is therefore a difference of the two inputs. To alleviate microphonic induced vibrations on the optics the interferometer is operated under a vacuum.

The optical efficiency of the two TAFTS input ports is not identical and each requires separate calibration. Each input port therefore consists of separate pairs of calibration blackbodies at hot and ambient temperature, with separate steering mirrors enabling switching of the view between the blackbodies and the scene in a predefined sequence. These calibration optics are open to the ambient air and are external to the evacuated interferometer section. To derive calibrated radiances a series of calibration measurements are made viewing the zenith and nadir calibration targets, usually on a 5-6 min repeat cycle during which between 1-4 min may be allocated to zenith and nadir observations.

To enhance detectivity, minimise acquisition time and hence enhance spatial sampling

TAFTS employs liquid helium cooled photoconductive detectors. GeGa (Gallium doped Germanium) detectors cover 80 to 300 cm^{-1} and SiSb (Antimony doped Silicon) detectors cover 320 to 600 cm^{-1}. In each output the two waveband regions covered by the detectors are separated using a dichroic filter with a cut-off at about 330 cm^{-1}. TAFTS is usually operated at a spectral resolution of 0.12 cm^{-1}, but for these flights this was reduced to 0.24 cm^{-1}. For CIRCCREX, this was achieved by spectral averaging. For PiknMix-F the maximum optical path of the interferometer scan mirror was reduced to improve signal-to-noise and increase the acquisition rate, from about 1 scan every 2.5 s to one scan every 1.5 s.

To calibrate TAFTS spectra, the instrument gain and offsets over the range of calibration cycles associated with the nadir observations under consideration are compared. Over the flight periods analysed here no significant change in the gain or offsets were identified. The random noise level of the calibration runs was reduced by averaging and then combined with the uncertainties associated with knowledge of the black-body temperatures to give the overall calibration error.

For both campaigns, TAFTS covered the spectral range 80-300 cm^{-1} and 320-540 cm^{-1}.

6.1.2   ARIES

ARIES (Wilson et al., 1999) is based around a rugged commercial interferometer fabricated by BOMEM of Canada, and was specifically designed for aircraft deployment as an airborne demonstrator and for future validation studies for the then proposed MIR satellite sounders such as Infrared Atmospheric Sounding Interferometer (IASI) (Klaes et al., 2007) and Atmospheric InfraRed Sounder (AIRS) (Chahine et al., 2006). The MIR spectral coverage of ARIES allows the employment of solid optical components that minimise sensitivity to vibration effects from the environment.

ARIES, also a four-port system, utilises a KBr (potassium bromide) beam-splitter acting as a division of amplitude device and is sensitive over the spectral range from approximately 550 cm^{-1} to 3000 cm^{-1}. For ARIES, the two input and two output ports are physically off-set. An internal reference blackbody of known temperature is positioned at one input. The other input port views the external scene via a window and steering mirror in the pointing head. This pointing mirror allows alternate views of calibration targets, used to derive the instrument spectral gain and offsets. This gain and offset are applied to the external nadir and zenith scene measurements to derive calibrated radiances referenced to the internal blackbody. The calibration-view scene cycle for ARIES can be changed during flight and was set to nadir for above cloud periods during the CIRCCREX flight. TAFTS was set to a fixed observation cycle and acquired fewer above cloud nadir measurements. There was no synchronisation between the two instruments resulting in poor temporal co-location of measurements. For PiknMix-F a greater proportion of TAFTS observations were dedicated to the nadir with improved

co-temporal measurements between TAFTS and ARIES.

ARIES uses two detectors to cover the MIR. In this case, the detectors are Mercury Cadmium Telluride (MCT) covering 550 to 1800 cm^{-1} and Indium Antimonide (InSb) covering 1700 to 3000 cm^{-1}. These detectors are mounted in a sandwich and are cooled to approximately 80 K by a Stirling cooler. Only a single output is used for the acquisition of spectra with the second unused output directed towards the interferometer input window. ARIES was operated at its maximum resolution of 1.0 cm^{-1} for both flights considered here, with individual scans acquired every 0.25 s.

Each ARIES nadir observation is calibrated using a set of calibration views made at the closest time to that observation. Calibration drift during a nadir observation cycle is established using information from the calibration views before and after the nadir observations. The random noise in individual views can be estimated from the standard deviation in the calibration views which consist of 120 spectra over a 30 s period. The calibration drift and random noise are treated separately and then combined to derive the overall uncertainty.

For both campaigns, ARIES covered the spectral range 600-2800 cm^{-1}.

6.1.3 FAAM Flights and Data Selection

Two sets of airborne measurements were collected, for both ARIES and TAFTS, during two independent experimental campaigns in 2015 and 2019.

Firstly, data from the CIRCCREX (Fox, 2015; Pickering et al., 2015) that comprises of several field campaigns are used. Here, a single flight from March 13th 2015 is considered (flight B895). The FAAM aircraft overflew a decaying band of cirrus clouds associated with an occluded front over the North Sea, off northern Scotland. 63 TAFTS spectra and 2729 ARIES spectra were collected during three Straight and Level Runs (SLRs) above the cloud deck, at about 9.4 km altitude, between 09:33 and 10:28 UTC. Six dropsondes were dropped to characterize the atmospheric column below the aircraft, and a LIDAR Leosphere ALS450 onboard the aircraft provided backscatter data, allowing the vertical extent and the optical depth at 355 nm of the observed cirrus clouds to be characterised.

The PiknMix field campaigns have been on-going for a number of years with the most recent, PiknMix-F, occurring in March 2019. These cloud physics and radiation campaigns are aimed at a variety of atmospheric states, providing measurements to improve numerical weather prediction and climate models. Here, data of the single flight C153 are used, from PiknMix-F on 13th March 2019. The FAAM aircraft overflew the North Sea in clear sky conditions, off eastern Scotland. 162 TAFTS spectra and 2399 ARIES spectra were collected during two SLRs at about 8.8 km altitude, between 12:30 and 12:55 UTC. Eight dropsondes were dropped during the flight. In Figure 6.1 is shown the tracks of flights B895 (black) and C153 (red) and the corresponding SLRs.

The timing of each spectrum is recorded as part of the instrument house-keeping for both TAFTS and ARIES and this information is used to co-locate TAFTS and ARIES
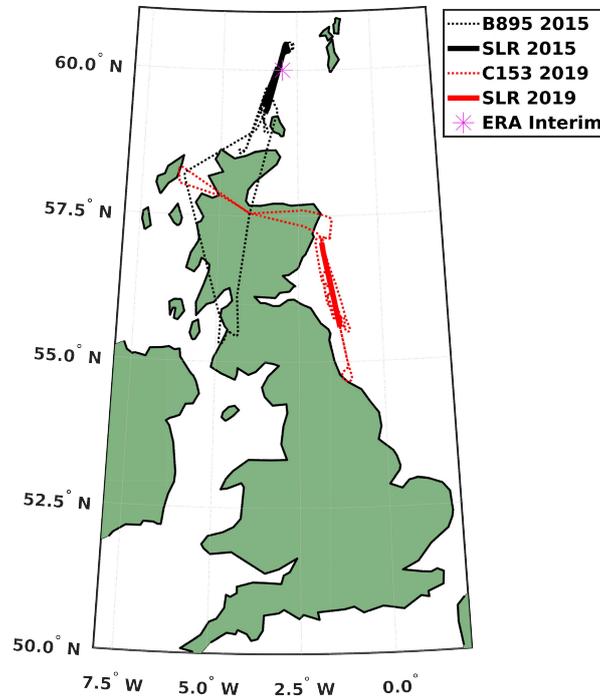
Figure 6.1: Flight tracks (dotted lines) and straight level runs (solid lines) of the two experimental campaigns described in the text. Cirrus cloud data have been collected during flight B895 (black) while clear sky data have been collected during flight C153 (red). Synthetic spectra are computed from ERA-Interim data referred to the location marked by the magenta asterisk.

observations. The selection criterion adopted requires that TAFTS and ARIES spectra are collected within few tenths of a second of each other to be ascribed to a given scene, corresponding to an average distance of the centre of the sensors' field of view on the ground of about 10 m, and never higher than 100 m. This constraint limits the final samples to 30 cloudy cases from flight B895 and to 90 clear sky cases from flight C153. As a result, a dataset of radiances spanning the IR spectrum from FIR to MIR is available and thus it is possible to apply the classification algorithm in different configurations, accounting for multiple spectral intervals. Figure 6.2 shows the temporal coverage of the observations for the two instruments (black marks). The red symbols mark TAFTS and ARIES simultaneous measurements. The aircraft altitude during the measurements is also highlighted with the blue track in the upper part of the panels.
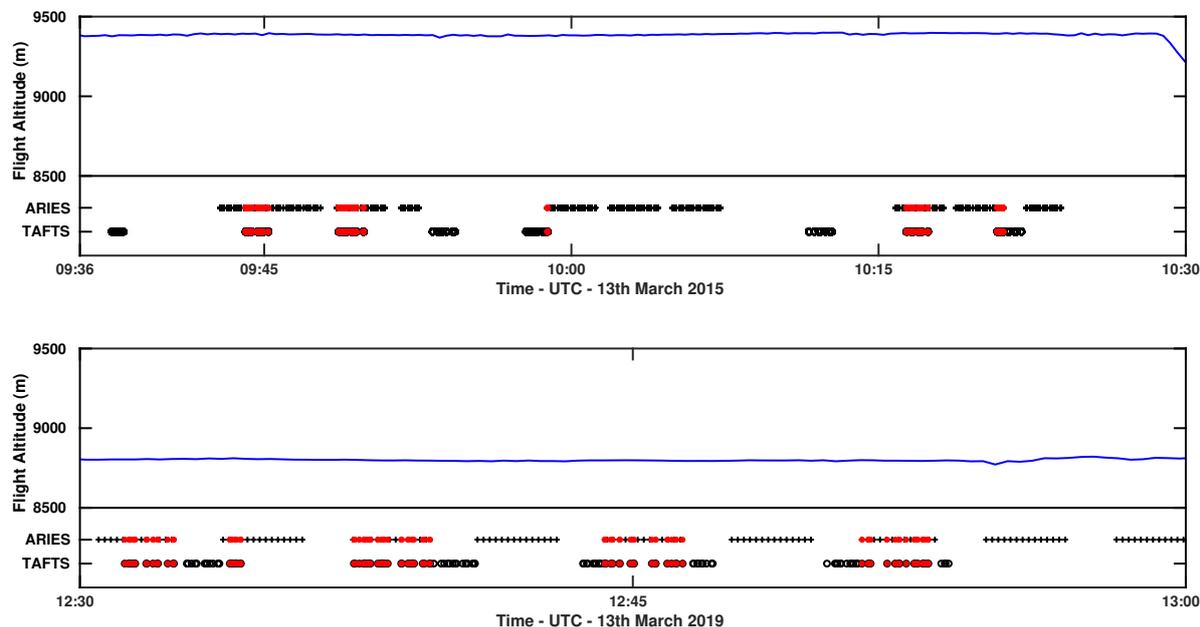
Figure 6.2: Tropospheric Airborne Fourier Transform Spectrometer (TAFTS) and Airborne Research Interferometer Evaluation System (ARIES) observation timestamps (black marks) for the Facility for Airborne Atmospheric Measurements (FAAM) flights B895 (top panel) and C153 (bottom panel). The red asterisks mark TAFTS and ARIES simultaneous measurements. Blue solid lines show the flights altitude during data collection.

6.2   The Synthetic Datasets

The number of TAFTS and ARIES colocated observations is limited to a total of 120 cases. In addition, 60 simulated cases are computed to expand the dataset in order to generate more robust statistics when testing the classification approach. Synthetic spectra in cloudy or clear sky conditions are produced to encompass a wider range of atmospheric states than those encountered during the two days of measurements of the field campaigns. These represent a better estimate of the upwelling radiance in clear sky conditions or in the presence of thin cirrus clouds for the time of the year and locations of the CIRCCREX and PiknMix-F experiments.

To define a synthetic dataset consistent with the field campaign data and, at the same time, representative of the natural variability of the experimental area, a radiance generator code, based in a methodology similar to that used for SGM, is developed.

In this case, data from ERA-Interim reanalysis are used. Considering a grid point over the sea, at 60°N-3°W (magenta asterisk in Figure 6.1), the radiance generator code is used to compute 30 synthetic radiances in clear sky conditions and 30 in cloudy conditions, spreading over the months February-March-April of the years 2015 and 2019, at different synoptic hours (00:00, 06:00, 12:00, or 18:00 UTC).

Table 6.1: Cloud parameter ranges used to define the cirrus cloud layers for the synthetic spectra.

Parameter	Min	Max	Units
Top Height	5.1	9.5	km
Geometric Thickness	0.2	2.8	km
Optical Depth (900 cm^{-1})	0.2	2.0	
Effective Dimension	2	40	µm

The simulations are made by using the LBLRTM model v12.7, with the water vapour continuum parametrization MT_CKD v3.2 and the AER v3.6 line parameter database, and the LBLDIS v3.0, which combines the LBLRTM and DISORT models for scattering atmospheres.

The clear sky spectra are computed with LBLRTM to cover the TAFTS+ARIES spectral range, from 50 to 2850 cm^{-1}, with spectral resolution 0.01 cm^{-1}, fixing the observer altitude at 9.5 km, which is the maximum altitude of the flights. The cloudy sky spectra are computed with LBLDIS accounting for cloud layers with parameters characteristic of cirrus clouds (Veglio and Maestri, 2011). A summary of the cloud properties is reported in Table 6.1.

To run the radiance generator, auxiliary information is needed to define the surface parameters, the atmospheric thermodynamic state, and the cloud properties. Specifically:

(a) Since the FAAM flights measurements are taken over the sea, the surface emissivity is assumed to be that of sea water. A sea water in calm wind conditions (Masuda et al., 1988) is accounted for, in the range 10-3000 cm^{-1} at 5 cm^{-1} spectral resolution. The sea surface temperature is derived from the geolocated ERA-Interim archive (Dee et al., 2011), publicly available at 0.75° horizontal resolution.

(b) The ERA-Interim archive data are also used to define the vertical profiles of temperature, pressure, water vapour and ozone concentration on specific dates. However, since the dataset is limited to about 50 km altitude, the vertical profiles are completed with the climatological database IG2 v5.7 (Remedios et al., 2007) up to 0.01 hPa. The IG2 is also used to define the entire volume mixing ratio profile of the same 10 more minor gases from SGM (CO_2, N_2O, CO, CH_4, O_2, NO, SO_2, NO_2, NH_3, and HNO_3). The CO_2 mixing ratio, estimated by IG2 in 2013, is linearly corrected to the 2018 concentration level with a scaling factor. All the profiles are defined with a vertical resolution that gradually increases from ground (200 m) to top of the atmosphere (5000 m) placed at 60 km as usually done for satellite view simulations.

(c) The cloud microphysical properties are generated for aggregates of pristine columns with a smooth roughness by integrating, over a large set of gamma type size distributions, the database by Yang et al. (2013), that contains single particle single scattering
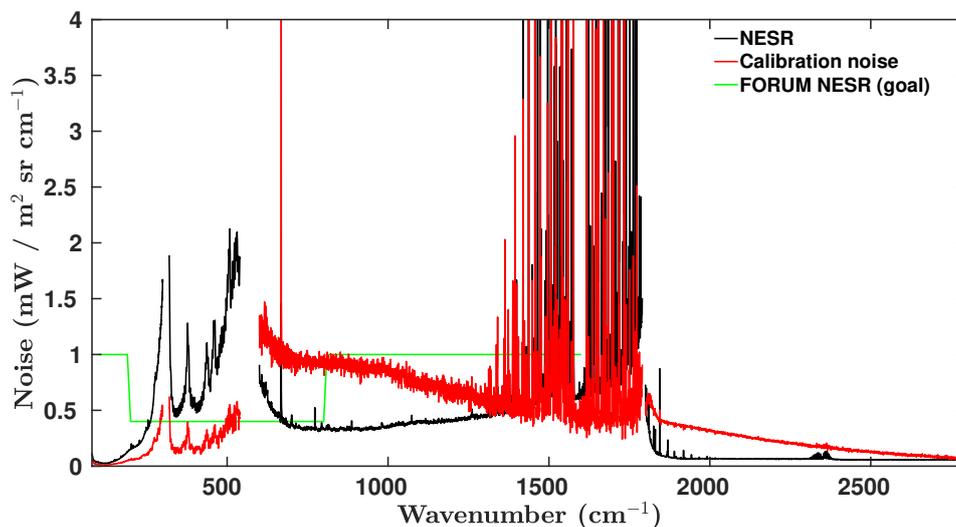
Figure 6.3: Calibration noise (red) and noise equivalent spectral radiance (NESR) (black) associated to the TAFTS and ARIES observations. The Far-infrared Outgoing Radiation Understanding and Monitoring (FORUM) NESR (in light green) is also reported for reference.

properties of randomly oriented non-spherical ice crystals.

The synthetic high spectral resolution radiance fields are convolved with the appropriate spectral response function associated with the respective instrument apodisation functions, a Kaiser and a top-hat function, to mimic the TAFTS and ARIES instrument response functions respectively.

Finally, a component related to the total measurement uncertainty for TAFTS and ARIES (see details in Sections 6.1.1 and 6.1.2) is added to the spectra. The total uncertainty for each of the two instruments is computed, for any synthetic spectrum, as the sum of the calibration uncertainty (red line in Figure 6.3), assumed to be systematic for all the spectra, and a white Gaussian noise whose 1σ corresponds to the tabulated NESR at each wavenumber (black line in Figure 6.3). In the same figure the FORUM NESR goal (ESA, 2019) is also reported for comparison. The comparison shows that FORUM will have better performance, with respect to the analysed observations, in most of the FIR spectral range, in the CO_2 ν_2 band and above 1300 cm^{-1}. The measurement uncertainties (Figure 6.3) become sizeable at wavenumbers larger than 1300 cm^{-1} so these data are not used in the classification.

6.3 The Training Set and the Test Set

The synthetic spectra, described in the previous section, are used to define the TSs needed by the CIC to perform the classification of the observed scenes. However, to

Table 6.2: Number of synthetic and observed spectra in the training and test sets.

	Synthetic	Observed	Total
TS CLEAR	30	6	36
TS CLOUD	30	6	36
TEST SET CLEAR/CLOUD	0/0	84/24	108
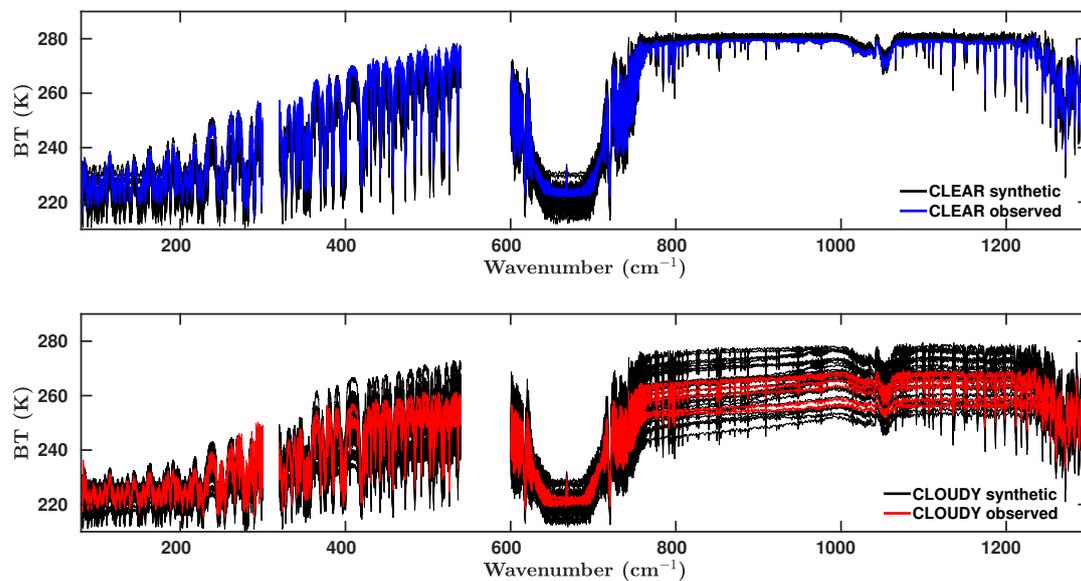
Figure 6.4: Synthetic (black) and observed (color blue or red) brightness temperature (BT) spectra in the TSs for clear (top) and cloudy (bottom) sky. The gaps in the spectra separate, from left to right, TAFTS long wave, TAFTS short wave, and ARIES long wave spectral ranges.

improve the representativeness of the considered scenarios, few observations provided by the FAAM flights B895 (cloudy) and C153 (clear) are merged into the TSs. The TSs are thus composed of few observed spectra and of all the synthetic ones while the test set comprises the rest of the observations. See Table 6.2 for a summary.

Figure 6.4 shows the spectral BT of the clear (top panel) and cloudy (bottom panel) TS elements. Synthetic spectra are plotted in black while observations are blue (clear) or red (cloudy). The variability is larger across the cloudy spectra than for the clear sky cases for both observed and simulated data. However, the variability is markedly larger in the simulations compared to the observations in both scenarios. In particular, the largest variations in BT are encountered in the atmospheric window between about 800 and 1200 cm^{-1} and at FIR wavenumbers larger than 350 cm^{-1}.

6.4   Results

6.4.1   Classification Algorithm Set Up

The performance of the CIC algorithm depends on the TS characteristics and the spectral intervals of application. Assuming that the TSs are good representatives of the variability of the considered scenarios, and seeking to optimize the CIC algorithm performance, the classification of the test set is made by applying the *distributional approach*.

As previously noted, the clear and cloudy training sets have been set up by using a limited number of observations (6) and 30 synthetic spectra, in order to mimic the variability of atmospheric and sea surface conditions for the locations and times of year of the flights. Three different spectral intervals are considered in this study to evaluate the ability of TAFTS and ARIES to identify clear and cloudy scenes and to assess the advantages of using FIR and MIR spectral radiances synergistically. The wavenumber limits of the intervals considered are summarised in Table 6.3 and described in detail below.

1. The *TAFTS* interval covers radiances in the 80-540 cm^{-1} band, with a gap between 300 and 320 cm^{-1}. It only includes the FIR spectral range, as observed by TAFTS.

2. The *ARIES* interval (600-1300 cm^{-1}) covers a key band of the MIR spectral range. This is a commonly used spectral range, since most of the available infrared sounders sense the atmosphere in the MIR (e.g., AIRS: Chahine et al. (2006); IASI: Lerner et al. (2002); Cross-track Infrared Sounder (CrIS): Bloom (2001)). As explained in Section 6.2, the ARIES sensor performed with a low SNR in the 1300 cm$^{-1} < \tilde{\nu} < 2800$ cm^{-1} band and therefore radiances in this part of the spectrum are discarded. Moreover, the strong 667 cm^{-1} CO$_2$ absorption band is limited to wavenumbers lower than 620 cm^{-1} and higher than 667 cm^{-1}; this was observed in Section 4.2 to increase the CIC performance likely due to the removal of correlated information.

3. The *OPTIMAL* interval (320-1300 cm^{-1}) is chosen to mimic as closely as possible the interval 300-1300 cm^{-1} selected in Section 4.2 as a very efficient infrared interval for the CIC application to spectral radiance fields simulating FORUM measurements over the globe that will span the 100-1600 cm^{-1} spectral band. However, since TAFTS measurements do not cover wavenumbers between 300 and 320 cm^{-1} or higher than 540 cm^{-1}, the range used here is 320-1300 cm^{-1}, with a gap between 540 and 600 cm^{-1}. Moreover, as in the previous *ARIES* case, the 620-667 cm^{-1} band is not considered. The OPTIMAL interval is considered to evaluate the advantage of fully exploiting the MIR and the FIR synergy, and to account for radiance coming from both sensors.

