



**HAL**  
open science

# The role of clouds and surface processes on local temperature variability : analysis of multi-variables observations

Oscar Javier Rojas Munoz

► **To cite this version:**

Oscar Javier Rojas Munoz. The role of clouds and surface processes on local temperature variability : analysis of multi-variables observations. Ocean, Atmosphere. Université Paris-Saclay, 2021. English. NNT : 2021UPASJ016 . tel-03524679

**HAL Id: tel-03524679**

**<https://theses.hal.science/tel-03524679>**

Submitted on 13 Jan 2022

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Le rôle des nuages et des processus de surface sur la variabilité de la température à l'échelle locale : analyse d'observations multivariées

*The role of clouds and surface processes on local temperature variability: analysis of multi-variable observations*

## Thèse de doctorat de l'université Paris-Saclay

École doctorale n° 129, Sciences de l'Environnement d'Ile-de-France (SEIF)  
Spécialité de doctorat : Océan, atmosphère, climat et observations spatiales  
Graduate School : Géosciences, climat, environnement, et planètes. Référent : Université de Versailles-Saint-Quentin-en-Yvelines

Thèse préparée dans l'unité de recherche OVSQ (Université Paris-Saclay, UVSQ, CNRS), sous la direction de **Marjolaine CHIRIACO**, Maitresse de Conférence

Thèse soutenue à Guyancourt, le 13 décembre 2021, par

**Oscar Javier ROJAS MUÑOZ**

## Composition du Jury

<b>Cécile MALLET</b> Professeure des Universités, UVSQ/LATMOS	Présidente
<b>Céline CORNET</b> Professeure des Universités, UDL/LOA	Rapporteuse & Examinatrice
<b>Fabienne LOHOU</b> Maitresse de Conférence, HDR, UPS/LAERO	Rapporteuse & Examinatrice
<b>Vincent NOEL</b> Directeur de Recherche, UPS/LAERO	Examinateur
<b>Romain ROEHRIG</b> Ingénieur des Ponts, des Eaux, et des Forêts, CNRM	Examinateur
<b>Marjolaine CHIRIACO</b> Maitresse de Conférence, UVSQ/LATMOS	Directrice de thèse

**Titre :** Le rôle des nuages et des processus de surface sur la variabilité de la température à l'échelle locale : analyse d'observations multi-variables

**Mots clés :** Nuages, Processus de surface, Variabilité, Température, Observations

**Résumé :** La circulation des masses d'air à grande échelle domine au premier ordre le climat européen, mais elle n'explique pas à elle seule l'ensemble de la variabilité des températures et des précipitations. Au second ordre, cette variabilité dépend de processus de petite échelle via la couche limite atmosphérique et le bilan d'énergie à la surface (SEB), qui dépend lui-même en grande partie du rayonnement et donc des propriétés des nuages.

Ce travail consiste à développer une méthode (OTEM) pour estimer les termes contrôlant les variations horaires de température à 2 m à l'échelle locale à l'observatoire francilien du SIRTA, en s'appuyant quasiment uniquement sur des observations, grâce au jeu de données SIRTA-ReOBS. Les contributions de ces termes ainsi que leur importance (estimée à l'aide de la méthode des forêts aléatoires) sur les variations de températures

sont détaillées, avec une attention particulière sur le rôle des nuages.

La méthode développée est aussi appliquée au site de la Météopole à Toulouse pour comprendre comment les conditions spécifiques locales affectent chacun des termes impliqués. De plus, la variabilité de ces termes en fonction des conditions de circulation de l'air à grande échelle en séparant nos résultats en fonction des régimes de temps nord-atlantique est analysé. Cette méthode peut aider à améliorer les paramétrisations déjà existantes des termes SEB, ainsi qu'à faire une comparaison entre les simulations climatiques régionales multi-modèles pour évaluer si les modèles sont capables de bien reproduire ces comportements, et les améliorer.

**Title:** The role of clouds and surface processes on local temperature variability: analysis of multi-variable observations

**Keywords:** Clouds, Surface processes, Temperature, Variability, Observations

**Abstract:** Large-scale air mass circulation dominates the European climate at the first order, but it does not explain by itself all of the temperature and precipitation variability. At the second order, this variability depends on small-scale processes via the atmospheric boundary layer and the surface energy balance (SEB), which itself depends largely on radiation and thus on cloud properties.

This work consists in developing a method (OTEM) to estimate the terms controlling the hourly temperature variations at 2 m at the local scale at the SIRTA observatory in Paris, relying almost only on observations, thanks to the SIRTA-ReOBS dataset. The contributions of these terms as well as their importance (estimated using the random forest method) on temperature variations are detailed, with

a particular focus on the role of clouds.

The developed method is also applied to the Meteopole site in Toulouse to understand how the local specific conditions affect each of the involved terms. Furthermore, the variability of these terms as a function of large-scale air circulation conditions by separating our results into North Atlantic weather patterns is analyzed. This method can help to improve the already existing parameterizations of the SEB terms, as well as to make a comparison between multi-model regional climate simulations to assess whether the models are able to reproduce these behaviors well, and improve them.