Table 6.3: Spectral intervals of application of the CIC algorithm.

Interval Name	Initial Wavenumber (cm^{-1})	Final Wavenumber (cm^{-1})
TAFTS	80	540
ARIES	600	1300
OPTIMAL	320	1300

Table 6.4: Hit rate scores of the CIC algorithm.

Interval Range * (cm^{-1})	Scene	HR
TAFTS (80-540)	CLEAR	0.46
	CLOUDY	1.00
	TOTAL	0.58
ARIES (600-1300)	CLEAR	1.00
	CLOUDY	1.00
	TOTAL	1.00
OPTIMAL (320-1300)	CLEAR	1.00
	CLOUDY	1.00
	TOTAL	1.00

* See text for gaps in the intervals.

6.4.2 Clear and Clouds Identification

All classification results can be assessed by defining a quality flag for the CIC performance. In this case, the statistical measure named *sensitivity* or Hit Rate (HR) is used, in accordance with the following definition:

$$HR(i) = \frac{TP(i)}{TP(i) + FN(i)} \tag{6.1}$$

where i is the identifier of the set (CLEAR, CLOUDY or TOTAL), $TP(i) + FN(i)$ is the number of analysed scenes, $TP(i)$ is the number of correctly classified scenes, and $TOTAL = CLEAR + CLOUDY$. The results of the classification, for the three wavenumber intervals considered (TAFTS, ARIES, OPTIMAL), are listed in Table 6.4 and shown in Figure 6.5 in terms of the SID parameter, which allows a visual inspection of the classification outcomes.

When the FIR spectral interval (*TAFTS*) was used alone, the overall classification performance was poor, having a total HR limited to 58% (first panel of Figure 6.5). Part of the clear sky spectra were misclassified, and the clear and cloudy spectra partially overlap in terms of the SID values. This interval was not expected to provide an accurate
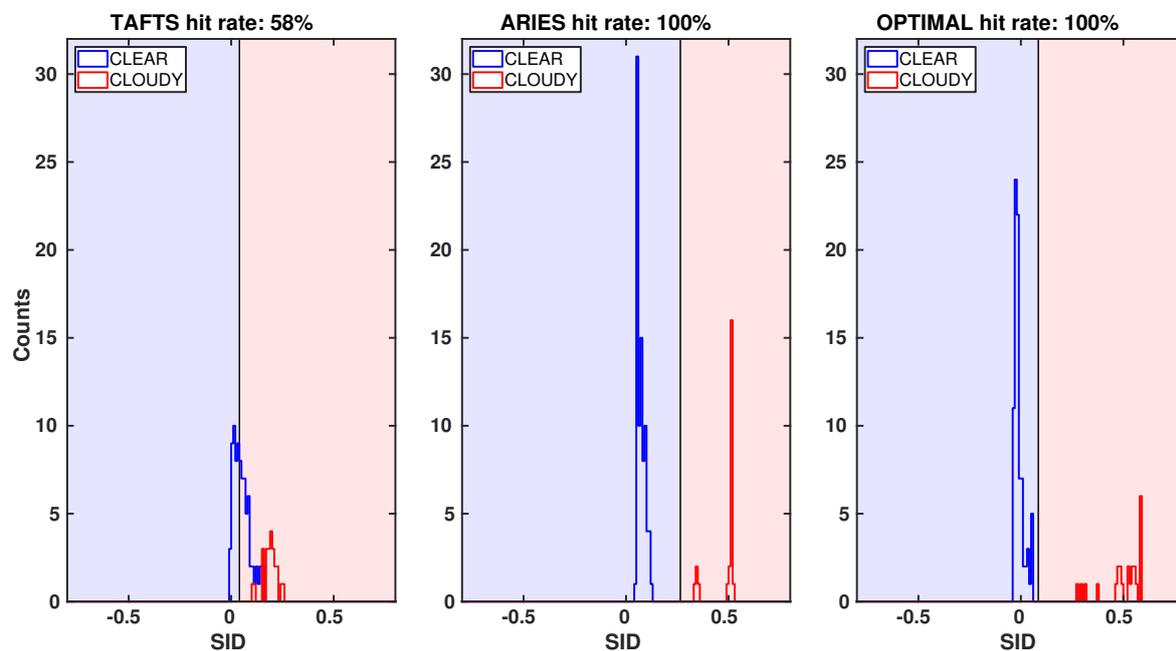
Figure 6.5: CIC classification results on the observed FAAM flights spectra. CIC is run, from left to right, over the intervals *TAFTS* (80-540 cm^{-1}), *ARIES* (600-1300 cm^{-1}), and *OPTIMAL* (320-1300 cm^{-1}). Histogram elements within the blue/red shaded area are classified as clear/cloudy respectively.

classification of the scenes due to the limited amount of information coming from the surface and the lowest atmospheric levels in the FIR. In fact, the upwelling radiance in this spectral region shows large sensitivity to water vapour (especially in the highest tropospheric levels) and to cirrus clouds properties.

The MIR spectral interval (*ARIES*) alone had greater information content in terms of cloud classification than the FIR alone, and, in the present case, the entire set of observed spectra was correctly classified (HR = 100%). The second panel of Figure 6.5 shows that the clear sky dataset (blue histogram) was classified with very similar SID values for the different spectra; this is expected since the observations were collected in clear sky over the same marine area within less than 30 min, and the spectra are all representative of very similar conditions. On the other hand, the cloudy dataset (red histogram) has higher, even if limited, variability due to the clouds' characteristics, that affect the classification results producing slight variations in the SID values and displaying as two distinct histogram elements. The first group of four cloudy spectra, corresponding to the lower SID values, refers to the last four measured spectra, in chronological order, of the analysed dataset. This suggests that different features, present within the cloud layer during the time of observation, are identified in the classification results.

Finally, the synergy between MIR and FIR data was evaluated by applying the CIC

over the *OPTIMAL* interval. Despite the limited information content from the FIR alone, in Chapter 4 is possible to note that its contribution is significant when used in combination with the MIR. Unfortunately, due to the limited variability of the observed dataset, it is not possible to appreciate significant differences between the *OPTIMAL* and the *ARIES* interval, since both have a HR equal to 100%. However, the last panel of Figure 6.5 shows that the SID values for the clear sky spectra are slightly lower than in the *ARIES* case and spread over a slightly larger interval of SID values, indicating that the CIC is capable of detecting greater variability among the clear spectra analysed. The cloudy spectra provide almost the same SID values as in the *ARIES* case but again with a wider distribution over SID values, suggesting the greater sensitivity of the *OPTIMAL* interval. The distribution of the SID values reflects the variability within the cloud layer: the first group of spectra corresponding to the lowest SID values refers to the last spectra of the dataset, in chronological order; the six spectra with the highest SID values refer to the first collected spectra during the observation. Despite the limited sample, this highlights that the use of the FIR part of the spectrum enhances the sensitivity of the CIC algorithm.

6.4.3 Thin Clouds Detectability

An additional analysis is performed to investigate the capability of the algorithm to detect very thin cirrus clouds when the same experimental conditions are considered. To facilitate this, a set of synthetic spectra were generated for a cirrus cloud layer 400 m thick, with cloud top at different heights between 5.1 and 9.1 km, and varying the optical depth (OD) at 900 cm^{-1} in the range 0.03-1. The ice particle effective dimension is 40 µm. All simulations account for the TAFTS and ARIES noise as previously described. The results are summarized in Figure 6.6 that shows the minimum OD value allowing cloud detection using the CIC algorithm, as a function of cloud top height.

All spectra obtained from clouds with similar features (i.e., altitude, geometrical thickness, effective particle dimensions, etc.) but larger OD than the cloud OD limit shown in Figure 6.6 were well classified. Thinner clouds are misclassified and the corresponding spectra are incorrectly tagged as clear sky. The result that defines a lower limit of OD detectability as a function of cloud altitude, cannot be assumed as general since it only applies to the specific atmospheric and cloud properties considered here.

The *OPTIMAL* range (black solid line) performs better than the *ARIES* range (MIR only, red dashed line), allowing the detection of 50-60% thinner clouds, in the present conditions. The thinnest clouds detected by exploiting the *OPTIMAL* and the *ARIES* spectral ranges have OD = 0.09 and OD = 0.18 respectively. This result suggests that the FIR contains useful information to improve upon the detection of a thin cloud. The result also confirms the findings in Chapter 4 regarding simulations of the future FORUM mission. This aspect is not fully appreciated in the analysis of the observed airborne dataset since the estimated ODs of the detected clouds are all higher than about 0.3
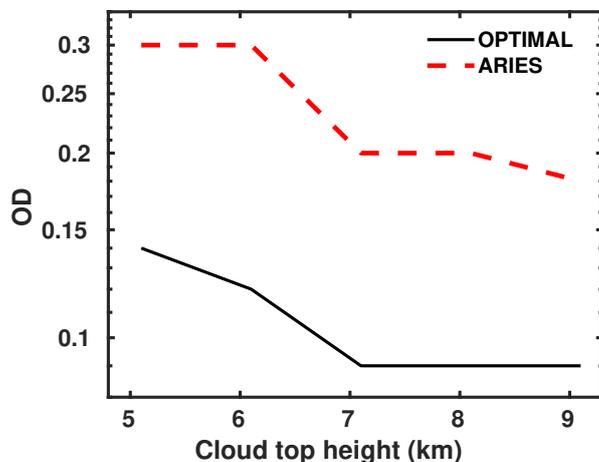
Figure 6.6: Minimum cloud optical depth for which the CIC, applied using different wavenumber intervals, classifies the scene as cloudy. The cloud is computed with a fixed geometrical thickness of 400 m and cloud ice particle effective dimensions of 40 μm. Cloud top altitude is varied as reported on the x-axis.

(internal communication). For such OD values the CIC correctly classifies cloud spectra either using the *ARIES* or the *OPTIMAL* spectral range. Nevertheless, the exploitation of a spectral interval from FIR to MIR is expected to be of great value in view of the global observations that will be performed by the FORUM sensor.

As expected, the cloud detection limit increases (OD decreases) with increasing cloud altitude. This is because, for the same cloud OD, an increase in cloud altitude implies a decrease of the cloud temperature (and thus an increase of the contrast with respect to a clear sky spectrum) and an increase of the transmissivity of the atmospheric layers between the cloud and the sensor (due to a decrease in water vapour content), increasing the radiance sensitivity to cloud properties.

To highlight the impact of the improvement in the scene classification performed by the *OPTIMAL* interval, Figure 6.7 compares three synthetic spectra used in the sensitivity test: the spectrum computed in clear sky conditions (blue line); the cloudy spectrum, with cloud top at 9.1 km, corresponding to the thinnest detected cloud in the *OPTIMAL* (OD = 0.09, black line) and *ARIES* (OD = 0.18, red line) interval. BT differences with respect to the clear sky spectrum are shown in the bottom panel of the same figure where the spectral noise associated with the simulations is also plotted (in light blue). The importance of using the *OPTIMAL* interval, instead of the MIR interval only, for improving the cirrus clouds identification, can be demonstrated by a simple example on the estimate of the error produced by a misclassification on retrieved quantities such as the sea surface temperature that is usually retrieved from radiances in the 800-1200 cm^{-1} spectral range. For the analysed conditions, the classification using the *ARIES* interval misclassifies spectra with clouds with OD ≤ 0.18, represented by a
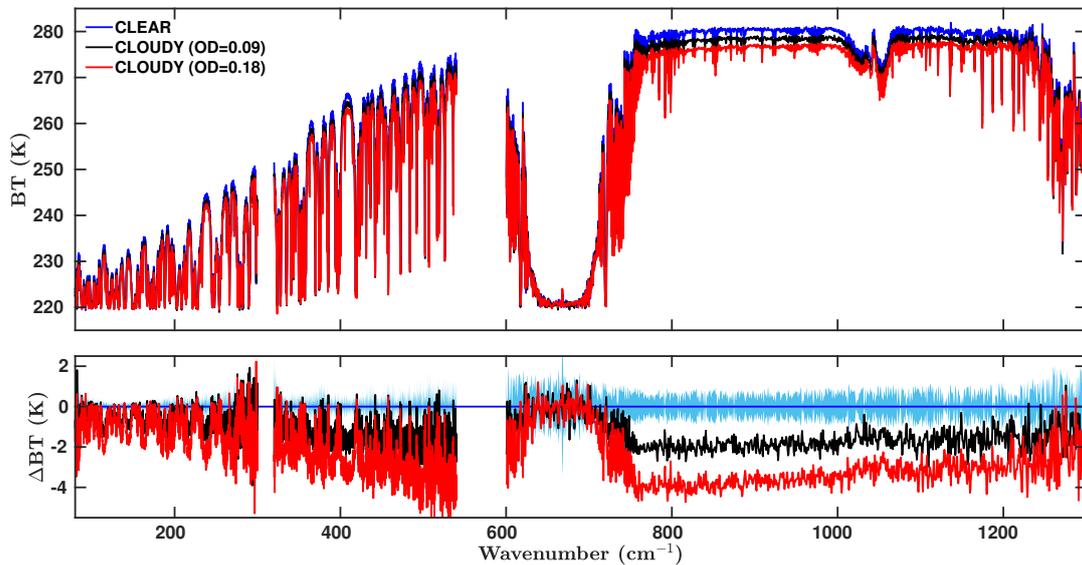
Figure 6.7: Upper panel: BT synthetic spectra of a clear sky (blue) and two cloudy skies (black/red) in the same atmospheric conditions. The clouds are placed at 9.1 km altitude, are 400 m thick, with an effective particle dimension of 40 µm, for two different optical depths (ODs) that are representative of the detection limits of CIC using the MIR and FIR (OD = 0.09, black line) or the MIR only (OD = 0.18, red line). Bottom panel: BT differences with respect to the clear sky spectrum for the two ODs. The noise level associated with the simulated spectra is also plotted in light blue around the zero line.

BT difference, in the 800-1200 cm^{-1} range, of up to 3.5 K with respect to the clear sky. When the *OPTIMAL* interval is used the misclassification occurs in presence of thin cirri with OD $\leq$ 0.09, limiting the BT error to less than 2 K in the same spectral interval.

6.5 Summary

The CIC algorithm is applied, for the first time, to a dataset of airborne high spectral resolution infrared measurements in order to evaluate the capability of the algorithm to identify clear and cloudy sky spectra. In Chapter 4, the algorithm was tested only on synthetic datasets, therefore, its results on real data are a fundamental step forward the algorithm exploitation in operating environment. Data are collected by two different sensors (TAFTS and ARIES) that measured up-welling radiances in the FIR and MIR part of the spectrum. Two training sets are built by using a set of synthetic spectra that account for instrument noise (calibration and NESR) and capture the atmospheric conditions associated with the locations and times of the airborne measurements. A small number of observations are also added to the training sets.

The classification results show that employing only the FIR spectral region in the CIC

algorithm yields low overall hit rates, the statistical measure used here to evaluate the classification success. In contrast, when the MIR spectral region is used all the spectra are correctly classified. The same perfect hit score is obtained for a combination of far- and mid-infrared channels. However, detailed analysis of the classification parameter SID shows that the use of both the far- and the mid-infrared part of the spectrum provides a wider sensitivity to the variability of the observed spectra.

The benefit of exploiting FIR radiances for cirrus cloud detection may be particularly marked for optically thin cases. The analysed airborne measurements dataset does not include very thin clouds. Nevertheless, simulations show that the use of a combination of TAFTS and ARIES data would provide better classification scores with respect to using only one of the sensors. Specifically, it is shown that the synergistic use of far- and mid-infrared radiances increases the detectability of thin cirri (with respect to using the mid-infrared part of the spectrum only) and, for the specific observational conditions considered, the OD detection improves from about 0.2 to about 0.09 for cirri with cloud top altitude above 7 km. The simulation results also show that the ability to detect thin clouds increases with cloud altitude due to a lesser impact of the atmospheric gases (i.e., water vapour) on the spectra and to an increase in contrast with background radiation from sea surface.

Note that the CIC algorithm only relies on the information contained in the spectra that are selected for the training sets. This highlights its applicability very easy and adaptable to other types of sensors since no further ancillary or model data are required. Moreover, the CIC analyses the spectral variation of the radiance in key intervals of the spectrum instead of relying on methods accounting for brightness temperature differences (thresholds) at specific wavenumbers that again require accurate knowledge of the surface and lower atmosphere when applied to detect cirrus clouds. For the present case, only two training sets were required (clear and cirrus clouds) but others accounting for mid-level ice clouds, water clouds and aerosol layers can be considered for application to global observation as those foreseen from the FORUM mission.

Chapter 7

Ice and Mixed-Phase Cloud Statistics on Antarctic Plateau

The polar regions present several challenges for meteorology and climatology studies (Walsh et al., 2018). These regions are crucial components of the ERB (Liou, 2002; Kiehl and Trenberth, 1997), since they generally emit more energy to space in the form of infrared radiation than what is absorbed from sunlight, thereby behaving as heat sinks. Modelling studies have shown that changes in cloud properties (e.g., cloud amount, cloud thermodynamic phase, cloud height, cloud optical thickness) over Antarctica may impact different regions in the globe, highlighting the importance of Antarctic clouds for the global climate system (Lubin et al., 1998). However, obtaining measurements of cloud properties in the interior of the Antarctic continent is still a challenge (Town et al., 2005; Lachlan-Cope, 2010; Bromwich et al., 2012). Observations from synoptic weather stations require an experienced observing staff and sometimes become unavailable during 'white-out' conditions caused by blowing snow. Analysis of satellite measurements from both active and passive sensors must account for a number of problems in inferring the cloud properties. One issue is that the cloud radiative properties tend to be very similar to those of the background (the snow/ice surface). Optically thin cirrus clouds are often present in the Antarctic Plateau (King and Turner, 1997) but are difficult to identify and analyse due to their small cloud signals (the difference between cloudy and clear-sky radiances). Measurements become problematic during the long polar night (King and Turner, 1997), and some stations reduce the observing frequency in the winter time (Bromwich et al., 2012). Observations at solar wavelengths are not available for about half of the year, thus reducing the overall ability to recognize the presence of cloud layers and to derive their physical and optical features. Measurements at longer wavelengths (i.e. in the infrared, or IR) are available regardless of solar illumination but it is often the case that the cloud top temperature is similar to the ice surface temperature (King et al., 1992; King and Turner, 1997; Bromwich et al., 2012).

Active remote sensing techniques have been very helpful in overcoming the limita-

tions of the passive instruments in polar regions. Adhikari et al. (2012) investigated the seasonal and interannual variabilities of the vertical and horizontal cloud distributions over the southern high latitudes poleward of 60°S, using observations from CloudSat and Cloud-Aerosol Lidar and Infrared Pathfinder Satellite Observation (CALIPSO) satellites between June 2006 and May 2010. They found that the Antarctic Plateau has the lowest cloud occurrence of the Antarctic continent (< 30%). The sensors on board of the aforementioned satellites have been also used to investigate macro and microphysical Antarctic cloud properties (Verlinden et al., 2011; Adhikari et al., 2012; Listowski et al., 2019; Ricaud et al., 2020). Nevertheless, satellite active sensors are not lacking in problems when used for cloud detection in Polar regions. For example, Chan and Comiso (2011, 2013) discuss the difficulties encountered by either the Cloud Profiling Radar (CPR), on Cloudsat, and the Cloud-Aerosol Lidar with Orthogonal Polarization (CALIOP), on CALIPSO, in detecting low-level clouds in the Arctic. The difficulties arise from the CloudSat coarse vertical resolution (about 500 m) and its limited sensitivity (low signal-to-noise ratio) near the surface, and in the case of CALIOP are due to the geometrically thin nature of the cloud and its surface proximity. Bromwich et al. (2012) present a review on Antarctic tropospheric clouds. They discuss the instruments and methods to observe Antarctic clouds and the current data sets available. The authors highlight that there are relatively few measurements of clouds in the Antarctic, especially in the interior. They also indicate that better and more frequent remote sensing and *in situ* observations are needed.

The selection of the FORUM project (Palchetti et al., 2020) in 2019 as the 9th Earth Explorer Mission by European Space Agency (ESA) has revitalized studies in the FIR part of the spectrum. As discussed until now, many studies have shown that the FIR can be used to complement standard remote sensing measurements performed in the MIR and improve cloud detection, classification, and inference of cloud properties (Rathke et al., 2002; Palchetti et al., 2016; Di Natale et al., 2017; Maestri et al., 2019a). Moreover, ground-based remote sensing spectral upwards-looking measurements are very useful to determine the cloud properties relevant to the energy budget (Mahesh et al., 2001; Cox et al., 2014; Natale et al., 2020).

In this chapter a unique dataset derived from FIR and MIR downwelling spectral radiances measured at the Concordia station, Dome C, in the middle of the Antarctic Plateau is exploited. The measurements are performed by means of the REFIR-PAD Fourier transform spectroradiometer (Bianchini et al., 2019), in the scope of the projects Radiative Properties of Water Vapor and Clouds in Antarctica (PRANA) and Concordia Multi-Process Atmospheric Studies (CoMPASs), within the Italian National Program for Research in Antarctica (PNRA) (Palchetti et al., 2015). These projects represent the first long-term field campaigns to collect high spectral resolution radiances in the FIR, with continuity for an extended period (measurements started in 2012). REFIR-PAD is installed inside an insulated shelter, named the *Physics Shelter*, together with a backscattering Light Detection and Ranging (LiDAR). The LiDAR detects

backscattering and depolarization signals up to 7 km above the surface. Besides these measurements, the Antarctic Meteo-Climatological Observatory installed at Concordia (http://www.climantartide.it/, last access: 25th January, 2021) provides data from an Automatic Weather Station (AWS), and from daily radiosondes launches. These measurements are analyzed and correlated to the meteorological conditions observed at the Concordia station and considered representative of a large area of East Antarctica because of the horizontal uniformity in the Antarctic Plateau.

The CIC algorithm is applied to REFIR-PAD downwelling radiances to detect and classify Antarctic clouds between 2012 and 2015. The main goal of this effort is to obtain statistics on clear/cloud occurrence and in the investigation of the diurnal cycle and seasonality of clouds in the Antarctic Plateau. Both ice and mixed-phase clouds have been considered, the latter consisting of a geometrically thin supercooled liquid water layer that, in general, may have ice particles present either above or below this liquid layer.

The algorithm is first applied to a test set so that the CIC performances can be assessed. The excellent classification scores obtained in the testing phase provide a solid base for the application of the CIC to the entire dataset. In this study, an effort is made to link the meteorological state of the atmosphere to the cloud occurrence.

This chapter paper is organized as follows. Section 7.1 describes the instrumentation and measurements performed at Concordia Station. Section 7.2 describes the CIC set-up and its optimization to identify and classify clouds. Section 7.3 discusses the cloud occurrences results in different time scales. The study performed in this chapter is summarized in Section 7.4.

7.1 Concordia Station: Instrumentation and Measurements

Concordia station is an Antarctic research base located at Dome C over the Antarctic Plateau (75°06'S, 123°23'E, 3.230 m AMSL), in the East Antarctic region (Figure 7.1). The station opened in 2005 as part of an international cooperation project between the PNRA and the French Polar Institute Paul-Émile Victor (IPEV). A detailed description of the instrumentation available in the PRANA and CoMPASs experiments at Concordia station is given in Palchetti et al. (2015). A brief overview of the instruments and measurements made between 2012 and 2015 is provided in what follows.

Spectral measurements of the downwelling radiance are performed by REFIR-PAD, which provides spectrally resolved zenith-looking radiance measurements in the range 100-1500 cm^{-1} with a 0.4 cm^{-1} spectral resolution, thus covering a large part of the atmospheric longwave emission including both the FIR and part of the MIR region. The instrument points at the zenith through a 1.5 m chimney. The measurement sequence
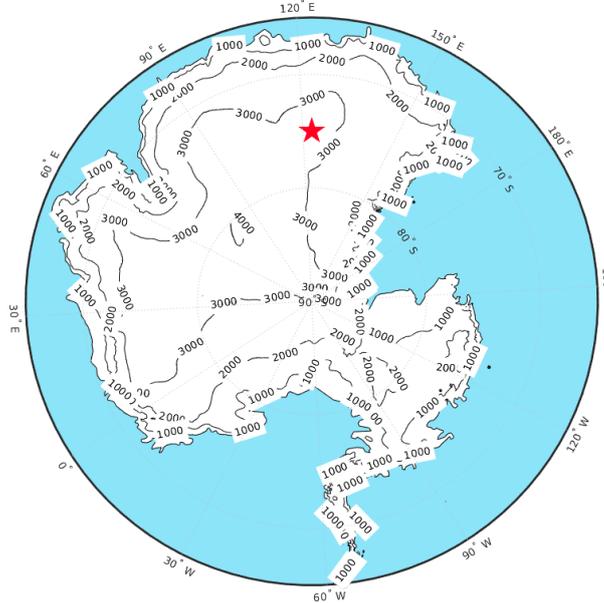
Figure 7.1: Antarctica elevation map, with the Concordia station indicated by the red star.

to obtain one complete spectrum is made of four calibration acquisitions, in which the instrument looks at the internal reference blackbody sources, and four sky observations. Each single acquisition takes about 80s. The entire sequence has a duration of about 14 min: 5.5 minutes of sky observations, 5.5 minutes of calibrations, and delays for detector settling after scene changes (Palchetti et al., 2015). 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 has been designed to operate with uncooled detectors and optics. The instrument operates full time, alternating cycles of 5-6 hours of measurements, with 1-3 hours of analysis. It is installed in the Physics Shelter, located 500 m southward from the main station, in what is called the *Clean Air Area*, where the predominant winds keep the air clean from the exhaust plume of the Concordia power generator. Between the years 2012 and 2015, a total of 87960 spectra were analysed. The spectra annual distribution is reported in Table 7.1.

Table 7.1: Number of analysed REFIR-PAD spectra for each year.

Year:	2012	2013	2014	2015
Spectra:	16177	19298	25089	27396

Atmospheric backscattering and depolarization ratio profiles are measured from a LiDAR system every 5 minutes (Palchetti et al., 2015). This instrument operates at 532 nm and uses a Quantel laser (Brio). The measurements are made in the range 30-

7000 m above the surface, with 7.5 m of vertical resolution. The LiDAR telescope has refractive optics with 10 cm diameter and 30 cm focal length. There is an interference bandpass filter of 0.15 nm bandwidth. The line of sight is zenith-looking through a window, enabling measurements in all-weather conditions.

A set of REFIR-PAD spectra (1928) is co-located with LiDAR measurements. The co-location criterion is defined by the time of measurements: each REFIR-PAD spectrum is associated to the LiDAR data that is closest in time. Co-located measurements are used to visually classify the REFIR-PAD spectra. This process requires a human intervention for the analysis of each co-located measurement and is not routinely applicable to the whole database. For these cases, cloud layers are detected from the analysis of the backscatter profiles and the depolarization ratio is used to determine the thermodynamic phase of the particles. The visually classified spectra are then used to set up training and test sets as described in more detail in the next section. In the Antarctic environment the determination of cloud thermodynamic phase is not trivial. According to Liou and Yang (2016), liquid water droplets retain the polarization state of the incident energy, while the light beam backscattered from non-spherical ice particles is partially depolarized as a result of internal reflections and the transformation of coordinate systems governing the electric vector. A theoretical analysis performed by the same authors shows that in presence of a liquid water cloud the depolarization remains at about 2-4%, whereas radiation backscattered from non-spherical ice particles is strongly depolarized, varying between 30 and 40%. However, the threshold to determine the water physical state in real clouds can vary depending on the atmosphere and the cloud microphysical parameters. Sassen and yu Hsueh (1998) evaluate ground-based LiDAR data in presence of contrail cirri, during the Subsonic Aircraft: Contrail and Cloud Effects Special Study (SUCCESS) field campaign. They found depolarization ratios in persisting contrails ranging from about 0.3 to 0.7. Freudenthaler et al. (1996) observed depolarization ratios of 0.1 to 0.5 for contrails with temperatures ranging from -60 to $-50°$C, depending on the stage of their growth. In this study a depolarization ratio of 0.15 is used as a threshold for the discrimination of the liquid water clouds and ice clouds over the Concordia Station.

An example for the observation of a clear sky (red triangle), ice cloud (blue triangle), and mixed-phase cloud (green triangle) is provided for the LiDAR backscatter and depolarization ratio in the upper and middle panels of Figure 7.2. The lower panel of the same figure provides the correspondent REFIR-PAD spectra. In clear sky conditions, the LiDAR backscattering signal decreases with altitude, while the signal increases in the presence of cloud particles. As shown in the figure, clouds can be composed of multiple layers, each one with different depolarization features. When the depolarization ratio is higher than 15%, the cloud is classified as an ice cloud (blue triangle). For lower values of the depolarization ratio, the cloud is assumed to be a mixed-phase cloud (green triangle). In particular, it is frequently observed that mixed-phase clouds are composed of a layer of liquid water near the cloud base and an ice layer at the cloud top.

Since 2005, the Concordia station has provided hourly measurements of air temper-
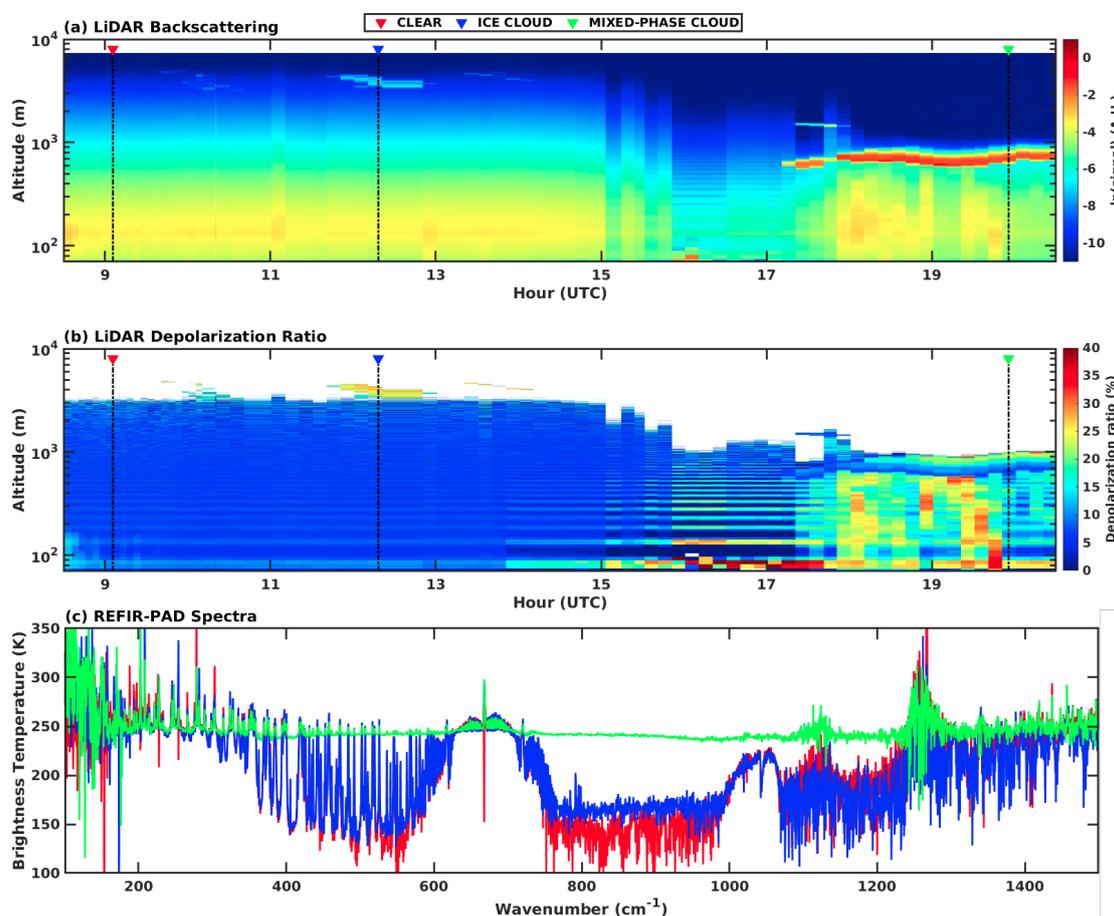
Figure 7.2: (a) LiDAR backscattering and (b) depolarization ratio for 2013, 2nd January. (a) and (b): Different sky conditions are highlighted in correspondence of vertical dashed line. A red triangle indicate clear sky, blue triangle is used for ice cloud, and green triangle for mixed-phase cloud. (c): REFIR-PAD spectra in correspondence of the three sky conditions highlighted in the upper and middle panels. The same color code is used.

ature, pressure at the surface level, relative humidity, wind speed, and wind direction. The snow temperature is measured at different depths from 5 cm to 10 m. These measurements began in December 2012. Radiosondes Vaisala RS92 are routinely released every day at 12 UTC, since 2006. They reach an altitude of about 18 km in wintertime and about 25 km in the summer. All these data are made available by the Antarctic Meteo-Climatological Observatory and a subset of them is used in this study.

7.2   CIC Algorithm: set up

The CIC algorithm is described in Section 4.1. For this study the CIC classification consider three classes: clear sky, ice clouds and mixed-phase clouds. Consequently, three TSs are arranged, each one containing spectra representatives of one class. The classification of an observed spectrum is performed in accordance with the the logical diagram of Figure 7.3, by using the *distributional approach*. The diagram shows the comparison between specific couples of classes. The partial results of each comparison are represented by white boxes. If one class prevails over the other two class a classification is reached and the final output is provided (green boxes in the figure).
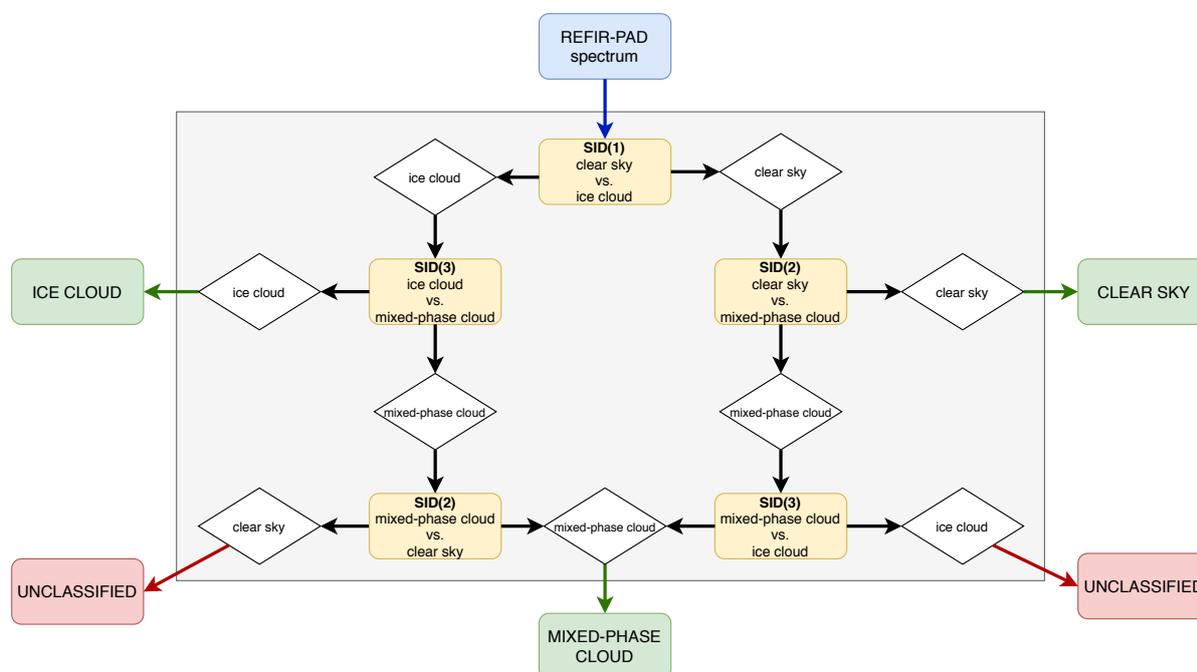
Figure 7.3: Logical diagram of the classification process performed by the CIC algorithm for the definition of clear sky, ice cloud and mixed-phase cloud class.

7.2.1   Training and Test Sets

Spectra used to populate the training sets are chosen from a set of manually classified observations. The identification is performed by visually inspecting the co-located LiDAR backscatter and depolarization profiles in accordance with the criteria described in Section 7.1. Each training set contains a limited number of spectra from the REFIR-PAD database, aiming at describing the variability of atmospheric conditions over the Concordia station. Due to the intense variations of the environmental conditions, the training

sets are defined for two macro-seasons: a warm season (November-March) and a cold season (April-October). The choice is also supported by the fact that mixed-phase clouds are extremely rare in the cold macro-season. Ricaud et al. (2020) observed the occurrence of supercooled liquid water clouds during the warm macro-season only, with the largest frequency occurring in December and January. Listowski et al. (2019) also observed that the fraction of supercooled liquid water-containing clouds in the Antarctic Plateau varies between 10%, in the summertime, and 0%, in winter. Therefore, three training sets for the warm macro-season are defined: clear sky, ice cloud, and mixed-phase cloud. For the cold macro-season, only the clear sky and ice cloud training sets are used. Table 7.2 summarizes the number of spectra for each TS and macro-season.

Mean spectra in terms of BT (solid lines) and their standard deviations (shaded area) are presented in Figure 7.4 for the training sets used for both macro-seasons. Differences between the mean spectra of the different classes are observed in the window channels located between 400 and 600 cm^{-1}, and 800 and 1000 cm^{-1}. Note that in IR window regions (transparent channels) the standard deviation of the clear sky spectra is usually lower than that of the cloudy spectra, which account for a wider signal variability in these bands. Furthermore, the clear sky signal is very low at window wavenumbers, that results in a very low signal to noise ratio.

Table 7.2: Number of spectra used in each TS, according with the macro-season.

Season / class	clear sky	ice cloud	mixed-phase cloud
November-March	49	30	22
April-October	64	37	–

Once the TSs are defined, the CIC is ready to ingest the REFIR-PAD spectra and provide their classification. To evaluate the CIC performance and optimize its set-up, a test set is analysed of 1726 pre-classified spectra collected in 2013. The test set is composed of 559 clear sky, 1022 ice cloud, and 145 mixed-phase cloud spectra. These spectra were previously classified by using the co-located LiDAR backscatter and depolarization profiles. An example is provided in Figure 7.2. The sky condition is defined as that observed when the REFIR-PAD starts its measurement. Then, the visually classified REFIR-PAD spectra are associated to the sky conditions encountered at the beginning of each measurement.

7.2.2   CIC performances and optimization

The CIC algorithm is applied to the test set spectra by accounting for their BT in different spectral intervals. This operation is performed to find the optimal spectral interval that maximizes the classification results for each class (clear sky, ice cloud,
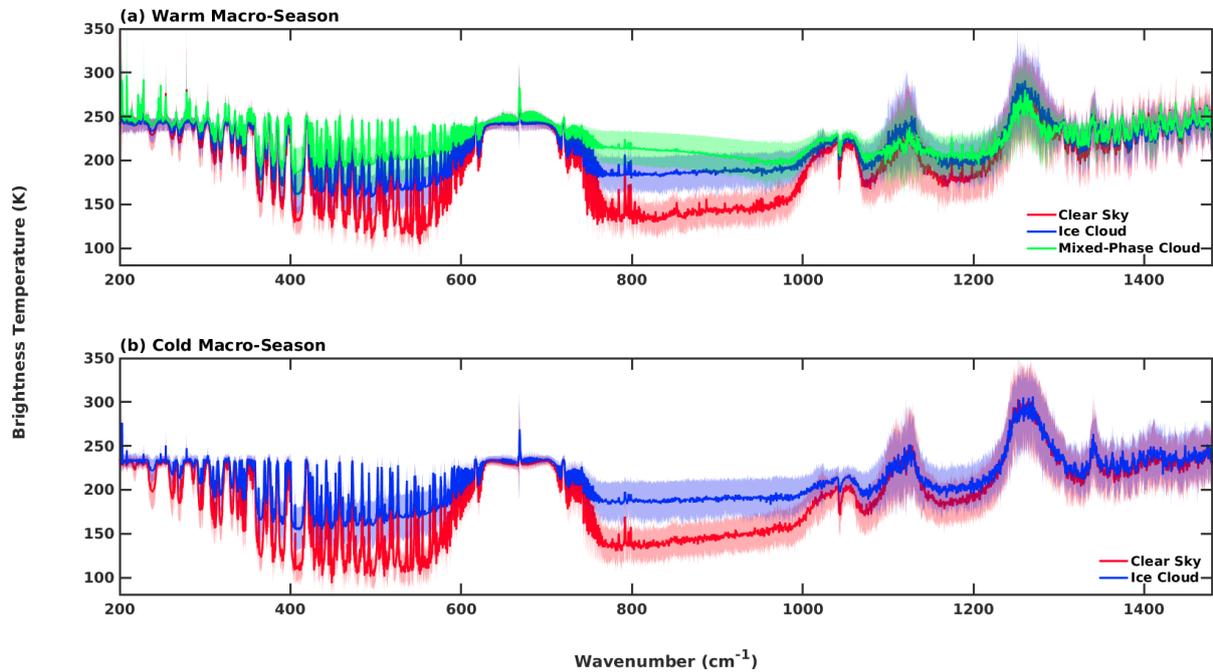
Figure 7.4: Average BT (solid line) ± 1 standard deviation (shaded area) for TS elements of the (a) Warm and (b) Cold macro-seasons.

mixed-phase cloud). Multiple runs of the CIC algorithm are performed on the same test set, by applying it to different spectral ranges. Specifically, the starting wavenumber is moved, at steps of 20 cm^{-1}, in the $200 - 600 cm^{-1}$ band, and the ending wavenumber is moved between 960 and 1480 cm^{-1}.

The algorithm performance during this process is assessed by evaluating the Threat Score (ThS). A confusion matrix is used to compute the ThS of each class and for each considered spectral interval. Each individual spectrum can be classified correctly as a member of its class (i.e class A), or incorrectly as a member of a different class (i.e. class B or C). With this symbolism, the spectrum classification is interpreted in terms of:

- True positive (TP): the spectrum belong to class A and it is properly classified in class A.

- True negative (TN): the spectrum does not belong to class A and it is properly classified in its class of pertinence (B or C).

- False positive (FP): the spectrum belongs to class B or C but it is misclassified in class A.

- False negative (FN): the spectrum belongs to class A but it is misclassified in class B or C.

Given the above possibilities, for each class the threat score is defined as:

$$ThS = \frac{TP}{TP + FN + FP} \tag{7.1}$$

that accounts for the correctly classified spectra (TP) in the class and penalizes all the misclassified occurrences (FN and FP). A ThS value of 1 means that there are no misclassified spectra.

Based on the results obtained for each of the combinations of starting and ending wavenumbers, the ThS is calculated for each class (clear sky, ice cloud, mixed-phase cloud). The weighted mean ThS values, that account for the total number of cases in each class, are also calculated. In the upper left panel (a) of Figure 7.5 the mean ThSs are plotted as a function of the starting and ending wavenumbers. The other panels in this figure (b-c-d) show results for the three specific classes. The ThS values span from 0.487 to 0.966 in accordance with the selected interval and the given class. For intervals ending with wavenumbers larger than 1140 cm^{-1}, the ThS decreases considerably for all the classes. This is likely associated to the noise of the REFIR-PAD sensor which increases considerably above 1200 cm^{-1} and degrades the classification results. When the ending wavenumber is set to values between 980 and 1080 cm^{-1}, the ThS is very high (larger than 0.9) for all the starting wavenumbers below 400 cm^{-1}, both for clear sky and ice clouds. The spectral interval 380-1000 cm^{-1} performs the best for classification of both clear sky and ice clouds, where the ThS values are 0.963 and 0.966, respectively. The classification of mixed-phase clouds is slightly less robust compared to the other two classes, and the best spectral interval is 540-1020 cm^{-1} with a ThS of 0.927.

When accounting for all the classes, the most performing spectral range for clear and cloud identification and classification is the 380-1000 cm^{-1} interval. The result is dependent on sensor characteristics and for this study it is specifically driven by the REFIR-PAD spectral resolution and noise features. The optimal interval for the classification is also dependent on many other parameters, among which are the type and number of classes considered, the observation geometry (e.g. satellite or ground based), the observing location, and the mean atmospheric conditions. Because the water vapor content is extremely low, the ground-based measurements on the Antarctic Plateau allow the full exploitation of the FIR spectral range. These channels would be totally opaque for upward observations in regions of increased water vapor content such as the tropics. The selected spectral range (380-1000 cm^{-1}) highlights the fundamental role of the FIR part of the spectrum in the cloud identification and classification.

The results of the CIC classification applied to the test set using the 380-1000 cm^{-1} are summarized in Table 7.3. The Table reports the number of spectra per class in the test set, the CIC HRs, in percentage, and the threat scores. The HR for a class (i.e. A) is defined as:
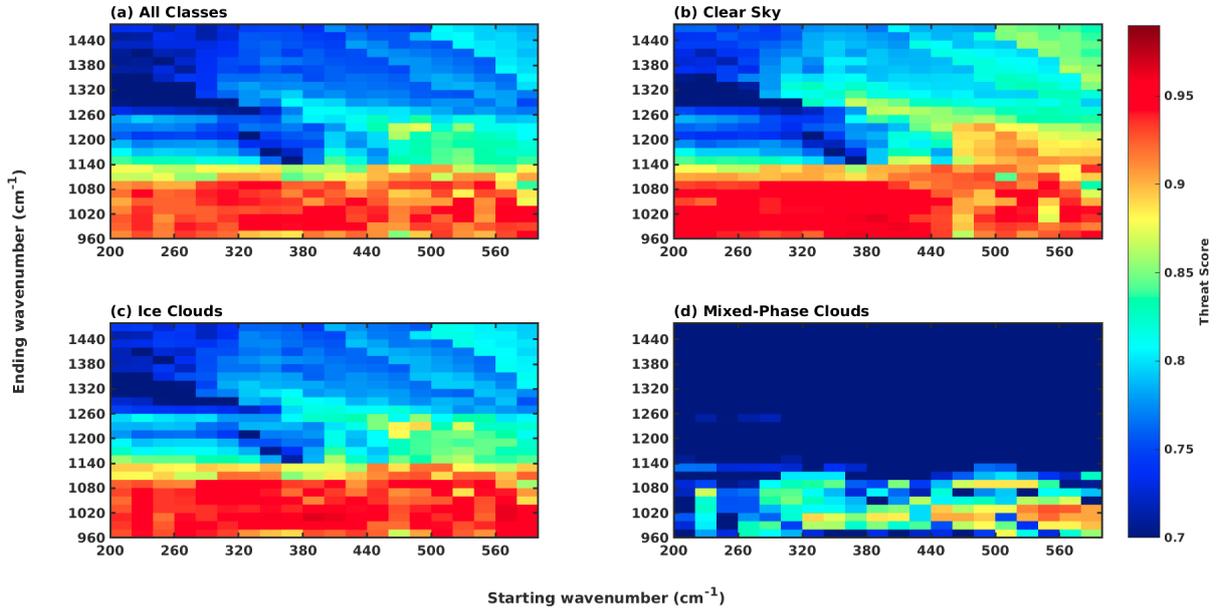
Figure 7.5: Threat Scores for the test set as a function of different spectral intervals ingested by the CIC algorithm. The (a) All Classes threat score is a weighted mean of the threat scores computed for each class: (b) clear sky, (c) ice cloud, and (d) mixed-phase cloud.

$$HR_A = \frac{N_A^{\mathrm{CIC}}}{N_A^{\mathrm{true}}} = \frac{TP}{TP + FN} \qquad (7.2)$$

where N_A^{CIC} is the number of occurrences of the class A that are correctly identified by the CIC (corresponding to the True Positive (TP) in the confusion matrix). N_A^{true} is the total number of elements in class A of the dataset, and corresponds to TP+False Negative (FN) of the class A.

The overall performance is that almost 98% of spectra are correctly classified. Only a small percentage (less than 1%) of cloudy spectra (ice clouds plus mixed-phase clouds) is misclassified as clear sky, and about 2% of the clear sky spectra are erroneously identified as ice clouds. Note that in case of mixed-phase clouds the CIC is able to identify the presence of the cloud in 99.3% of the cases even if for the 8.3% the cloud phase is classified as ice instead of mixed-phase. This is actually a very reasonable performance considering that, as noted before, most of the mixed-phase clouds are composed of a layer of super-cooled liquid phase near the cloud base and ice phase particles close to the cloud top. For optically thin clouds, the definition of the phase can be problematic.

Table 7.3: Test set classification performed by CIC, using the optimal spectral range 380–1000 cm^{-1}.

Class	# Spectra	Hit Rate	Threat Score
Clear Sky	**559**	**98.0% - Clear Sky** 2.0 % - Ice Cloud 0.0 % - Mixed-Phase Cloud	**0.963**
Ice Cloud	**1022**	0.9 % - Clear Sky **98.7 % - Ice Cloud** 0.4 % - Mixed-Phase Cloud	**0.966**
Mixed-Phase Cloud	**145**	0.7 % - Clear Sky 8.3 % - Ice Cloud **91.0 % - Mixed-Phase Cloud**	**0.886**
Total	1726	97.9 % - Correct 2.1 % - Misclassified	**Weighted mean:** **0.958**

Test Set Misclassified Spectra

Each of the misclassified cases is inspected visually to understand the main causes of error in the CIC classification. It appears that the misclassification of clear sky as ice cloud, and vice-versa, occurs primarily for spectra taken during the cold macro-season. The misclassification in this case is associated with the: (a) presence of a very thin cirrus cloud; (b) REFIR-PAD measurements taken over a period of time in which the observed scene is changing (i.e. the measuring time encompasses both clear sky and cloudy sky); (c) presence of suspended particles near the surface (e.g., diamond dust, wind-blown snow, or combustion products produced by the generator that heats Concordia Station).

During the warm macro-season, a small percentage of mixed-phase clouds are misclassified as either clear sky or ice clouds. In some cases, ice clouds are misclassified as mixed-phase clouds; this happens mostly when the ice cloud spectra are characterized by a high BT in the main window region.

7.3   Results

The 380-1000 cm^{-1} spectral interval is used to run the CIC algorithm over the entire REFIR-PAD dataset, comprising measurements from year 2012 through 2015. In Figure 7.6, the CIC classifications are compared with co-located LiDAR depolarization data for two different days. For each REFIR-PAD observation, the classification is reported as a colored triangle in the upper part of each panel. As previously discussed, low values of LiDAR depolarization together with large values of the backscattering signal (not shown) indicate the presence of liquid water phase in the cloud layer, while high

depolarization values are observed in presence of ice clouds. The upper panel of Figure 7.6 shows the presence of a mixed-phase cloud over the Concordia station from about 10:00 UTC until the nighttime of the 3rd of January 2014. The presence of the cloud and its thermodynamic phase are correctly identified and classified by the algorithm. Between hours 21:00 and 22:30 UTC, CIC identifies a spectral signal characteristic of ice clouds that corresponds to larger values of the depolarization ratio measured by the LiDAR. On the 1st of August 2014 (lower panel of Figure 7.6), the LiDAR depolarization shows that the day starts with a precipitating ice cloud, followed by clear sky conditions from 15:00 UTC. For this case, both the clear sky and the ice cloud are correctly detected by the CIC algorithm.
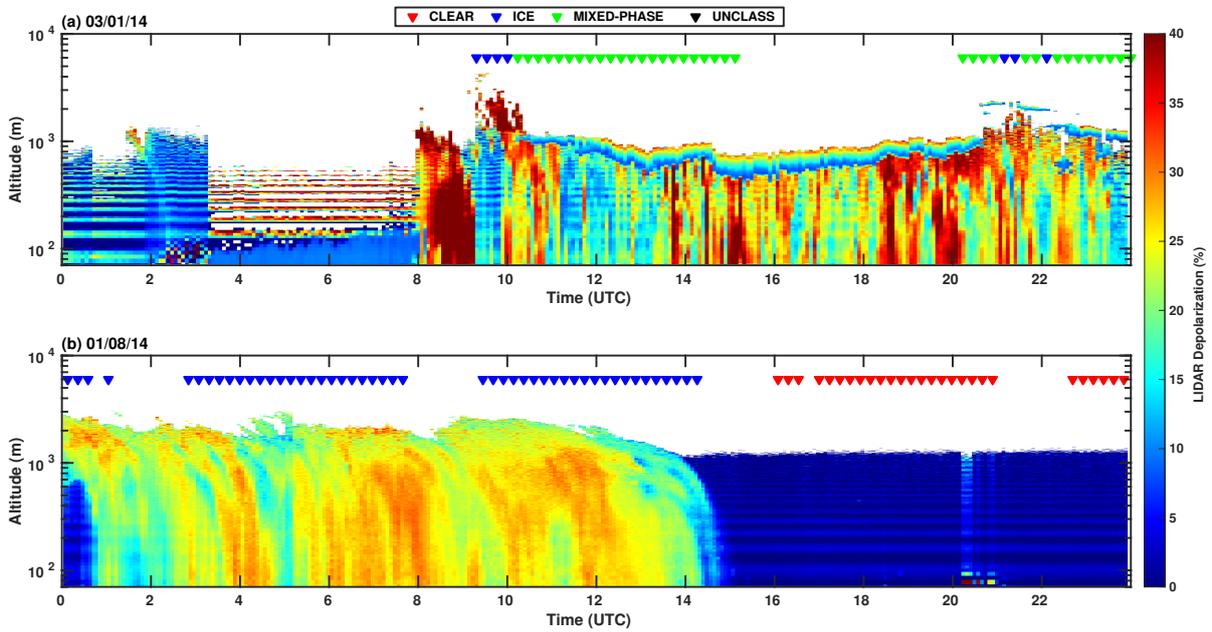
Figure 7.6: LiDAR depolarization ratio on (a) 3rd of January 2014 and (b) 1st of August 2014. The triangles mark the REFIR-PAD observations. The color code indicates the CIC classification: red for clear sky, blue for ice cloud, green for mixed-phase cloud, and black for unclassified spectra.

The results of applying the CIC to the full available REFIR-PAD dataset are provided in terms of percentages, defining the occurrence of each class with respect to the total number of analysed spectra. An error can be associated to the percentage occurrence, exploiting the HRs derived in the analysis of the test set. With the use of Equation 7.2 for the HR definition for the class A:

$$N_A^{\text{CIC}} = N_A^{\text{true}} \times HR_A \tag{7.3}$$

the number of misclassified spectra (N_A^{err}) of class A can be written as:

$$N_A^{\text{err}} = N_A^{\text{true}} \times (1 - HR_A) \tag{7.4}$$

Through combination of Equation 7.3 and Equation 7.4, it is possible to remove the term N_A^{true} which is unknown for results applied to the entire dataset. The following relation is then derived:

$$N_A^{\text{err}} = N_A^{\text{CIC}} \times \frac{(1 - HR_A)}{HR_A} = N_A^{\text{CIC}} \times \left(\frac{1}{HR_A} - 1 \right) \tag{7.5}$$

The relative error (ϵ), associated to the classification of the elements of class A, is obtained by dividing the number of misclassified A spectra to the total number of spectra $N_{A+B+C} = N_{\text{TOT}}$:

$$\epsilon_A = \frac{N_A^{\text{err}}}{N_{\text{TOT}}} = \frac{N_A^{\text{CIC}}}{N_{\text{TOT}}} \times \left(\frac{1}{HR_A} - 1 \right) \tag{7.6}$$

Note that the HR values associated with the individual classes for the entire dataset are unknowns. However, it is assumed that the CIC scores over the test set spectra are representatives of the performances that are obtained over the full dataset. Therefore, the HRs obtained for the test set analysis (see back Table 7.3) are used in place of the dataset HR in Equation 7.6. Thus, for the class A, the percentage classification error is simply:

$$\epsilon_A\% = \frac{N_A^{\text{CIC}}}{N_{\text{TOT}}} \times \left(\frac{1}{HR_A} - 1 \right) \times 100 \tag{7.7}$$

where N_A^{CIC} is the number of spectra identified by CIC as member of class A and N_{TOT} is the total number of spectra in the entire dataset. The HR_A is obtained from the application of CIC to the test set and is thus known a-priori. Note that for a small number of false positives ($FP \ll TP$) the HR for class A is very similar to the ThS for the same class. CIC provides very small values of FP when applied to the test set, with respect to TP values: 2% for ice clouds and clear sky, and about 3% for mixed-phase clouds.

7.3.1 Sky classification: 4 years averages and inter-annual variability

A total of 87960 REFIR-PAD spectra are analysed from the dataset spanning over the time range 2012-2015. From this set, only 202 spectra are used as TS, and the other 87758, which include the test set elements, are ingested by the CIC algorithm to evaluate the cloud occurrence over the Concordia station. The classification results are shown in Table 7.4 as percentages for clear sky, ice clouds, mixed-phase clouds, and unclassified spectra. The entire dataset and individual years classifications are presented, as

well as the estimated percentage uncertainties (see Equation 7.7). On average, the clear sky is detected in almost 72% of the cases, with ice cloud occurrence of about 25% and mixed-phase cloud occurrence of less than 3%. The inter-annual variability of total cloud occurrence in the Antarctic Plateau, the sum of ice and mixed-phase clouds, spans between about 23 and 31%. This percentage interval is in accordance with the observations from Adhikari et al. (2012), who analysed data from CloudSat and CALIPSO between 2006 and 2010 and reported percentages spanning between 20–30% interval. From our analysis the cloudiest year in the 2012-2015 period is 2012, with a value of 31.2%. This is almost identical to what observed in 2015 (cloud occurrence is 31.1% in this case), with the difference being that in 2012 there was a significantly larger fraction of mixed-phase clouds than in 2015 (5.8% and 1.5% respectively).

Mean temperatures at the surface level for the entire dataset and for each single year are also reported in Table 7.4. Temperatures are measured every hour at the Concordia station and are linearly interpolated in time to be associated with the REFIR-PAD measurements and the corresponding CIC classifications. The last row of Table 7.4 provides information only for the months of the warm macro-season from November to March. The results suggest a positive correlation between mean air temperatures at surface level in the Warm macro-season and the occurrence of mixed-phase clouds. Note that mixed-phase clouds are present only for months from November to March. The temperature and mixed-phase cloud correlation could indicate that warm temperatures are favorable for mixed-phase clouds formation or that the presence of warm liquid clouds implies a stronger cloud forcing at the surface and, consequently, an increase in the temperature values near the ground. Ice clouds are observed during the entire year. In contrast with mixed-phase clouds, their occurrence does not seem correlated to the mean air temperature at the surface. Note that the maximum occurrence of ice clouds is observed during year 2015, which had the lowest mean value of surface air temperature in the 4 years time range.

7.3.2 Seasonal clear sky and cloud occurrence

Seasonal averages of cloud occurrence are computed for the entire dataset and presented in Table 7.5. The Table also reports the number of spectra observed in each season, which show that the data are homogeneously distributed over the course of the year, and the mean air temperatures. The mean total cloud occurrence varies from the minimum value of 23.9% detected in austral spring (SON) to the maximum value of 33.2% in the cold winter season (JJA). The dominant cloud occurrence and thermodynamic phase is ice. During the austral summer, the occurrence of ice clouds is the smallest. However, for the same season, the occurrence of mixed-phase clouds reaches its maximum over Concordia Station (10.9%). It is interesting that during summer, more than one third of the clouds over Concordia is of the mixed-phase type. The occurrence of mixed-phase clouds in summer is in line with the analysis performed by Listowski et al. (2019),

Table 7.4: CIC classification results for the whole REFIR-PAD spectra dataset (2012–2015) and for single years. The associated uncertainties are computed using Equation 7.7. Mean air temperatures at surface level for the entire period and for each year are also reported. The last row refers to mean air temperatures at surface level computed for the months from November to March (Warm season).

CIC Classification	Entire dataset (%)	2012 (%)	2013 (%)	2014 (%)	2015 (%)
Clear sky	**72.3 ± 1.5**	68.6 ± 1.4	75.1 ± 1.5	76.3 ± 1.5	68.8 ± 1.4
Ice cloud	**24.9 ± 0.3**	25.4 ± 0.3	22.8 ± 0.3	21.1 ± 0.3	29.6 ± 0.4
Mixed-phase cloud	**2.7 ± 0.3**	5.8 ± 0.6	2.0 ± 0.2	2.5 ± 0.2	1.5 ± 0.2
Unclassified	**0.1 ± 0.1**	0.2 ± 0.1	0.1 ± 0.1	0.1 ± 0.1	0.1 ± 0.1
Mean T ($^\circ C$)	-53.5	-49.6	-54.5	-53.4	-55.0
Warm season Mean T ($^\circ C$)	-40.2	-37.6	-41.0	-40.7	-41.1

who analysed DARDAR data (Delanoë and Hogan, 2010; Ceccaldi et al., 2013) based on combined observations from CloudSat and CALIPSO satellites, in the period 2007–2010. The same authors, by performing a visual analysis of the geographical distribution of the clouds containing liquid water particles, estimate that during the other seasons (MAM, JJA, and SON), the occurrence of mixed-phase clouds is close to 0%, in the region around the Concordia station.

Table 7.5: Mean seasonal occurrences of clear sky, ice clouds, and mixed-phase clouds at Concordia station. Mean surface air temperatures are reported for each season.

	DJF	MAM	JJA	SON
# spectra	21209	21093	22395	23061
CIC CLASSIFICATION	(%)	(%)	(%)	(%)
CLEAR SKY	71.1	75.1	66.8	76.1
ICE CLOUD	17.6	24.7	33.2	23.8
MIXED-PHASE CLOUD	10.9	0.2	0.0	0.1
UNCLASSIFIED	0.4	0.0	0.0	0.0
Mean T ($^\circ$C)	-34.9	-61.0	-65.0	-52.2

Seasonal occurrences for each class are analysed in combination with meteorological parameters encountered during the corresponding REFIR-PAD measurements. In Figure 7.7, the percentage distribution of each class seasonal occurrence is reported as a function of the air surface temperature, with histograms binning of 7°C. The same color code is adopted here that was used previously: clear sky in red, ice clouds in blue, and mixed-phase clouds in green. The number of REFIR-PAD measurements for each bin is reported

at the base of the histograms. Over the four years, the surface air temperature (in correspondence of REFIR-PAD measurements) varies between a minimum of $-81.3°C$ and a maximum of $-15.8°C$. With the exception of the spring season (SON, lower right panel of Figure 7.7), the results show that the detected cloudy sky occurrence increases (clear skies decrease) as surface air temperature increases. This holds both for ice and mixed-phase clouds. In the winter season (JJA, lower left panel of Figure 7.7), for surface air temperature larger than $-43.3°C$ the CIC identifies only ice cloud conditions. Note that the winter and spring seasons have the largest variation in the air surface temperatures. In the winter season, extremely low temperatures (below $-70°C$) are very frequent and result from the lack of insolation, the dry atmospheric conditions, and the absence of clouds. In the same season, higher surface temperatures are measured mainly in correspondence of cloudy sky conditions. The downwelling longwave radiation from cloud layers contributes to the surface radiative forcing and mitigates the temperature of the cold season. Over the four-year period the average winter surface temperature in clear sky conditions is $-67.9°C$, while in presence of ice clouds is $-59°C$.
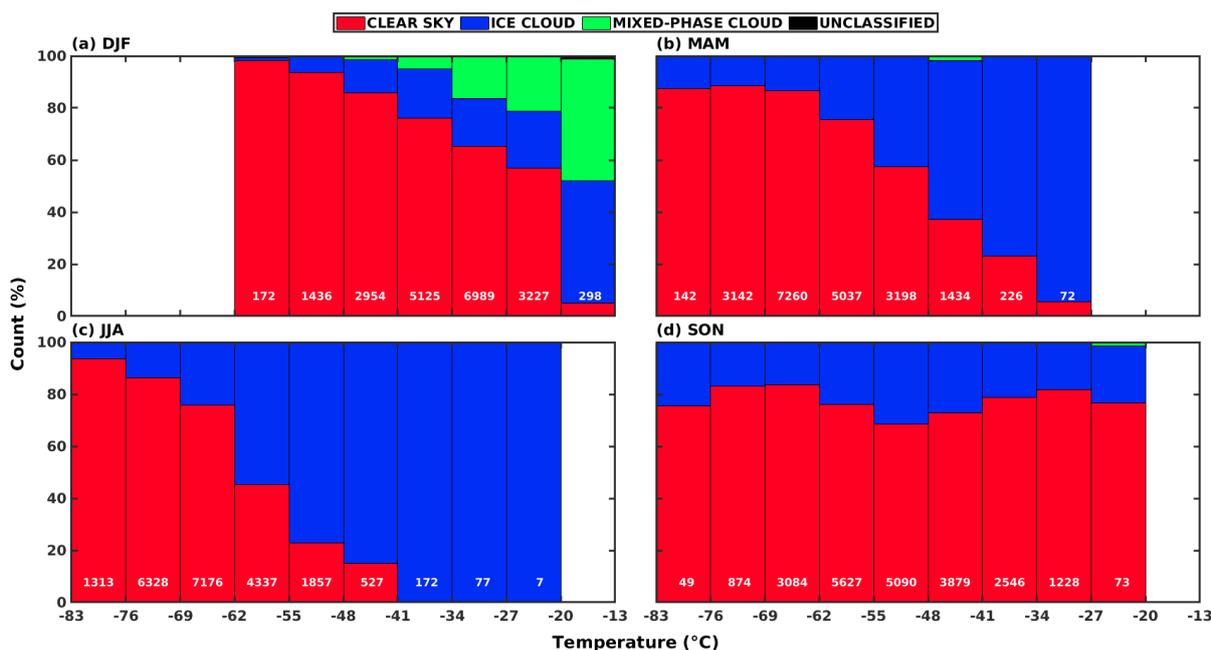
Figure 7.7: Histograms of the seasonal occurrence of the analysed sky conditions as a function of the surface air temperature. (a) DJF - summer, (b) MAM - autumn, (c) JJA - winter, and (d) SON - spring. The number of observations for each $7°C$ bin is reported at the base of each histogram.

A similar analysis is performed by relating clear and cloudy sky occurrences to measurements of surface relative humidity and surface pressure.

The relative humidity and surface pressure histograms are shown in Figures 7.8 and
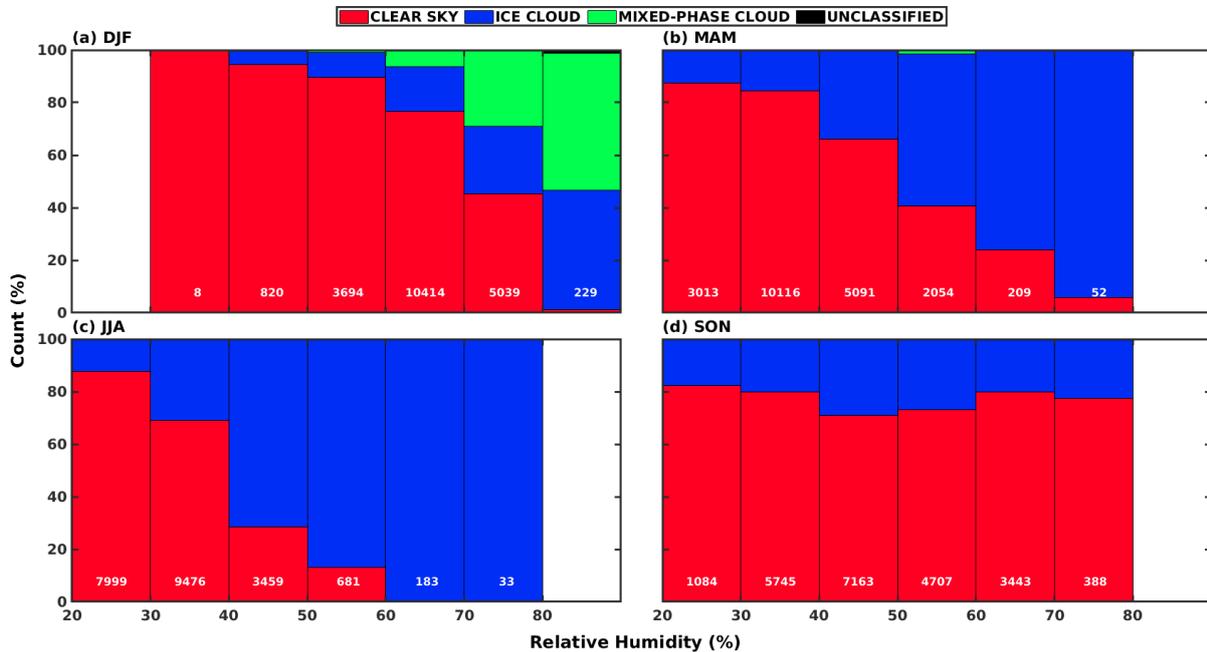
Figure 7.8: Histograms of seasonal occurrence for multiple bins of relative humidity. (a) DJF - Summer, (b) MAM - Autumn, (c) JJA - Winter, and (d) SON - Spring. The number of observations for each 10% bin relative humidity is reported in white at the base of each histogram.

7.9, respectively. The histograms show the same structure of that presented for the air temperature (Figure 7.7), however, in these cases the bin values are 10% to relative humidity and 8 hPa to surface pressure. Over the 4 years, the relative humidity varies between 21% and 84%, and the pressure at surface level ranges from 607.8 to 672.4 hPa. The results indicate that the highest values of relative humidity tend to occur with the highest percentage of clouds for all the seasons except spring. The highest mean values of surface pressure in the summer season tend to occur with the highest percentages of mixed-phase clouds. Unclassified spectra are obtained only in the summer season and in correspondence of very high values of surface pressure, air temperature, and relative humidity.

Surface wind measurements are also analysed and related to CIC classification results for each season. The values of wind speed and direction closest in time to the REFIR-PAD measurements are used. Wind roses are built considering the bias correction methodology proposed by Droppo and Napier (2008), which indicates the necessity of weighting the contribution of each direction to correctly represent them in the wind roses.

In Figure 7.10, the wind roses for each season and class are shown. Clear sky cases correspond to about 70% of all occurrences in all seasons and are associated with a
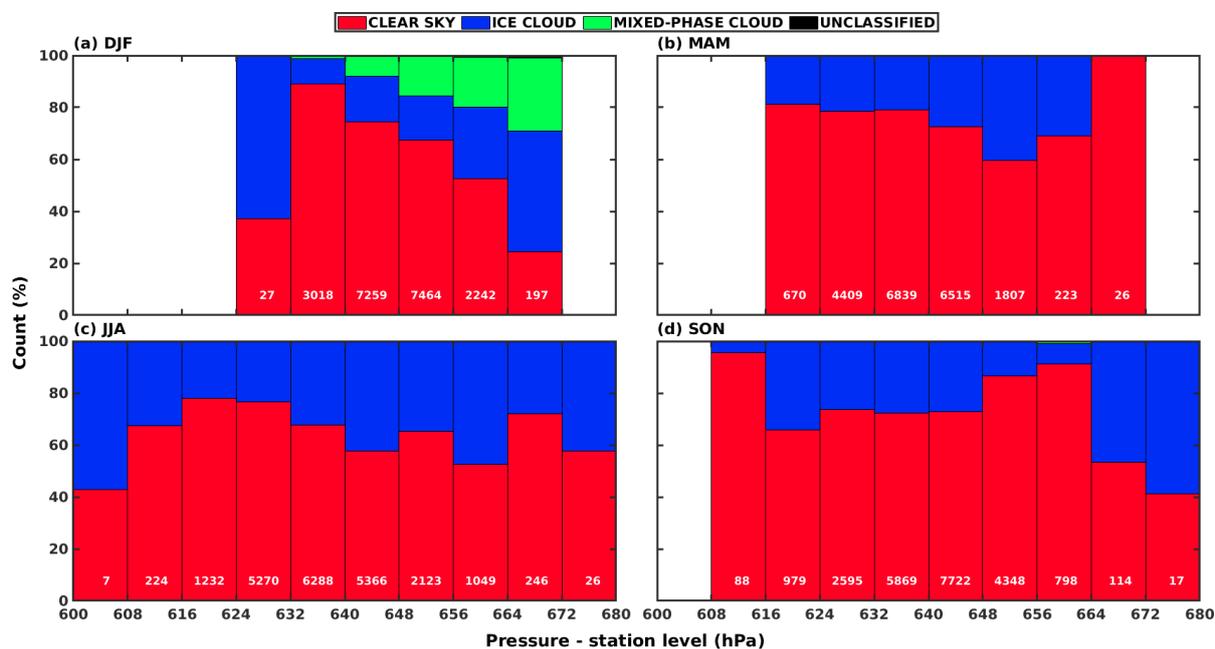
Figure 7.9: Histograms of seasonal occurrence for multiple bins of surface pressure. (a) DJF - Summer, (b) MAM - Autumn, (c) JJA - Winter, and (d) SON - Spring. The number of observations for each 8 hPa bin pressure is reported in white at the base of each histogram.

surface level wind that blows predominantly from South and South-West. Higher wind intensities are found in springtime. An additional wind component from the West is observed in summer, but is negligible in the other seasons. When ice clouds are present, the dominant surface wind direction is from the South-East, and the wind intensity is larger than in clear sky conditions on average (7.7 m/s versus 6.1 m/s). Note that non-negligible occurrences of surface wind from the North-East are observed only when mixed-phase clouds are detected, especially during the fall season (MAM). This component overlaps with the dominant South-East wind component found both in summer and autumn. The wind rose for mixed-phase clouds in the spring season (SON) is reported for completeness but is affected by the very few number of cases detected. Even if very preliminary, the analysis of the surface wind direction in presence of different sky conditions highlights some correlations between the wind component and the clear sky or cloud occurrence. Winds from the South and West directions at the Concordia station are from the Antarctic Plateau, where the drier air is supposedly found. Winds from the South-East and East directions suggest a moisture transport from the Ross sea to the region of the Concordia station. The correlations are far from being conclusive since the upper level winds and the back trajectories of the air masses have not been analyzed yet.
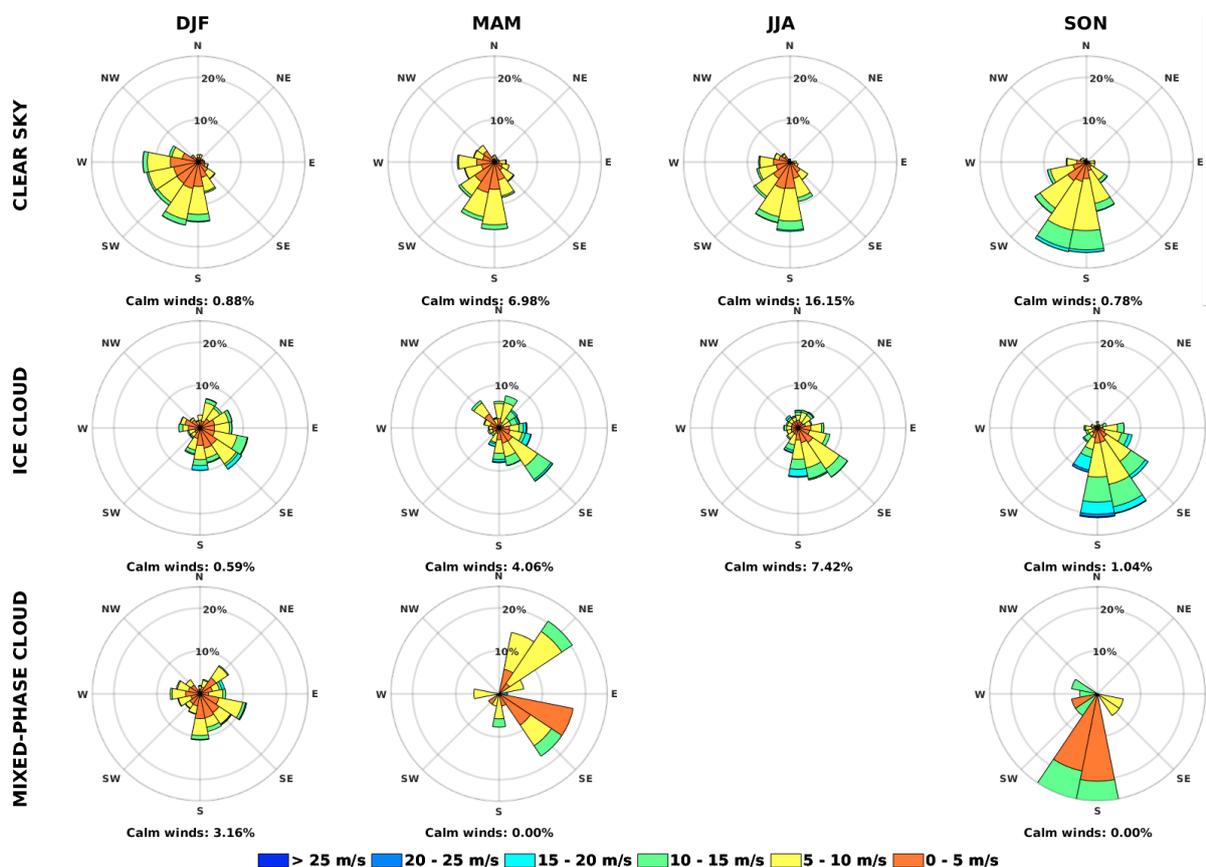
Figure 7.10: Wind roses at Concordia station, for the four seasons (DJF, MAM, JJA, and SON) of the period 2012–2015. Clear sky, ice cloud, and mixed-phase cloud conditions are split into separate rows.

7.3.3 Monthly mean cloud occurrence: comparison with satellite data

CIC monthly mean cloud percentages (including ice and mixed-phase) for the period 2012-2015 are shown in Figure 7.11. The black curve corresponds to the 4-years monthly average cloud occurrence, and the shaded grey area indicates the minimum and maximum CIC monthly values. The lowest average value is found in November (17%), while higher occurrences are observed during the winter months. The peak is located in August, with an average value of 39%. For the same month, the inter-annual variability is quite large as indicated by the extent of the grey area. As examples, in August the monthly mean values span from 31% to 62%, which is the highest derived occurrence, and in November from 1% (lowest registered value) to 37%.

For comparison, the monthly mean cloud occurrences/fractions derived from level-3 (L3) satellite products are reported for the same period of time. According to World Me-

teorological Organization (WMO) [1], the L3 satellite products are composed of variables mapped on uniform space-time grid scales, and are constructed to provide completeness and consistency for the anticipated users. These products type are frequently used to perform climate analysis and model evaluation (e.g. Stubenrauch et al., 2013; Webb et al., 2017). In practice, different data sets present specific strengths and limitations that are briefly described below.

The L3 products used in this work are derived from passive radiometric observations performed by the MODIS on board the TERRA and the AQUA satellite platforms, by the CALIOP on board the CALIPSO satellite, and by the CPR on board CloudSat satellite.

For MODIS L3 products, the occurrence by cloud type is not available, and the cloud fraction is used. This variable is computed as the ratio between the cloud covered pixels and the total number of pixels observed by both satellites platforms each month and is mapped in a global grid of $1°$ of latitude and longitude, which corresponds to an area of about 3000 km^2 in the region of the Concordia station.

Two types of MODIS L3 products are used in this study: MCD06COSP and MYD08/MOD08. The first one combines the observations from both AQUA and TERRA platforms (MCD06COSP_L3, MODIS Atmosphere Science Team, 2020). This product is based on a cloud mask which uses bands at visible and infrared wavelengths. Thus the cloud fraction information (solid blue in Figure 7.11) is not available in the Antarctic region during the dark season (between May and July). The second product is derived from each MODIS sensors on platforms separately (MYD08 for AQUA, and MOD08 for TERRA, MODIS Atmosphere Science Team, 2017). The MOD08/MYD08 L3 product is based on a cloud mask which exploits infrared bands when in absence of solar illumination. In this case, the monthly mean cloud fraction is available for all the seasons (dashed and dotted blue curves in Figure 7.11, for MODIS TERRA and AQUA L3 products, respectively).

In contrast to MODIS, the CALIOP and the CPR active sensors detect the cloud occurrence within vertical profiles. The L3 product from these sensors is a volume cloud occurrence, which considers the number of cloud observations along the vertical profiles that are mapped monthly on a regular grid. The CALIOP L3 product (CAL_LID_L3_Cloud_Occurrence-Standard-V1-00, Winker, 2018) is built on a grid map of $2.5°$ of longitude and $2.0°$ of latitude, which corresponds to an area of about 15000 km^2 in the region surrounding Concordia station. CALIOP mean occurrence is plotted in green in Figure 7.11. The L3 product from CloudSat (3S-RMCP, Haynes, 2019) is available in a grid of $5°$x$5°$ of latitude and longitude, that covers an area of about 75000 km^2 around the Concordia station. Cloudsat results are reported as a red curve in the Figure. From year 2011, the CPR on CloudSat collected data only in daylight hours due to a battery anomaly, so there is no record of cloud occurrence from CloudSat from

---

[1]World Meteorological Organization - http://www.wmo.int/pages/prog/sat/dataandproducts_en.php, last access: 21th January 2021.

April to August.

For each one of the MODIS, CALIOP and CPR sensors, the grid point that includes the Concordia station is used to retrieve the monthly L3 satellite product. Monthly time series of the cloud fractions, in the case of MODIS data, and of cloud occurrences, in case of CALIOP and CPR observations, are computed for the period 2012-2015. Results are compared with the cloud occurrence derived by the CIC algorithm over the Concordia station (Figure 7.11). Since the L3 products of the three sensors refers to multiple extent areas of observations (of the order of tens of thousands of km^2), some differences are expected not only between the ground-based observations analysed by CIC but also among the mean values of the L3 satellite products.

In months in which insolation is present, the lowest values are those derived from CALIOP products, in green in Figure 7.11. Despite the very low values, CALIOP is able to identify the maximum of cloud occurrence during the austral winter (specifically August) also detected by the CIC algorithm applied to the REFIR-PAD data. In August, a maximum in the cloud fraction is also observed in the MODIS mean L3 combined product (MCD06COSP) even if the very high value (close to 100%) looks unrealistic and biased by the observational conditions. The MODIS MCD06COSP cloud mask performance is degraded by the low insolation reaching the region around Dome C during August. This is observed also in April through September. In the same period the MODIS MYD08 and MOD08 products provides very low values of cloud fraction likely due to the low efficiency of the cloud mask algorithm based on infrared bands only.

For the months from November to March (the warm season), the CIC cloud occurrence is comparable to that found by MODIS (both MCD06COSP and MYD08/MOD08) and the CPR sensors. Nevertheless, higher percentage of cloudiness is found by the CIC algorithm with respect to the CPR. The reason is likely that the CIC is very sensitive to the identification of optically thin ice clouds which are often present in the Antarctic Plateau (Maestri et al., 2019a) and that are missed by radar measurements (Henderson et al., 2013; L'Ecuyer et al., 2008).

7.3.4 Diurnal variability of cloud occurrence

The almost continuous REFIR-PAD measurements, during the four year period provide an opportunity to investigate an hourly mean cloud occurrence. The time collocation of each CIC classification is obtained by associating each spectra to the hourly time of observation. For instance, observations performed between 1:00:00 UTC and 1:59:59 UTC are associated to the time 1:00:00 UTC. For each hour, the percentage of occurrence of each class is computed and results are reported in Figure 7.12. Results are also presented as seasonal means.

In the austral summer (upper left panel of Figure 7.12), a diurnal cycle is observed and related to the hourly mean insolation, also reported in the same Figure with a black dotted curve. The clear sky occurrence is characterized by a maximum value of about
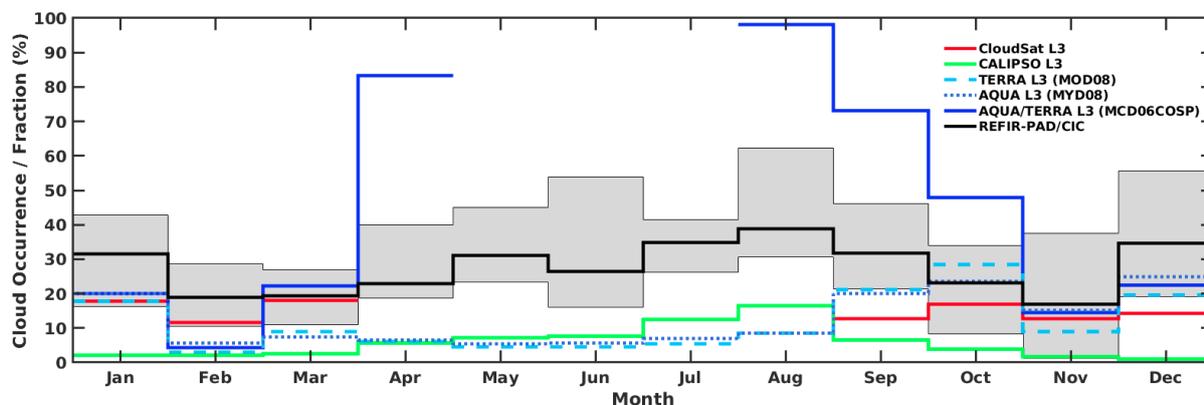
Figure 7.11: Percentage fraction of CIC monthly mean cloud occurrence (in black) compared with CloudSat L3 product (red line), CALIPSO L3 product (green line), and MODIS L3 products (solid blue line for combined AQUA and TERRA L3 product - MCD06COSP, dashed blue line for TERRA L3 product – MOD08, and dotted blue line for AQUA L3 product - MYD08). The shaded grey area indicates the minimum and maximum CIC monthly values in the interval 2012–2015.

78% at around 5:00 UTC (13:00 local time). This maximum is very close in time to the maximum of insolation for the same period of the year. In the summer season, the highest percentage of occurrence of cloudiness (about 36%) is obtained during nighttime hours, that correspond to the coldest time of the day. For the other seasons, a clear diurnal cycle of the percentage occurrences is not observed. Note that for the fall and spring seasons the daily variation of the insolation is much less intense that for austral summer, and in winter is almost null. In the austral autumn (MAM, panel b) and spring (SON, panel d) seasons the clouds are almost entirely composed of ice since mixed-phase clouds are very rare. In the austral winter (JJA, panel c) the insolation is close to zero and the ice cloud occurrence reaches its seasonal maximum.

In Figure 7.13, the hourly mean surface air temperature is plotted for the four seasons for clear sky, ice clouds, and mixed-phase clouds. The hourly mean temperatures are also presented for all-sky conditions (magenta line) and the hourly mean top of the atmosphere insolation (dashed black curve). The all-sky hourly mean air surface temperature is driven by the diurnal cycle of insolation in summer and spring: a lag of about two hours is observed between the maximum in insolation and the maximum in temperature. The all-sky surface air temperature has a 11.2°C amplitude in summer, when the top of the atmosphere diurnal cycle of insolation is the largest. This amplitude decreases as the insolation cycle becomes weaker and it is almost null in winter.

The surface mean air temperature is higher in cloudy sky conditions (ice cloud or mixed-phase cloud) than for clear sky at all hours of the day, suggesting a positive cloud forcing at the surface level. Mean values of surface air temperature are higher in presence of mixed-phase clouds than ice clouds at all times of the day. Observations of
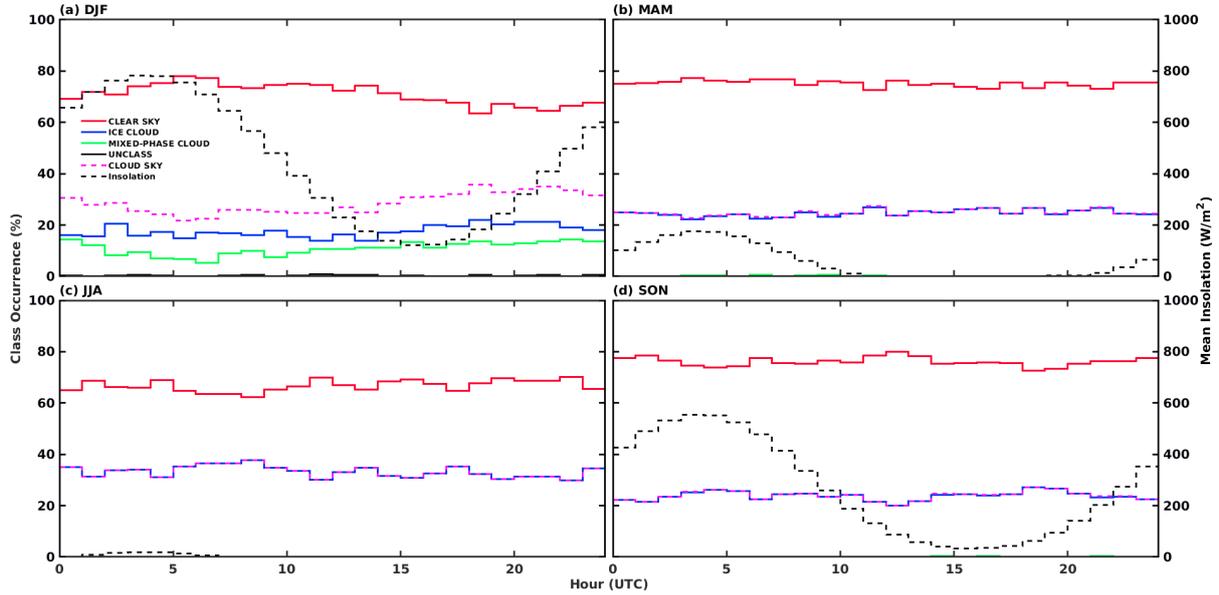
Figure 7.12: Hourly mean cloud occurrence of clear sky (red lines), ice clouds (blue lines), and mixed-phase clouds (green line). Unclassified spectra are in black and total cloud occurrence in magenta. Percentages of occurrence are provided for each season: (a) DJF, (b) MAM, (c) JJA, and (d) SON. The top of the atmosphere hourly mean insolation is also reported in Wm^{-2} (black dashed line), referring to values reported on the right ordinate axis. Local time is UTC+8.

mixed-phase clouds (green curves) are rare in autumn and spring and the data do not cover the full day in these seasons. Note that when mixed-phase clouds are present, the daily thermal amplitude is smoothed with respect to the other sky conditions. Potential explanations for this are this could be due to the decrease in surface insolation caused by the liquid phase clouds or perhaps caused by the larger cloud forcing of mixed phase clouds in comparison to ice clouds (Natale et al., 2020). The hourly mean surface temperature is larger when ice clouds are present than in clear sky conditions. This difference is, on average, larger in winter (about 9°C) and autumn (about 7°C), diminishes in summer (about 4°C) and becomes very small in spring (about 1°C). The cause of this low value needs further investigation. Possible explanations could be related to the optical thickness and position of the clouds and/or related to the circulation of the air in the area that is not accounted for in this analysis. For the spring and summer seasons, where the insolation diurnal cycle is larger, the surface temperature difference is greater between the clear sky and ice cloud conditions for low insolation but decreases for higher insolation.
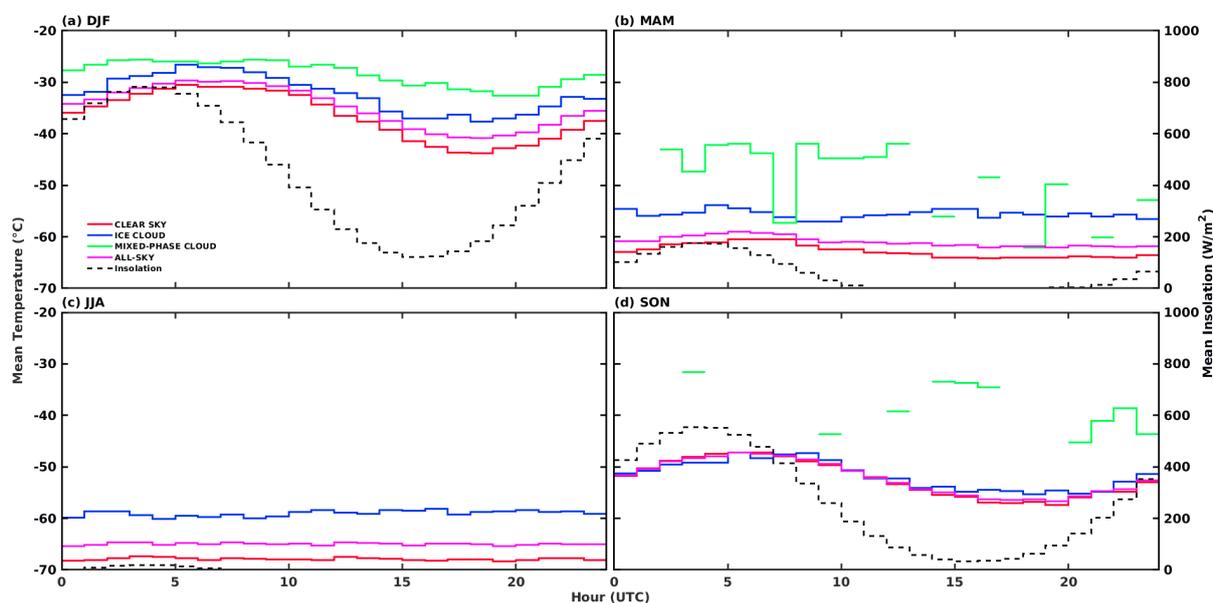
Figure 7.13: Hourly mean surface air temperature, according to the sky condition: clear sky (red line), ice cloud (blue line), mixed-phase cloud (green line), and all-sky (magenta line). The temperature is reported according to the season: (a) DJF, (b) MAM, (c) JJA, and (d) SON. The top of the atmosphere hourly mean insolation is also reported in Wm^{-2} (black dashed line), and values indicated by the ordinate axis on the right. Local time is UTC+8.

7.4 Summary

High spectral resolution downwelling radiances at FIR and MIR wavelengths are measured by the REFIR-PAD spectroradiometer located at Dome C on the Antarctic Plateau between 2012-2015. The radiance dataset comprising spectral measurements are ingested by the CIC algorithm to perform single spectrum classification. As previously discussed, CIC is developed to identify high spectral resolution observations and, in case of cloudy scene, to perform a classification. The algorithm is computationally very fast and only requires a limited number of spectra as training set, which makes it very flexible, efficient, user-friendly, and easy to adapt to different types of sensors. In this chapter, the algorithm is set-up to classify a REFIR-PAD spectrum as being a clear sky, ice cloud, or mixed-phase cloud. Mixed-phase clouds in the Antarctic Plateau are usually composed of a liquid water layer at cloud base and an ice layer close to the cloud top.

While an accurate description of clear and cloud properties is quite difficult in the Antarctic from passive measurements alone, this analysis of the REFIR-PAD data is greatly enhanced through coincident active measurents of atmospheric backscattering and depolarization ratio profiles measured by a LiDAR system that is temporally co-located with the REFIR-PAD radiance measurements. The coincident LiDAR and REFIR-PAD measurements are used to obtain accurate training sets for the CIC algo-

rithm. The training sets are formed by using a total of 202 spectra that are sufficient to characterize the large variability of the atmospheric conditions in the Antarctic Plateau region. An analysis of the LiDAR data and atmospheric vertical profiles of temperature and humidity, obtained from radiosondes launched every day at the Concordia station, is used to separate the training sets in two macro seasons. The first is named the warm season and ranges between November and March. Three training sets, defining three different classes of spectra, are considered for the warm season: clear sky, ice cloud, and mixed-phase cloud. The second macro-season is named the cold season and corresponds to the period from April to October. For the cold season, only two classes are considered (clear sky and ice cloud) since mixed-phase clouds are rarely observed during this period due to the extremely cold atmospheric temperatures.

A number of 1726 LiDAR co-located REFIR-PAD measurements are then used to select a test set of spectra, previously classified by visual inspection using the LiDAR backscatter and depolarization ratio vertical profiles. This sample is used to test the algorithm performance, to estimate the CIC classification uncertainty, and to optimize the classification results for each class. For the optimization process, the CIC algorithm is applied to classify the test set by considering different spectral intervals. A weighted Threat Score (ThS) is used to select the optimal spectral range for the classification. Results show that the spectral interval 380-1000 cm^{-1} provides the best score due to the experimental and observational conditions. This result highlights the fundamental role of the FIR part of the spectrum to improve the process of clear/cloud identification and cloud type classification in the Antarctic.

The optimized CIC algorithm is then applied to the entire REFIR-PAD dataset from 2012 to 2015 consisting of 87758 spectra.

On average, clear sky conditions are detected in almost 72% of the cases with an associated uncertainty of the order of 1.5%. The ice cloud occurrence is about 25% and the mixed-phase clouds are identified in less than 3% of the observations. The uncertainty is 0.3% in cloudy conditions. The cloud occurrence over the Antarctic Concordia station is analysed at different temporal scales: inter-annual, seasonal, monthly, and daily variability. The inter-annual variability of total cloud occurrence spans between about 23 and 31%. A positive correlation is observed between mean air temperatures at surface level, in the warm macro-season, and the occurrence of mixed-phase clouds. This result suggests that (a) warm temperatures are favorable for the mixed-phase clouds formation or that (b) the occurrence of warm cloud layers enhances the cloud radiative forcing at the surface with a consequent increase in the surface temperature. Further work will help to further our understanding of mixed phase clouds in the Antarctic.

Seasonal analysis indicate that the mean total cloud occurrence varies from 23.9% in the spring (SON) to 33.2% in the cold winter season (JJA) when only ice clouds are present. In fact, most of the mixed-phase clouds are observed in summer season, where they amount to more than one third of the total clouds over Concordia station. The seasonal scene classification is analysed in accordance with meteorological parameters.

Results show that highest values of surface air temperature (and relative humidity) are found in correspondence of the highest amounts of cloud for the summer, fall, and winter seasons; in the spring this relationship is minimal. The influence of the longwave radiative forcing of ice clouds on surface temperature is most observed in the winter months where the insolation is negligible. For this season, the mean surface temperature is about $-68°C$ in clear sky and $-59°C$ in presence of clouds. Furthermore, surface level winds from the South and South-West are more frequently observed in clear sky conditions, while in presence of ice clouds the surface wind is primarily from the South-East. When mixed-phase clouds are identified, surface winds from the East quadrant are more frequent. The mean wind intensity is about 2 ms^{-1} higher in presence of ice clouds than in clear atmospheric conditions.

CIC monthly mean cloud occurrences show, on average, a maximum in August and a minimum in November. The inter-annual variability of monthly mean cloud occurrences can be very high. Noteworthy is the November case that registers a cloud occurrence variation spanning from 0 to 40% among the four years of analysis. The monthly mean data are compared with Level-3 gridded satellite products derived from the MODIS (passive imager), CALIOP (LiDAR) and CPR (RADAR) sensors. Some differences are observed among the analyzed products. In periods of intense insolation the lowest values of monthly cloud occurrence are those derived from CALIOP. Despite the low scores, CALIOP collocates the maximum cloud occurrence in winter (August) similarly to what derived by the CIC algorithm. For the months from November to March, which correspond to the warm season, the CIC cloud occurrence is larger but comparable to what is found by the MODIS and CPR sensors. The higher values detected by the CIC are probably due to the greater sensitivity of the algorithm for the identification of thin cirrus clouds and cloud layers near the surface from MIR and FIR data. The added value of both the local and continuous measurements is demonstrated. The CIC results, by exploitation of REFIR-PAD FIR and MIR spectral data available at all times during the year, provide a continuous record of cloud occurrence with excellent classification scores.

Finally, a hourly cloud occurrence analysis is performed that shows the presence of a diurnal cycle with maximum of about 36% and minimum of 22% during the austral summer that follows the hourly mean insolation. The highest cloud occurrences are observed during nighttime hours. Conversely, the season maximum of clear sky occurrence is observed in correspondence to the local noon time. For all the other seasons, diurnal cycles are not observed for either cloud or clear sky conditions. An analysis between the daily sky condition and the surface mean air temperature reveals higher surface temperatures in cloudy sky conditions, especially for mixed-phase clouds, than in clear sky for all the seasons and hours of the day. In summer, the mean surface air temperature in the presence of clouds is on average 5°C warmer than in clear sky. This difference is larger during the night but it is reduced during the day probably due to the amount of insolation. The same effect, although smaller, is observed in fall and spring due to a weaker insolation cycle. In the winter, where the insolation is almost null, the differ-

ence between surface air temperature measured in cloudy sky and clear sky is constant at about 9°C throughout the day, which quantifies the effect of the longwave radiative forcing of the antarctic winter clouds.

The results of this chapter provide a basis for understanding of cloud occurrence at different time scales on the Antarctic Plateau where cloud identification and classification from satellites is challenging. The obtained results provide a useful benchmark for satellite and model product comparisons and open the path to new investigations.

Chapter 8

Concluding Remarks

FORUM was selected as the ninth ESA Earth Explorer Mission, which is a research program devoted to innovative measurement techniques for the exploration and the understanding of different aspects of the Earth system. FORUM mission will measure outgoing radiances in the FIR part of the spectrum, which has never been observed from space, spectrally resolved and in its entirety. These measurements will allow to evaluate the role of the FIR in the ERB, and will improve our knowledge of the climate system and the atmospheric processes governing it.

In the framework of the FORUM mission, this thesis aimed at studying the FIR spectral radiances in order to:

- characterize the radiance signal in presence of ice clouds, especially cirrus clouds;

- set-up an algorithm for the detection of optically thin and sub-visible ice-clouds, and to classify clouds in general;

- contribute in the retrieval of ice-cloud optical, microphysical and geometric properties.

In this context, sensitivity tests are performed to investigate the radiance signal at FIR wavelengths and are functional for the definition of the information content concerning atmospheric and cloud properties in the FIR part of the spectrum.

Results from the sensitivity tests highlight that in the presence of the cirrus cloud distinctive spectral features are observed. Although both MIR and FIR are sensible to the cloud thermodynamics, geometrical and optical properties, the radiance sensitivity to the particle water phase and ice habit is found to be larger in the FIR than in the MIR. These features are mainly associated with the local minimum of the refractive index imaginary part, which makes the scattering process more significant at around 410 cm^{-1} for ice clouds.

It is noted that the combined use of spectrally resolved radiances at both FIR and MIR wavelengths would allow better performances in the identification of clouds and their

classification. Thereby, an innovative machine learning algorithm (CIC) is developed and applied for the identification and classification of clouds.

Machine learning techniques, based on principal component analysis, artificial neural networks, or support vector machines, among others, have been extensively applied to identify clouds from high resolution satellite imagers. Most of these techniques exploit data spanning from infrared to shortwave, which limits their applicability to daytime hours only. It is less common to find cloud detection algorithms based on hyperspectral infrared sounders. The existing ones are generally applied by selecting specific spectral channels and/or evaluating brightness temperature thresholds and brightness temperature differences, which can limit their applicability and require elaborate preliminary calibrations. They are also commonly employed in conjunction with high spatial resolution imagers, which implies a relatively high computational cost due to the necessity to combine two different types of measurements.

In this sense, the CIC algorithm presents some advantages in comparison with other ones. CIC is a machine learning algorithm based on PCA, which performs the cloud identification and classification by only using high spectral resolution radiances (either simulated or observed), without using any additional ancillary or model data. While most supervised machine learning techniques use a metric to establish the distance between the analyzed element and the reference training sets, the CIC algorithm exploits the PCA analysis to extract the information content of the reference training sets and their changes after the addition of the analyzed element. This allows to rely on a computationally very fast algorithm and requires a limited number of spectra as training sets, which makes it very flexible, efficient, user-friendly, and easy to adapt to different types of sensors at infrared wavelengths.

CIC is initially tested against a large synthetic dataset, computed to simulate high spectral resolution radiances from satellite, as those observed by the FORUM mission. Cloud detection tests are performed both ingesting the MIR part of the spectrum only or the FIR and MIR bands, so that the detection performances could allow an evaluation of far-infrared channels information content. The results highlight that the combined use of FIR and MIR provides a better classification of the FORUM simulated spectra all over the globe.

The CIC algorithm is then applied to a dataset of airborne high spectral resolution infrared measurements, in order to evaluate their capability to identify clear and cloudy sky spectra in real measurements conditions. The classification results show that either employing the MIR spectral region only or a combination of far- and mid-infrared channels, all the aircraft measured spectra are correctly classified by the CIC code. However, a detailed analysis of the classification results shows that the combined use of FIR and MIR allows a wider sensitivity of the algorithm to possible variability of the experimental conditions. At this regard a theoretical sensitivity study is brought on, which exploits simulated radiances for cirrus clouds thinner than those observed during the aircraft field campaigns. Results show that the performance of the CIC algorithm improves with

cloud altitude and also show that, for conditions encompassing those sampled by the flight campaign, the additional information contained within the far-infrared improves the algorithm's performance when compared to using mid-infrared data only. Therefore, the CIC is used to show that the exploitation of radiance in the FIR will allow the detection of thinner cirrus clouds and thus reduce the error on retrieved atmospheric and surface parameters from IR sounding sensors.

This thesis also presents a brief introduction about the FE2ES, a tool to predict and assess the Level-1 and Level-2 performances of the FORUM mission. This tool is used in the definition of parametric, scientific, and industrial studies related to FORUM. Moreover, it is also used to understand and quantify the contribution of the FIR part of the spectrum in atmospheric retrievals. As an example, a test performed in this regarding shows that the use of FIR part of the spectrum in the CIC can reduce the error on the surface temperature retrieved by the FE2ES, since the cloud detection improves when FIR wavenumbers are used.

Finally, the algorithm is applied to a large and unique dataset of downwelling spectral radiances measured at the Concordia station, Dome C, in the middle of the Antarctic Plateau, between 2012 and 2015. The continuous measurements, performed by means of the REFIR-PAD Fourier transform spectroradiometer, allow to evaluate the CIC performances, besides developing a punctual statistic on clear/cloud occurrence and investigate the diurnal cycle and seasonality of ice and mixed-phase clouds in the Antarctic Plateau. The results highlight the importance of a proper algorithm calibration in accordance with the analysed sensor and the fundamental role of the FIR part of the spectrum in cloud identification. Averagely, the clear sky is detected in almost 72% of the spectra classified, while the ice cloud occurrence is about 25% and the mixed-phase clouds result in less of the 3% of the observations. The cloud occurrence over the Antarctic Concordia station is analysed in different temporal scales (inter-annual, seasonal, monthly, and daily variability). Seasonal and monthly analyses indicate a maximum of cloud occurrence during the austral winter. Monthly statistics are also compared with satellite L3 products of MODIS, CALIOP, and CPR sensors, which allows highlighting the CIC capabilities. Furthermore, the cloud occurrences results are correlated with meteorological variables measured in the Antarctic Plateau. Positive correlations between mean temperatures at surface level and the cloud occurrence are identified in different time scales. These correlations point out some applications of the CIC algorithm, that contributes to performing different analyses type. The CIC classification results from the REFIR-PAD dataset are also used to retrieve Antarctic cloud geometrical, microphysical, and optical features in the work developed by Natale et al. (2020).

In conclusion, the studies presented in this work showed the potentiality and the importance of the exploitation of spectral radiance measurements in the FIR part of the spectrum for clear/cloud identification and for cloud classification.

Acronyms

AER Atmospheric and Environmental Research

AIRS Atmospheric InfraRed Sounder

ARIES Airborne Research Interferometer Evaluation System

AWS Automatic Weather Station

BT Brightness Temperature

CALIOP Cloud-Aerosol Lidar with Orthogonal Polarization

CALIPSO Cloud-Aerosol Lidar and Infrared Pathfinder Satellite Observation

CERES Clouds and the Earth's Radiant Energy System

CESM Community Earth System Model

CIC Cloud Identification and Classification

CIRCCREX Cirrus Coupled Cloud-Radiation Experiment

CoI Consistency Index

CoMPASs Concordia Multi-Process Atmospheric Studies

CPR Cloud Profiling Radar

CrIS Cross-track Infrared Sounder

CSID Corrected Similarity Index Difference

DISORT Discrete Ordinates Radiative Transfer

DP Detection Performance

ERA ECMWF Re-Analysis

ESA European Space Agency

ERB Earth's Radiation Budget

ETS Extended Training Set

ϵETS Extended Training Set Eigenvector Matrix

ϵTS Training Set Eigenvector Matrix

FAAM Facility for Airborne Atmospheric Measurements

FE2ES FORUM end-to-end simulator

FEI FORUM embedded imager

FIR Far InfraRed

FN False Negative

FORUM Far-infrared Outgoing Radiation Understanding and Monitoring

FoV Field of View

FSI FORUM sounding instrument

GM Geometry Module

GCM Global Climate Models

HITRAN High-resolution transmission molecular absorption

HR Hit Rate

IASI Infrared Atmospheric Sounding Interferometer

IBECs Information-BEaring principal Components

IG2 Initial Guess Climatological Database no. 2

ILS Instrument Line Shape

IND Indicator function

IPEV French Polar Institute Paul-Émile Victor

ISCCP International Satellite Cloud Climatology Project

IWC Ice Water Content

L2M Level-2 Module

LiDAR Light Detection and Ranging

LTE Local Thermodynamic Equilibrium

LWC Liquid Water Content

MERIS Medium Resolution Imaging Spectrometer Instrument

MIDWIN Mid-latitude Winter

MIR Mid InfraRed

MODIS Moderate Resolution Imaging Spectroradiometer

NESR Noise Equivalent Spectral Radiance

OD Optical Depth

OLR Outgoing Long Wave Radiation

OSS Observing System Simulator

PAM Performance Assessment Module

PCs Principal Components

PCA Principal Component Analysis

PNRA Italian National Program for Research in Antarctica

POLSUM Polar Summer

POLWIN Polar Winter

PRANA Radiative Properties of Water Vapor and Clouds in Antarctica

PSD Particle Size Distribution

REFIR-PAD Radiation Explorer in the Far InfraRed: Prototype for Applications and Development

RETS Reference Training Set

RT Radiative Transfer

RTE Radiative Transfer Equation

SGM Scene Generator Module

SI Similarity Index

SID Similarity Index Difference

SLR Straight and Level Run

SNR Signal to Noise Ratio

SUCCESS Subsonic Aircraft: Contrail and Cloud Effects Special Study

TAFTS Tropospheric Airborne Fourier Transform Spectrometer

ThS Threat Score

TOA Top of Atmosphere

TP True Positive

TraNC Training set Number Configuration

TS Training Set

UTLS Upper Troposphere/Lower Stratosphere

WMO World Meteorological Organization

Bibliography

Adhikari, L., Wang, Z., and Deng, M. (2012). Seasonal variations of antarctic clouds observed by CloudSat and CALIPSO satellites. *Journal of Geophysical Research: Atmospheres*, 117(D4):n/a–n/a. ↑28, ↑102, ↑115

Ahmad, A. and Quegan, S. (2012). Cloud masking for remotely sensed data using spectral and principal components analysis. *Engineering, Technology and Applied Science Research*, 2(3):221–225. ↑39

Ahrens, C. (2011). *Essentials of meteorology : an invitation to the atmosphere.* Brooks/Cole, Belmont, CA. ↑5

Alvarado, M. J., Payne, V. H., Mlawer, E. J., Uymin, G., Shephard, M. W., Cady-Pereira, K. E., Delamere, J. S., and Moncet, J.-L. (2013). Performance of the line-by-line radiative transfer model (LBLRTM) for temperature, water vapor, and trace gas retrievals: recent updates evaluated with IASI case studies. *Atmospheric Chemistry and Physics*, 13(14):6687–6711. ↑15

Bantges, R., Russell, J., and Haigh, J. (1999). Cirrus cloud top-of-atmosphere radiance spectra in the thermal infrared. *Journal of Quantitative Spectroscopy and Radiative Transfer*, 63(2-6):487–498. ↑7

Bantges, R. J., Brindley, H. E., Murray, J. E., Last, A. E., Russell, J. E., Fox, C., Fox, S., Harlow, C., O'Shea, S. J., Bower, K. N., Baum, B. A., Yang, P., Oetjen, H., and Pickering, J. C. (2020). A test of the ability of current bulk optical models to represent the radiative properties of cirrus cloud across the mid- and far-infrared. *Atmospheric Chemistry and Physics*, 20(21):12889–12903. ↑7

Baran, A. J. (2005). The dependence of cirrus infrared radiative properties on ice crystal geometry and shape of the size-distribution function. *Quarterly Journal of the Royal Meteorological Society*, 131(607):1129–1142. ↑7

Baran, A. J. (2009). A review of the light scattering properties of cirrus. *Journal of Quantitative Spectroscopy and Radiative Transfer*, 110(14-16):1239–1260. ↑6

Baran, A. J. (2012). From the single-scattering properties of ice crystals to climate prediction: A way forward. *Atmospheric Research*, 112:45–69. ↑7

Bellisario, C., Brindley, H. E., Murray, J. E., Last, A., Pickering, J., Harlow, R. C., Fox, S., Fox, C., Newman, S. M., Smith, M., Anderson, D., Huang, X., and Chen, X. (2017). Retrievals of the Far Infrared Surface Emissivity Over the Greenland Plateau Using the Tropospheric Airborne Fourier Transform Spectrometer (TAFTS). *Journal of Geophysical Research (Atmospheres)*, 122(22):12,152–12,166. ↑86

Bianchini, G., Castagnoli, F., Di Natale, G., and Palchetti, L. (2019). A fourier transform spectroradiometer for ground-based remote sensing of the atmospheric downwelling long-wave radiance. *Atmospheric Measurement Techniques*, 12:619–635. ↑102

Bianchini, G., Palchetti, L., and Carli, B. (2006). A wide-band nadir-sounding spectroradiometer for the characterization of the earth's outgoing long-wave radiation. In Meynart, R., Neeck, S. P., and Shimoda, H., editors, *Sensors, Systems, and Next-Generation Satellites X*. SPIE. ↑85

Bloom, H. J. (2001). The cross-track infrared sounder (cris): a sensor for operational meteorological remote sensing. In *IGARSS 2001. Scanning the Present and Resolving the Future. Proceedings. IEEE 2001 International Geoscience and Remote Sensing Symposium (Cat. No.01CH37217)*, volume 3, pages 1341–1343 vol.3. ↑94

Bontemps, S., Defourny, P., Van Bogaert, E., Arino, O., Kalogirou, V., and Ramos Perez, J. J. (2011). Globcover 2009 products description and validation report. Technical report, Université catholique de Louvain (UCL) and European Space Agency (esa). ↑72

Borbas, E. E., Seemann, S., Li, Z., Li, J., Kern, A., and Menzel, W. P. (2016). Modis atmosphere profiles product (07 l2). ↑73

Bromwich, D. H., Nicolas, J. P., Hines, K. M., Kay, J. E., Key, E. L., Lazzara, M. A., Lubin, D., McFarquhar, G. M., Gorodetskaya, I. V., Grosvenor, D. P., Lachlan-Cope, T., and van Lipzig, N. P. M. (2012). Tropospheric clouds in antarctica. *Reviews of Geophysics*, 50(1). ↑28, ↑101, ↑102

Canas, T. A., Murray, J. E., and Harries, J. E. (1997). Tropospheric airborne fourier transform spectrometer (TAFTS). In Haigh, J. D., editor, *Satellite Remote Sensing of Clouds and the Atmosphere II*. SPIE. ↑85

Carli, B. (1984). The high resolution submillimetre spectrum of the stratosphere. *Journal of Quantitative Spectroscopy and Radiative Transfer*, 32(5-6):397–405. ↑2

Ceccaldi, M., Delanoë, J., Hogan, R. J., Pounder, N. L., Protat, A., and Pelon, J. (2013). From CloudSat-CALIPSO to EarthCare: Evolution of the DARDAR cloud classification and its comparison to airborne radar-lidar observations. *Journal of Geophysical Research: Atmospheres*, 118(14):7962–7981. ↑116

Chahine, M. T., Pagano, T. S., Aumann, H. H., Atlas, R., Barnet, C., Blaisdell, J., Chen, L., Divakarla, M., Fetzer, E. J., Goldberg, M., Gautier, C., Granger, S., Hannon, S., Irion, F. W., Kakar, R., Kalnay, E., Lambrigtsen, B. H., Lee, S.-Y., Le Marshall, J., McMillan, W. W., McMillin, L., Olsen, E. T., Revercomb, H., Rosenkranz, P., Smith, W. L., Staelin, D., Larrabee Strow, L., Susskind, J., Tobin, D., Wolf, W., and Zhou, L. (2006). AIRS: Improving Weather Forecasting and Providing New Data on Greenhouse Gases. *Bulletin of the American Meteorological Society*, 87(7):911–926. ↑87, ↑94

Chan, M. A. and Comiso, J. C. (2011). Cloud features detected by MODIS but not by CloudSat and CALIOP. *Geophysical Research Letters*, 38(24):n/a–n/a. ↑102

Chan, M. A. and Comiso, J. C. (2013). Arctic cloud characteristics as derived from MODIS, CALIPSO, and CloudSat. *Journal of Climate*, 26(10):3285–3306. ↑102

Chandrasekhar, S. (1950). *Radiative transfer*. Dover Publications, New York. ↑11

Choi, Y.-S. and Ho, C.-H. (2006). Radiative effect of cirrus with different optical properties over the tropics in MODIS and CERES observations. *Geophysical Research Letters*, 33(21). ↑6

Chwolson, O. (1889). Grundzüge einer mathematischen theorie der inneren diffusion des lichtes. *Bull. Acad. Imp. Sci. St. Petersbourg*, 33:221–256. ↑10

Clough, S., Kneizys, F., and Davies, R. (1989). Line shape and the water vapor continuum. *Atmospheric Research*, 23(3-4):229–241. ↑2

Clough, S., Shephard, M., Mlawer, E., Delamere, J., Iacono, M., Cady-Pereira, K., Boukabara, S., and Brown, P. (2005). Atmospheric radiative transfer modeling: a summary of the AER codes. *Journal of Quantitative Spectroscopy and Radiative Transfer*, 91(2):233–244. ↑15, ↑26

Clough, S. A., Iacono, M. J., and Moncet, J.-L. (1992). Line-by-line calculations of atmospheric fluxes and cooling rates: Application to water vapor. *Journal of Geophysical Research*, 97(D14):15761. ↑14, ↑15

Clough, S. A., Kneizys, F. X., Rothman, L. S., and Gallery, W. O. (1981). Atmospheric spectral transmittance and radiance: Fascod1 b. In Fan, R. W., editor, *Atmospheric Transmission*. SPIE. ↑14, ↑15

Cox, C. J., Turner, D. D., Rowe, P. M., Shupe, M. D., and Walden, V. P. (2014). Cloud microphysical properties retrieved from downwelling infrared radiance measurements made at eureka, nunavut, canada (2006–09). *Journal of Applied Meteorology and Climatology*, 53(3):772–791. ↑102

Cox, C. V., Harries, J. E., Taylor, J. P., Green, P. D., Baran, A. J., Pickering, J. C., Last, A. E., and Murray, J. (2010). Measurement and simulation of mid- and far-infrared spectra in the presence of cirrus. *Quarterly Journal of the Royal Meteorological Society*, 136(648):718–739. ↑6, ↑7, ↑86

Dee, D. P., Uppala, S. M., Simmons, A. J., Berrisford, P., Poli, P., Kobayashi, S., Andrae, U., Balmaseda, M. A., Balsamo, G., Bauer, P., Bechtold, P., Beljaars, A. C. M., van de Berg, L., Bidlot, J., Bormann, N., Delsol, C., Dragani, R., Fuentes, M., Geer, A. J., Haimberger, L., Healy, S. B., Hersbach, H., Hólm, E. V., Isaksen, L., Kållberg, P., Köhler, M., Matricardi, M., McNally, A. P., Monge-Sanz, B. M., Morcrette, J.-J., Park, B.-K., Peubey, C., de Rosnay, P., Tavolato, C., Thépaut, J.-N., and Vitart, F. (2011). The ERA-interim reanalysis: configuration and performance of the data assimilation system. *Quarterly Journal of the Royal Meteorological Society*, 137(656):553–597. ↑26, ↑91

Delanoë, J. and Hogan, R. J. (2010). Combined CloudSat-CALIPSO-MODIS retrievals of the properties of ice clouds. *Journal of Geophysical Research*, 115. ↑116

Demmel, J., Dumitriu, I., and Holtz, O. (2007). Fast linear algebra is stable. *Numerische Mathematik*, 108(1):59–91. ↑59

Dessler, A. E., Zhang, Z., and Yang, P. (2008). Water-vapor climate feedback inferred from climate fluctuations, 2003–2008. *Geophysical Research Letters*, 35(20). ↑3

Di Natale, G., Palchetti, L., Bianchini, G., and Del Guasta, M. (2017). Simultaneous retrieval of water vapour, temperature and cirrus clouds properties from measurements of far infrared spectral radiance over the Antarctic Plateau. *Atmospheric Measurement Techniques*, 10(3):825–837. ↑3, ↑7, ↑102

Diem, M. (1948). Messungen der grosse von wolkenelementen ii. *Meteor. Rundschau*, 1:261–273. ↑22

Downing, H. D. and Williams, D. (1975). Optical constants of water in the infrared. *Journal of Geophysical Research*, 80(12):1656–1661. ↑iii, ↑20, ↑26

Droppo, J. G. and Napier, B. A. (2008). Wind direction bias in generating wind roses and conducting sector-based air dispersion modeling. *Journal of the Air & Waste Management Association*, 58(7):913–918. ↑118

Eresmaa, R. (2014). Imager-assisted cloud detection for assimilation of Infrared Atmospheric Sounding Interferometer radiances. *Quarterly Journal of the Royal Meteorological Society*, 140(684):2342–2352. ↑39

ESA (2019). Report for mission selection: Forum. Technical Report ESA-EOPSM-FORM-RP-3549, European Space Agency,. ↑vi, ↑1, ↑2, ↑3, ↑5, ↑28, ↑60, ↑63, ↑64, ↑76, ↑85, ↑92

Evans, K. and Stephens, G. (1991). A new polarized atmospheric radiative transfer model. *Journal of Quantitative Spectroscopy and Radiative Transfer*, 46(5):413–423. ↑14, ↑18, ↑19

Feldman, D. R., Collins, W. D., Pincus, R., Huang, X., and Chen, X. (2014). Far-infrared surface emissivity and climate. *Proceedings of the National Academy of Sciences*, 111(46):16297–16302. ↑5

Fox, C. (2015). *Far-Infrared Spectral Radiance Studies: Application to Water Vapour and Cirrus*. PhD thesis, Imperial College London. ↑20, ↑86, ↑88

Freudenthaler, V., Homburg, F., and Jäger, H. (1996). Optical parameters of contrails from lidar measurements: Linear depolarization. *Geophysical Research Letters*, 23(25):3715–3718. ↑105

Green, P. D., Newman, S. M., Beeby, R. J., Murray, J. E., Pickering, J. C., and Harries, J. E. (2012). Recent advances in measurement of the water vapour continuum in the far-infrared spectral region. *Philosophical Transactions of the Royal Society of London Series A*, 370(1968):2637–2655. ↑86

Hahn, C. J., Rossow, W. B., and Warren, S. G. (2001). ISCCP cloud properties associated with standard cloud types identified in individual surface observations. *Journal of Climate*, 14(1):11–28. ↑28

Hansen, J. E. (1971). Multiple scattering of polarized light in planetary atmospheres part II. sunlight reflected by terrestrial water clouds. *Journal of the Atmospheric Sciences*, 28(8):1400–1426. ↑21, ↑22

Harries, J., Carli, B., Rizzi, R., Serio, C., Mlynczak, M., Palchetti, L., Maestri, T., Brindley, H., and Masiello, G. (2008). The far-infrared earth. *Reviews of Geophysics*, 46(4). ↑1, ↑2, ↑3, ↑15

Hartmann, D. (1994). *Global physical climatology*. Academic Press, San Diego. ↑5, ↑6

Haynes, J. (2019). Cloudsat level 3 rmcp gridded data product. ↑121

Held, I. M. and Soden, B. J. (2000). Water vapor feedback and global warming. *Annual Review of Energy and the Environment*, 25(1):441–475. ↑4

Henderson, D. S., L'Ecuyer, T., Stephens, G., Partain, P., and Sekiguchi, M. (2013). A multisensor perspective on the radiative impacts of clouds and aerosols. *Journal of Applied Meteorology and Climatology*, 52(4):853–871. ↑122

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., Nicolas, J., Peubey, C., Radu, R., Schepers, D., Simmons, A., Soci, C., Abdalla, S., Abellan, X., Balsamo, G., Bechtold, P., Biavati, G., Bidlot, J., Bonavita, M., Chiara, G., Dahlgren, P., Dee, D., Diamantakis, M., Dragani, R., Flemming, J., Forbes, R., Fuentes, M., Geer, A., Haimberger, L., Healy, S., Hogan, R. J., Hólm, E., Janisková, M., Keeley, S., Laloyaux, P., Lopez, P., Lupu, C., Radnoti, G., Rosnay, P., Rozum, I., Vamborg, F., Villaume, S., and Thépaut, J.-N. (2020). The ERA5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society*. ↑66

Hong, Y. and Liu, G. (2015). The characteristics of ice cloud properties derived from CloudSat and CALIPSO measurements. *Journal of Climate*, 28(9):3880–3901. ↑6

Hong, Y., Liu, G., and Li, J.-L. F. (2016). Assessing the radiative effects of global ice clouds based on CloudSat and CALIPSO measurements. *Journal of Climate*, 29(21):7651–7674. ↑6

Huang, X., Chen, X., Zhou, D. K., and Liu, X. (2016). An observationally based global band-by-band surface emissivity dataset for climate and weather simulations. *Journal of the Atmospheric Sciences*, 73(9):3541–3555. ↑vii, ↑27, ↑66, ↑70, ↑72

Huang, Y., Leroy, S., Gero, P. J., Dykema, J., and Anderson, J. (2010). Separation of longwave climate feedbacks from spectral observations. *Journal of Geophysical Research*, 115(D7). ↑4

IPCC (2013). *Climate change 2013 : the physical science basis : Working Group I contribution to the Fifth assessment report of the Intergovernmental Panel on Climate Change.* Cambridge University Press, New York. ↑4, ↑5, ↑6

Kiehl, J. T. and Trenberth, K. E. (1997). Earth's Annual Global Mean Energy Budget. *Bulletin of the American Meteorological Society*, 78(2):197–197. ↑6, ↑101

King, J. C. and Turner, J. (1997). *Antarctic Meteorology and Climatology.* Cambridge University Press. ↑101

King, M., Kaufman, Y., Menzel, W., and Tanre, D. (1992). Remote sensing of cloud, aerosol, and water vapor properties from the moderate resolution imaging spectrometer (MODIS). *IEEE Transactions on Geoscience and Remote Sensing*, 30(1):2–27. ↑101

Kirchhoff, G. (1860). Ueber das verhältniss zwischen dem emissionsvermögen und dem absorptionsvermögen der körper für wärme und licht. *Annalen der Physik und Chemie*, 185(2):275–301. ↑10

Klaes, K. D., Cohen, M., Buhler, Y., Schlüssel, P., Munro, R., Luntama, J.-P., Engeln, A. V., Clérigh, E. Ó., Bonekamp, H., Ackermann, J., and Schmetz, J. (2007). An Introduction to the EUMETSAT Polar system. *Bulletin of the American Meteorological Society*, 88(7):1085. ↑87

Kurucz, R. L. (1994). Synthetic infrared spectra. In *Infrared Solar Physics*, pages 523–531. Springer Netherlands. ↑15

Lachlan-Cope, T. (2010). Antarctic clouds. *Polar Research*, 29(2):150–158. ↑28, ↑101

Laszlo, I., Stamnes, K., Wiscombe, W. J., and Tsay, S.-C. (2016). The discrete ordinate algorithm, DISORT for radiative transfer. In *Light Scattering Reviews, Volume 11*, pages 3–65. Springer Berlin Heidelberg. ↑17

Lavanant, L., Fourrié, N., Gambacorta, A., Grieco, G., Heilliette, S., Hilton, F. I., Kim, M. J., McNally, A. P., Nishihata, H., Pavelin, E. G., and Rabier, F. (2011). Comparison of cloud products within IASI footprints for the assimilation of cloudy radiances. *Quarterly Journal of the Royal Meteorological Society*, 137(661):1988–2003. ↑39

L'Ecuyer, T. S., Wood, N. B., Haladay, T., Stephens, G. L., and Stackhouse, P. W. (2008). Impact of clouds on atmospheric heating based on the r04 CloudSat fluxes and heating rates data set. *Journal of Geophysical Research*, 113. ↑122

Lerner, J. A., Weisz, E., and Kirchengast, G. (2002). Temperature and humidity retrieval from simulated Infrared Atmospheric Sounding Interferometer (IASI) measurements. *Journal of Geophysical Research (Atmospheres)*, 107(D14):4189. ↑94

Liou, K. (1986). Influence of cirrus clouds on weather and climate processes: A global perspective. *Monthly Weather Review*, 114(6):1167–1199. ↑6

Liou, K. (1992). *Radiation and cloud processes in the atmosphere : theory, observation, and modeling.* Oxford University Press, New York. ↑25

Liou, K. (2002). *An introduction to atmospheric radiation.* Academic Press, Amsterdam Boston. ↑iii, ↑2, ↑4, ↑5, ↑6, ↑12, ↑13, ↑101

Liou, K. and Yang, P. (2016). *Light scattering by ice crystals : fundamentals and applications.* Cambridge University Press, Cambridge. ↑105

Listowski, C., Delanoë, J., Kirchgaessner, A., Lachlan-Cope, T., and King, J. (2019). Antarctic clouds, supercooled liquid water and mixed phase, investigated with dardar: geographical and seasonal variations. *Atmospheric Chemistry and Physics*, 19(10):6771–6808. ↑102, ↑108, ↑115

Liu, C., Yang, P., Nasiri, S. L., Platnick, S., Meyer, K. G., Wang, C., and Ding, S. (2015). A fast visible infrared imaging radiometer suite simulator for cloudy atmospheres. *Journal of Geophysical Research: Atmospheres*, 120(1):240–255. ↑15

Lommel, E. (1889). Die photometrie der diffusen zurückwerfung. *Annalen der Physik*, 272(2):473–502. ↑10

Lubin, D., Chen, B., Bromwich, D. H., Somerville, R. C. J., Lee, W.-H., and Hines, K. M. (1998). The impact of antarctic cloud radiative properties on a GCM climate simulation. *Journal of Climate*, 11(3):447–462. ↑101

Luo, T., Zhang, W., Yu, Y., Feng, M., Duan, B., and Xing, D. (2019). Cloud detection using infrared atmospheric sounding interferometer observations by logistic regression. *International Journal of Remote Sensing*, 40(17):6530–6541. ↑39

Maestri, T., Arosio, C., Rizzi, R., Palchetti, L., Bianchini, G., and Guasta, M. D. (2019a). Antarctic ice cloud identification and properties using downwelling spectral radiance from 100 to 1,400 cm -1. *Journal of Geophysical Research: Atmospheres*, 124(8):4761–4781. ↑3, ↑7, ↑102, ↑122

Maestri, T., Cossich, W., and Sbrolli, I. (2019b). Cloud identification and classification from high spectral resolution data in the far infrared and mid-infrared. *Atmospheric Measurement Techniques*, 12(7):3521–3540. ↑7, ↑8, ↑25, ↑40, ↑75, ↑86

Maestri, T. and Rizzi, R. (2003). A study of infrared diabatic forcing of ice clouds in the tropical atmosphere. *Journal of Geophysical Research*, 108(D4). ↑7

Maestri, T., Rizzi, R., Tosi, E., Veglio, P., Palchetti, L., Bianchini, G., Girolamo, P. D., Masiello, G., Serio, C., and Summa, D. (2014). Analysis of cirrus cloud spectral signatures in the far infrared. *Journal of Quantitative Spectroscopy and Radiative Transfer*, 141:49–64. ↑7

Magurno, D., Cossich, W., Maestri, T., Bantges, R., Brindley, H., Fox, S., Harlow, C., Murray, J., Pickering, J., Warwick, L., and Oetjen, H. (2020). Cirrus cloud identification from airborne far-infrared and mid-infrared spectra. *Remote Sensing*, 12(13):2097. ↑7, ↑8, ↑86

Mahajan, S. and Fataniya, B. (2019). Cloud detection methodologies: variants and development—a review. *Complex & Intelligent Systems*. ↑39

Mahesh, A., Walden, V. P., and Warren, S. G. (2001). Ground-based infrared remote sensing of cloud properties over the antarctic plateau. part i: Cloud-base heights. *Journal of Applied Meteorology*, 40(7):1265–1278. ↑102

Malinowski, E. (2002). *Factor analysis in chemistry*. Wiley, New York. ↑43

Martin, G. M., Johnson, D. W., and Spice, A. (1994). The measurement and parameterization of effective radius of droplets in warm stratocumulus clouds. *Journal of the Atmospheric Sciences*, 51(13):1823–1842. ↑156

Masuda, K., Takashima, T., and Takayama, Y. (1988). Emissivity of pure and sea waters for the model sea surface in the infrared window regions. *Remote Sensing of Environment*, 24(2):313–329. ↑91

McNally, A. P. and Watts, P. D. (2003). A cloud detection algorithm for high-spectral-resolution infrared sounders. *Quarterly Journal of the Royal Meteorological Society*, 129(595):3411–3423. ↑39

Merrelli, A. (2012). *The Atmospheric Information Content of Earth's Far Infrared Spectrum*. PhD thesis, University of Winsconsin-Madison, Madison, Winsconsin. ↑2, ↑7

Merrelli, A. and Turner, D. D. (2012). Comparing information content of upwelling far-infrared and midinfrared radiance spectra for clear atmosphere profiling. *Journal of Atmospheric and Oceanic Technology*, 29(4):510–526. ↑3

Met Office (2019). Facility for Airborne Atmospheric Measurements; Natural Environment Research Council; Met Office (2019): FAAM C153 PIKNMIX-F flight: Airborne atmospheric measurements from core and non-core instrument suites on board the BAE-146 aircraft. Centre for Environmental Data Analysis, Jan 2019. https://catalogue.ceda.ac.uk/uuid/6a2bc7a1edc34650bd41e0f958cbd50a. ↑85

Mlawer, E. J., Payne, V. H., Moncet, J.-L., Delamere, J. S., Alvarado, M. J., and Tobin, D. C. (2012). Development and recent evaluation of the MT_CKD model of continuum absorption. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 370(1968):2520–2556. ↑2, ↑15, ↑26

Mlynczak, M. G., Johnson, D. G., Latvakoski, H., Jucks, K., Watson, M., Kratz, D. P., Bingham, G., Traub, W. A., Wellard, S. J., Hyde, C. R., and Liu, X. (2006). First light from the Far-Infrared Spectroscopy of the Troposphere (FIRST) instrument. *Geophysical Research Letters*, 33(7):L07704. ↑85

MODIS Atmosphere Science Team (2017). Modis/terra and modis/aqua aerosol cloud water vapor ozone monthly l3 global 1deg cmg. ↑121

MODIS Atmosphere Science Team (2020). Modis (aqua/terra) cloud properties level 3 monthly, 1x1 degree grid. ↑121

Natale, G. D., Bianchini, G., Guasta, M. D., Ridolfi, M., Maestri, T., Cossich, W., Magurno, D., and Palchetti, L. (2020). Characterization of the far infrared properties and radiative forcing of antarctic ice and water clouds exploiting the spectrometer-lidar synergy. *Remote Sensing*, 12(21):3574. ↑102, ↑124, ↑131

Nazaryan, H., McCormick, M. P., and Menzel, W. P. (2008). Global characterization of cirrus clouds using CALIPSO data. *Journal of Geophysical Research (Atmospheres)*, 113(D16):D16211. ↑7

Palchetti, L., Bianchini, G., Natale, G. D., and Guasta, M. D. (2015). Far-infrared radiative properties of water vapor and clouds in antarctica. *Bulletin of the American Meteorological Society*, 96(9):1505–1518. ↑102, ↑103, ↑104

Palchetti, L., Brindley, H., Bantges, R., Buehler, S. A., Camy-Peyret, C., Carli, B., Cortesi, U., Bianco, S. D., Natale, G. D., Dinelli, B. M., Feldman, D., Huang, X. L., C.-Labonnote, L., Libois, Q., Maestri, T., Mlynczak, M. G., Murray, J. E., Oetjen, H., Ridolfi, M., Riese, M., Russell, J., Saunders, R., and Serio, C. (2020). FORUM: unique far-infrared satellite observations to better understand how earth radiates energy to space. *Bulletin of the American Meteorological Society*, preprint(2020):1–52. ↑85, ↑102

Palchetti, L., Di Natale, G., and Bianchini, G. (2016). Remote sensing of cirrus cloud microphysical properties using spectral measurements over the full range of their thermal emission. *Journal of Geophysical Research (Atmospheres)*, 121(18):10,804–10,819. ↑3, ↑7, ↑102

Peixoto, J. and Oort, A. (1992). *Physics of climate*. American Institute of Physics, New York. ↑4

Peña, O. and Pal, U. (2009). Scattering of electromagnetic radiation by a multilayered sphere. *Computer Physics Communications*, 180(11):2348–2354. ↑20, ↑27, ↑71

Pickering, J. C., Fox, C., Murray, J. E., and Last, A. (2015). The cirrus coupled cloud-radiation experiment: Circcrex. In *Fourier Transform Spectroscopy and Hyperspectral Imaging and Sounding of the Environment*, page JM3A.14. Optical Society of America. ↑85, ↑88

Planck, M. and Masius, M. (1914). *The Theory of Heat Radiation*. Blakiston. ↑10

Platnick, S.and Ackerman, S. A., King, M. D., Meyer, K., Menzel, W. P., Holz, R. E., Baum, B. A., and Yang, P. (2015). Modis atmosphere l2 cloud product (06 l2). ↑72

Rathke, C., Fischer, J., Neshyba, S., and Shupe, M. (2002). Improving IR cloud phase determination with 20 microns spectral observations. *Geophysical Research Letters*, 29(8):50–1–50–4. ↑102

Raval, A. and Ramanathan, V. (1989). Observational determination of the greenhouse effect. *Nature*, 342(6251):758–761. ↑4

Raz, R. (2003). On the complexity of matrix product. *SIAM Journal on Computing*, 32(5):1356–1369. ↑59

Remedios, J. J., Leigh, R. J., Waterfall, A. M., Moore, D. P., Sembhi, H., Parkes, I., Greenhough, J., Chipperfield, M. P., and Hauglustaine, D. (2007). MIPAS reference atmospheres and comparisons to V4.61/V4.62 MIPAS level 2 geophysical data sets. *Atmospheric Chemistry & Physics Discussions*, 7(4):9973–10017. ↑26, ↑66, ↑91

Ricaud, P., Guasta, M. D., Bazile, E., Azouz, N., Lupi, A., Durand, P., Attié, J.-L., Veron, D., Guidard, V., and Grigioni, P. (2020). Supercooled liquid water cloud observed, analysed, and modelled at the top of the planetary boundary layer above dome c, antarctica. *Atmospheric Chemistry and Physics*, 20(7):4167–4191. ↑102, ↑108

Rizzi, R., Maestri, T., and Amorati, R. (2002). Quantitative role of far-infrared emission on diabatic forcing of the middle and upper troposphere in clear and cloudy conditions. In Larar, A. M. and Mlynczak, M. G., editors, *Optical Spectroscopic Techniques, Remote Sensing, and Instrumentation for Atmospheric and Space Research IV*. SPIE. ↑5, ↑25

Rizzi, R. and Mannozzi, L. (2000). Preliminary results on the planetary emission between 100 and 600 cm-1. Technical Report REFIR—Radiation Explorer in the Far Infrared, ENV4-CT6 0344, Eur. Comm. ↑7

Rodgers, C. D. (2000). *Inverse Methods for Atmospheric Sounding*. WORLD SCIENTIFIC. ↑77

Rossow, W. B. and Schiffer, R. A. (1999). Advances in understanding clouds from ISCCP. *Bulletin of the American Meteorological Society*, 80(11):2261–2287. ↑28

Rothman, L., Gordon, I., Babikov, Y., Barbe, A., Benner, D. C., Bernath, P., Birk, M., Bizzocchi, L., Boudon, V., Brown, L., Campargue, A., Chance, K., Cohen, E., Coudert, L., Devi, V., Drouin, B., Fayt, A., Flaud, J.-M., Gamache, R., Harrison, J., Hartmann, J.-M., Hill, C., Hodges, J., Jacquemart, D., Jolly, A., Lamouroux, J., Roy, R. L., Li, G., Long, D., Lyulin, O., Mackie, C., Massie, S., Mikhailenko, S., Müller, H., Naumenko, O., Nikitin, A., Orphal, J., Perevalov, V., Perrin, A., Polovtseva, E., Richard, C., Smith, M., Starikova, E., Sung, K., Tashkun, S., Tennyson, J., Toon, G., Tyuterev, V., and Wagner, G. (2013). The HITRAN2012 molecular spectroscopic

database. *Journal of Quantitative Spectroscopy and Radiative Transfer*, 130:4–50. ↑15, ↑26

Sassen, K., Wang, Z., and Liu, D. (2008). Global distribution of cirrus clouds from CloudSat/Cloud-Aerosol Lidar and Infrared Pathfinder Satellite Observations (CALIPSO) measurements. *Journal of Geophysical Research (Atmospheres)*, 113(D8):D00A12. ↑7

Sassen, K. and yu Hsueh, C. (1998). Contrail properties derived from high-resolution polarization lidar studies during SUCCESS. *Geophysical Research Letters*, 25(8):1165–1168. ↑105

Schuster, A. (1905). Radiation through a foggy atmosphere. *The Astrophysical Journal*, 21:1. ↑10

Schwarzschild, K. (1906). On the equilibrium of the sun's atmosphere. *Nachrichten Königlichen Gesellschaft Wissensch. Göttingen. Math.-Phys. Klasse*, 195:41–53. ↑10

Schwarzschild, K. (1914). Diffusion and absorption in the sun's atmosphere. *Sitzungsberichte Königlichen Preussichen Akad. Wissensch.*, pages 1183–1200. ↑10

Serio, C., Lubrano, A. M., Romano, F., and Shimoda, H. (2000). Cloud detection over sea surface by use of autocorrelation functions of upwelling infrared spectra in the 800–900-cmˆ-1 window region. *Applied Optics*, 39(21):3565. ↑39

Serio, C., Masiello, G., Esposito, F., Girolamo, P. D., Iorio, T. D., Palchetti, L., Bianchini, G., Muscari, G., Pavese, G., Rizzi, R., Carli, B., and Cuomo, V. (2008). Retrieval of foreign-broadened water vapor continuum coefficients from emitted spectral radiance in the h_2o rotational band from 240 to 590 cmˆ-1. *Optics Express*, 16(20):15816. ↑2

Sinha, A. and Harries, J. E. (1995). Water vapour and greenhouse trapping: The role of far infrared absorption. *Geophysical Research Letters*, 22(16):2147–2150. ↑2, ↑3

Solomon, S., Rosenlof, K. H., Portmann, R. W., Daniel, J. S., Davis, S. M., Sanford, T. J., and Plattner, G.-K. (2010). Contributions of stratospheric water vapor to decadal changes in the rate of global warming. *Science*, 327(5970):1219–1223. ↑4

Stamnes, K., Tsay, S.-C., Wiscombe, W., and Jayaweera, K. (1988). Numerically stable algorithm for discrete-ordinate-method radiative transfer in multiple scattering and emitting layered media. *Applied Optics*, 27(12):2502. ↑14

Stengel, M., Schlundt, C., Stapelberg, S., Sus, O., Eliasson, S., Willén, U., and Meirink, J. F. (2018). Comparing ERA-interim clouds with satellite observations using a simplified satellite simulator. *Atmospheric Chemistry and Physics*, 18(23):17601–17614. ↑x, ↑155, ↑156, ↑157, ↑158

Stengel, M., Stapelberg, S., Sus, O., Schlundt, C., Poulsen, C., Thomas, G., Christensen, M., Henken, C. C., Preusker, R., Fischer, J., Devasthale, A., Willén, U., Karlsson, K.-G., McGarragh, G. R., Proud, S., Povey, A. C., Grainger, R. G., Meirink, J. F., Feofilov, A., Bennartz, R., Bojanowski, J. S., and Hollmann, R. (2017). Cloud property datasets retrieved from AVHRR, MODIS, AATSR and MERIS in the framework of the cloud_cci project. *Earth System Science Data*, 9(2):881–904. ↑78

Stubenrauch, C. J., Rossow, W. B., Kinne, S., Ackerman, S., Cesana, G., Chepfer, H., Girolamo, L. D., Getzewich, B., Guignard, A., Heidinger, A., Maddux, B. C., Menzel, W. P., Minnis, P., Pearl, C., Platnick, S., Poulsen, C., Riedi, J., Sun-Mack, S., Walther, A., Winker, D., Zeng, S., and Zhao, G. (2013). Assessment of global cloud datasets from satellites: Project and database initiated by the GEWEX radiation panel. *Bulletin of the American Meteorological Society*, 94(7):1031–1049. ↑121

Sun, B., Kattawar, G. W., Yang, P., and Mlawer, E. (2017). An improved small-angle approximation for forward scattering and its use in a fast two-component radiative transfer method. *Journal of the Atmospheric Sciences*, 74(6):1959–1987. ↑14

Sun, Z. (2001). Reply to comments by greg m. McFarquhar on 'parametrization of effective sizes of cirrus-cloud particles and its verification against observations'. (october b, 1999, 125, 3037–3055). *Quarterly Journal of the Royal Meteorological Society*, 127(571):267–271. ↑157

Sun, Z. and Rikus, L. (1999). Parametrization of effective sizes of cirrus-cloud particles and its verification against observations. *Quarterly Journal of the Royal Meteorological Society*, 125(560):3037–3055. ↑157

Town, M. S., Walden, V. P., and Warren, S. G. (2005). Spectral and broadband longwave downwelling radiative fluxes, cloud radiative forcing, and fractional cloud cover over the south pole. *Journal of Climate*, 18(20):4235–4252. ↑101

Turner, D. D. (2005). Arctic mixed-phase cloud properties from AERI lidar observations: Algorithm and results from SHEBA. *Journal of Applied Meteorology*, 44(4):427–444. ↑14, ↑16

Turner, D. D., Ackerman, S. A., Baum, B. A., Revercomb, H. E., and Yang, P. (2003). Cloud phase determination using ground-based AERI observations at SHEBA. *Journal of Applied Meteorology*, 42(6):701–715. ↑14, ↑16

Turner, D. D., Knuteson, R. O., Revercomb, H. E., Lo, C., and Dedecker, R. G. (2006). Noise reduction of atmospheric emitted radiance interferometer (AERI) observations using principal component analysis. *Journal of Atmospheric and Oceanic Technology*, 23(9):1223–1238. ↑43

Turner, D. D., Tobin, D. C., Clough, S. A., Brown, P. D., Ellingson, R. G., Mlawer, E. J., Knuteson, R. O., Revercomb, H. E., Shippert, T. R., Smith, W. L., and Shephard, M. W. (2004). The QME AERI LBLRTM: A closure experiment for downwelling high spectral resolution infrared radiance. *Journal of the Atmospheric Sciences*, 61(22):2657–2675. ↑15

Veglio, P. (2013). *Global scale investigation of cirrus clouds properties using active and passive sensors*. PhD thesis, Alma Mater Studiorum – Università di Bologna. ↑6

Veglio, P. and Maestri, T. (2011). Statistics of vertical backscatter profiles of cirrus clouds. *Atmospheric Chemistry & Physics*, 11(24):12925–12943. ↑7, ↑28, ↑91

Verlinden, K. L., Thompson, D. W. J., and Stephens, G. L. (2011). The three-dimensional distribution of clouds over the southern hemisphere high latitudes. *Journal of Climate*, 24(22):5799–5811. ↑102

Walsh, J. E., Bromwich, D. H., Overland, J. E., Serreze, M. C., and Wood, K. R. (2018). 100 years of progress in polar meteorology. *Meteorological Monographs*, 59:21.1–21.36. ↑101

Wang, L., Zheng, Y., Liu, C., Niu, Z., Xu, J., Chen, W., and Jiang, R. (2019). Combination of AIRS Dual CO2 Absorption Bands to Develop an Ice Clouds Detection Algorithm in Different Atmospheric Layers. *Remote Sensing*, 12(1):6. ↑39

Warren, S. G. and Brandt, R. E. (2008). Optical constants of ice from the ultraviolet to the microwave: A revised compilation. *Journal of Geophysical Research*, 113(D14). ↑iii, ↑20, ↑26

Webb, M. J., Andrews, T., Bodas-Salcedo, A., Bony, S., Bretherton, C. S., Chadwick, R., Chepfer, H., Douville, H., Good, P., Kay, J. E., Klein, S. A., Marchand, R., Medeiros, B., Siebesma, A. P., Skinner, C. B., Stevens, B., Tselioudis, G., Tsushima, Y., and Watanabe, M. (2017). The cloud feedback model intercomparison project (cfmip) contribution to cmip6. *Geoscientific Model Development*, 10(1):359–384. ↑121

Wilks, D. (2011). *Statistical methods in the atmospheric sciences*. Elsevier/Academic Press, Amsterdam Boston. ↑49

Wilson, S. H. S., Atkinson, N. C., and Smith, J. A. (1999). The Development of an Airborne Infrared Interferometer for Meteorological Sounding Studies. *Journal of Atmospheric and Oceanic Technology*, 16(12):1912. ↑85, ↑87

Winker, D. (2018). Calipso lidar level 3 cloud occurrence data, standard v1-00. ↑121

Wylie, D., Jackson, D. L., Menzel, W. P., and Bates, J. J. (2005). Trends in global cloud cover in two decades of HIRS observations. *Journal of Climate*, 18(15):3021–3031. ↑6

Wylie, D. P., Menzel, W. P., Woolf, H. M., and Strabala, K. I. (1994). Four years of global cirrus cloud statistics using HIRS. *Journal of Climate*, 7(12):1972–1986. ↑6

Yang, P., Bi, L., Baum, B. A., Liou, K.-N., Kattawar, G. W., Mishchenko, M. I., and Cole, B. (2013). Spectrally consistent scattering, absorption, and polarization properties of atmospheric ice crystals at wavelengths from 0.2 to 100 μm. *Journal of the Atmospheric Sciences*, 70(1):330–347. ↑iii, ↑21, ↑22, ↑27, ↑71, ↑91

Appendix 1

In the case of the SGM's automatically generated clouds, the cloud properties are derived from the ERA 5 dataset. How this dataset does not provide any information about the cloud position within its grid point, it is considered that the cloud covers a sector of the scene positioned always from the western area of the FSI FoV.

The variable "Total Cloud Cover", from ERA 5 surface parameters, is used to define the SGM input parameter regard to the sector of the FSI FoV covered by the cloud. The other SGM cloud parameters are defined automatically from the specific cloud liquid and ice water content in the ERA 5 model levels vertical profiles. Since these ERA 5 model levels variables represent an integral value over the full ERA 5 grid point area, it is needed to divide them by the "Fraction of Cloud Cover", which is also an ERA 5 model level variable, to obtain the real in-cloud Liquid Water Content (LWC) and Ice Water Content (IWC) at each model level. This pre-processing step is the same made by Stengel et al. (2018).

The number of cloud types and the cloud(s) phase are determined by the presence of the LWC and IWC. If both water phases exist then the number of cloud types is assumed equal 2, while it is equal 1 for the case in which just one phase is identified in the vertical profiles. Moreover, in presence of ice clouds, i.e., if there is IWC in the ERA 5 vertical profiles, the ice particle shape is assumed as aggregate of pristine columns.

The cloud top height and geometrical thickness are assumed to span around the maximum value of the vertical profile of the LWC and IWC, considering the vertical levels of the SGM. This simplification is necessary, since the SGM assumes that a cloud type is vertically continuous and homogeneous in terms of its microphysical parameters (Cloud Optical Depth and Cloud Effective Radius).

In Figure A1.1 are shown four examples of the cloud layers (blue and orange boxes) with the LWC/IWC vertical profiles. It is possible to note that if there are two layers of the same water phase cloud in the vertical profile, just the layer where the maximum value has been found is considered (Figure A1.1d). On other hand, if both liquid water and ice phases are found in the same vertical profiles, it is considered that two clouds exist simultaneously in that profile and used to determine the cloud top height and geometrical thickness to each water phase.

In Figure A1.2 an example of the geolocation distribution of the cloud top height and
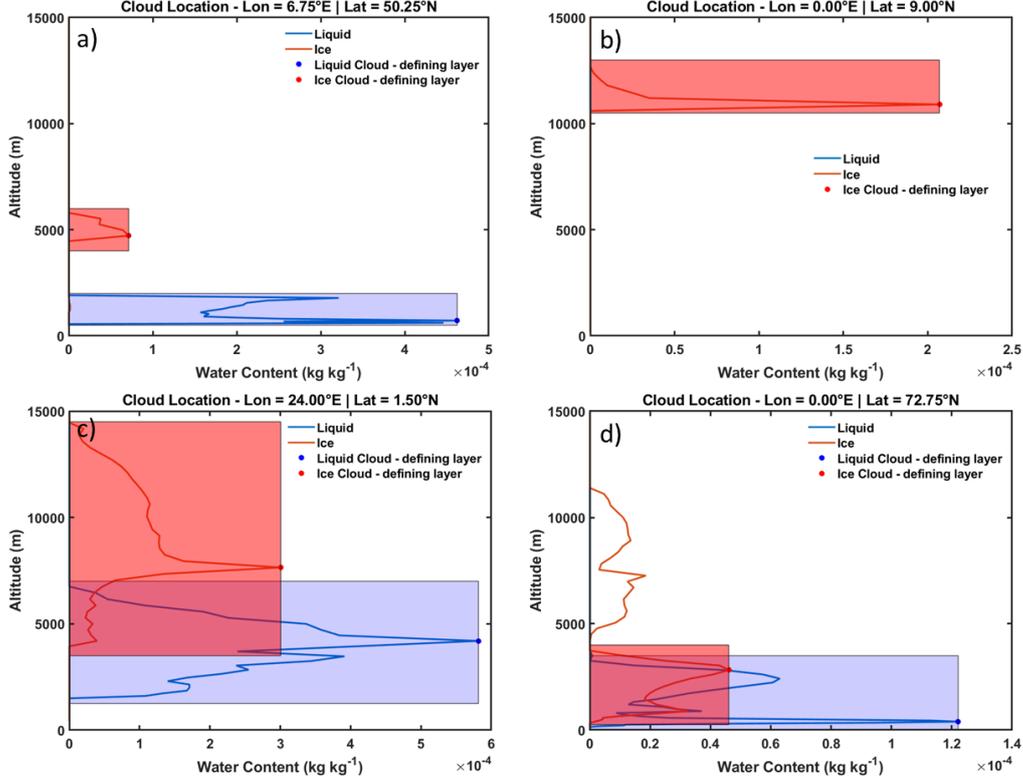
Figure A1.1: Vertical profiles of LWC (blue lines) and IWC (orange lines), from ERA 5 Model Levels, on 15th January 2018 at 00 UTC. The blue and orange boxes indicate the vertical extension, assumed automatically by the SGM, of liquid water and ice clouds, respectively.

geometrical thickness, for both liquid water and ice clouds in the SGM study area, on 15th January 2018 at 18 UTC, is shown.

Finally, the cloud effective radius and the cloud optical depth at 900 cm^{-1} are computed by using the parametrizations employed in Stengel et al. (2018), with a few differences, since these authors applied the parametrizations to data from the ERA Interim reanalysis, and here data from the ERA 5 reanalysis are used.

The liquid water droplet effective radius (r_{eff_LiqWat}, given in µm) is calculated by the parametrization proposed by Martin et al. (1994) – Equation A1.1.

$$r_{eff\_LiqWat} = \left(\frac{3LWC}{4\pi\rho_w N_d k} \right)^{(1/3)} \tag{A1.1}$$

where the LWC here is given in $kg m^{-3}$, obtained from the in-cloud specific liquid water content by multiplying by the density of air($\rho = p/R_w T$), which uses the air pressure (p), the air temperature (T), and gas constant in the moist air, that depends also of the specific humidity of the air.
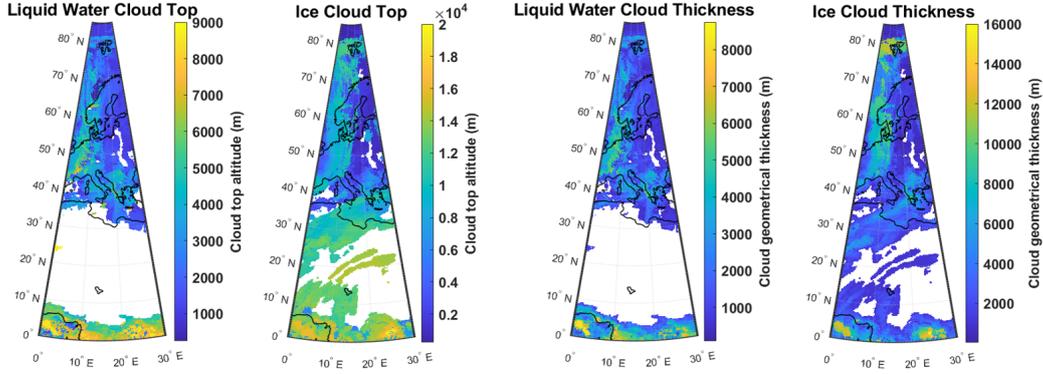
Figure A1.2: Cloud top height and geometrical thickness in the SGM study area, on 15th January 2018 at 18 UTC.

Since in the SGM the cloud microphysical parameters are assumed homogeneous in the cloud vertical profile the LWC used here assumes just one value given by the weighted mean of LWC in the liquid water cloud layer.

The other parameters of Equation A1.1 are: ρ_w - the density of water, assumed equal $10^3 kg m^{-3}$; N_d - the number of cloud condensation nuclei, assumed 300 cm^{-3} over land and 100 cm^{-3} over sea (Stengel et al., 2018); and k - a constant formulated in terms of the relative dispersion D of the particle size distribution - Equation A1.2.

$$k = \frac{\left(1 + D^2\right)^3}{\left(1 + 3D^2\right)^2} \tag{A1.2}$$

with D=0.33 over ocean and D=0.43 over land.

According to Stengel et al. (2018), the r_{eff_LiqWat} is limited in the range from 4 µm to 16 µm.

The ice particle effective radius (r_{eff_ice}, given in µm) is calculated from the effective diameter parametrized as a function of the temperature and IWC based on Sun and Rikus (1999) and revised by Sun (2001) – Equation A1.3.

$$r_{eff\_ice} = 0.64952 d_{eff\_ice} \tag{A1.3}$$

where:

$$d_{eff\_ice} = c\left(a + b\left(T - 83.15\right)\right)$$

$$a = 45.8966 IWC^{0.2214}$$

$$b = 0.7957 IWC^{0.2535}$$
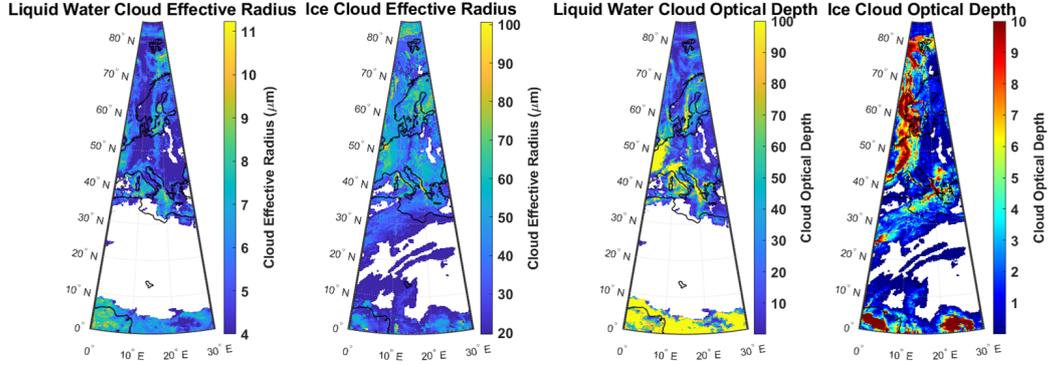
Figure A1.3: OD and r_{eff} of liquid water and ice clouds, calculated using the parametrization proposed by Stengel et al. (2018), on 15th January 2018 at 18 UTC.

$$c = 1.2351 + 0.0105\,(T - 273.15)$$

IWC also assumes just one value given by the weighted mean of IWC in the ice cloud layer, and now computed in gm^{-3}; and T is the temperature in K. As in the case of the r_{eff_LiqWat}, the d_{eff_ice} is also limited to be in the range $30\,\mu m$ to $155\,\mu m$ (Stengel et al., 2018).

After computed the cloud effective radius for liquid water cloud and ice cloud, the cloud optical depth at $900\ cm^{-1}$ is computed by Equation A1.4

$$OD = \frac{3LWC\,CL_{thick}Q_{ext}}{4r_{eff}\rho} \tag{A1.4}$$

where all the parameters are associated to the water phase: LWC is the in-cloud liquid water content or ice water content, in kgm^{-3}; r_{eff} is the effective radius of liquid water or ice particle, in m; CL_{thick} is the cloud geometrical thickness, in m; ρ is the density of liquid water or ice; and Q_{ext} is the extinction efficiency at $900\ cm^{-1}$, which depend of the r_{eff} and is extracted from the SGM cloud auxiliary database.

An example of the values obtained using this strategy is provided in Figure A1.3.

Acknowledgment

I would like to thanks everyone that has contributed in some way to me reach this dream.

First, I thanks my supervisor, Tiziano Maestri. Thank you for giving me this opportunity, to trust me, to share your knowledge and experiences, and guide me along this way. Thank you for your patience and for your friendship. I have no words to express my gratitude to you. Then, at this moment I just say thank you so much.

I wish also to acknowledge Davide Magurno and Michele Martinazzo for all discussion and support, either about scientific and general problems, and solutions. It is really a pleasure to work with you. I want to thanks also all people of the atmospheric physics research group from the University of Bologna. Thank you for all the moments, for each advice, and also for the weekly meeting during the last year, where I have really learned and opened my mind.

I would like also to thanks all groups involved with the FORUM mission. Thank you to work hard in this challenge, that now is a reality. Soon FORUM sensors will be flying and giving us much more work and pleasure to analyse its measurements. Especially, thank you to all research groups from the CNR (IAC, ISAC and INO) and from the University of Basilicata, that in all times share your knowledge, which helps me to understand each day a little bit more about the possibilities that this incredible mission is opening.

I wish to acknowledge Professor Helen Brindley, who hosted me for three months in the Imperial College London SPAT research group. Thank you for this opportunity to enjoy a great research experience in the UK, and also for your availability together with John Murray and Richard Bantges, that always supported and helped me. Thank you also to each one from the SPAT. It was an amazing experience that I lived with you.

I want also to thanks the University of Bologna for all structure and support available along this way. Particularly, I thanks to all people of the Geophysics PhD course. Thank you to the professors, for sharing your knowledge and experiences, to the administrative workers, for all support and availability to solve the problems, and also to the XXXIII

cycle students, who have divided their learning with me.

I really want thanks to all my friends. I will not nominate everyone, because this list would become endless. But I know that each one of you knows your importance in my life, and I hope that I can correspond with the same intensity and dedication that you give to me. Thank you to all my Brazilian friends, since those who studied with me from elementary school or lived near to me, and keep yet next to me, up to the great friends that UFRJ and the meteorology gave me. Thank you also to all friends that Bologna brought to me. Here, I need to mention a special thank to Professor Federico Porcu and his wife, Giovanna Corrado, who have made possible to us live in this wonderful city.

I dedicate also a special acknowledgment to my family. You are the most precious thing that I could have in my life. I am sorry to be far so long time. I can say now that we finish this cycle together. Thank you for your unconditional love, and I hope to hug everyone you as soon as possible. Special gratitude to my mother, Marcia Cossich, and my grandmother, Edna Cossich (in memoriam). We got it, and the dreams became a reality.

Finally, I thanks my wife, Alessandra Carbonel. Thank you for accompanying me in every moment of this challenge. Thank you for your patience and support, and for your love. I hope that it is the first stair of our life together. I have no words to express my love to you. I just can say I love you!