# Acknowledgements

I would like to start thanking my advisor Marjolaine, who has been an excellent tutor and most of all, an excellent human being. Thank you for being available when I needed you, when I had some doubts and/or questions not only about my work, but also about some aspects in my personal life. Despite the pandemic and the fact that we didn't see each other that often, you always tried to be there and you let me work at my own rhythm. Working by your side was a great experience and I hope someday our paths will cross again.

Thank you, Sophie, because even if officially you weren't a thesis advisor, during these 3 years you worked as if you were, and helped me to better understand and analyze some results. I appreciate the fact that you were also available to discuss plenty of times during quite hours of a day, this work done is also thanks you.

Thanks also to all the SPACE team who welcomed me at the beginning of my thesis with open arms and were an important companion for me to adapt to the French research environment. Thanks to: Nicolas, for being an amazing team boss and advising me concerning the defense, and for calming me down when I was preparing it; H el ene for your kindness; Julien who always said "good morning" every time he crossed my office; and of course, the other SPACE members who made me feel as home in this amazing laboratory.

Especial thanks to Guylaine Canut, a CNRM researcher with whom I collaborated for the last part of my thesis and whom provided me with the data to apply the model developed to the M et eopole site in Toulouse. Also, thanks to William Maurel who made the data I required available, sorted and ready to use.

Thanks then to all my PhD friends who I've met during these three years. Thank you Pragya, you've become an important companion and friend during these hard times, I appreciated all the moments we shared taking lunch, during the pause caf e, the hang outs, where we were discussing about our "dog house" problems and personal matters, and also thanks to open my eyes and taught me about the Indian culture, which I really love now. Thank you, Melissa, you were my first office mate and we had really good times and laughs sharing our Colombian heritage. To Erik, the kind Swede, you've been an awesome and wonderful friend. Thanks also to Chlo e, my second office mate, which whom I started an amazing friendship and to whom I wish all the best in the rest of the PhD in USA and here in LATMOS. To Cl emantyne, who helped me a lot recently and especially with my family, thank you so much from my part and from my parents. Also, thanks to Minh, Phoung and Vibolroth who made me discover further the Vietnamese and Cambodian culture. And thanks to the other PhD candidates with whom I spent my lunch breaks: Alexis, Valentin, Oph elie, C edric, and specially Nicolas, with whom I

shared the role of representative of doctoral students (but in reality, you did all the job), and also helped me through all the procedures at the end of my thesis. Thanks also to my Chilean friends Susana and Felipe.

I would like to thank the members of the jury: Fabienne et Céline, for having accepted to read very carefully my manuscript and suggested some valuable corrections; Romain, Vincent et Jordi for the great discussion we had at the end of my defense. Thanks especially for the two latter, whom were part of my thesis committee and advised me once in a year and gave me an outside perspective and point of view of my work.

A big thanks to my closest and best friends here in France, who I had the chance to meet the first year I came to Toulouse in 2015: Alberto and Sebastian, you were the biggest support during this amazing adventure, and I was able to carry out this work also thanks to you.

An especial thanks to Maria Victoria, who was without doubts the person who kept me in place and encouraged me during these three years with her love and kindness, and to whom I wish the best in her personal and professional life in Colombia.

I would like to dedicate some lines to Justine, who sadly left us this year: her help during my first year of this thesis was invaluable, she was always available with a big smile when I needed her to ask her even the dumbest question about coding. Life is not always fair, but I know you are in a better place now, and I wanted to thank you for everything.

I thank and dedicate my work to my parents, who always supported me and encouraged me years ago to come to France to accomplish my dreams, without you this probably would never happen. And to the rest of my family: my brothers and nieces and nephews, thank you for always been there.

## Résumé

La dynamique de grande échelle domine au premier ordre les variations de température de surface et les conditions atmosphériques en Europe occidentale. Cependant, cette circulation des masses d'air à grande échelle n'explique pas à elle seule l'ensemble de la variabilité des températures et des précipitations. Au second ordre, cette variabilité dépend de processus de petite échelle via la couche limite atmosphérique et le bilan d'énergie à la surface (SEB), qui dépend lui-même en grande partie du rayonnement et donc des propriétés des nuages.

L'objectif de cette thèse est de mieux comprendre les processus locaux et leur influence sur la variabilité climatique locale, avec une attention particulière sur le rôle des nuages.

Pour cela, le premier objectif est de quantifier la contribution locale spécifique des principaux termes du SEB et de l'advection de l'air agissant sur les variations de température à court terme (i.e. horaire) en Ile de France, et de déterminer leur importance et les conditions dans lesquelles l'un ou l'autre de ces termes sera prépondérant. Les cinq termes agissant sur les variations de température sont le rayonnement (que l'on peut séparer en contribution du ciel clair et des nuages), l'échange de chaleur avec l'atmosphère, l'échange de chaleur avec le sol, et l'advection de température. Nous développons le modèle OTEM qui permet d'estimer ces termes en s'appuyant quasiment exclusivement sur des observations, grâce au jeu de données SIRTA-ReOBS. Nous montrons que la somme de ces cinq termes donne une bonne estimation des variations horaires de température réelles. La contribution de chacun des termes en fonction de la saison et pour le jour et la nuit montre que le terme radiatif du ciel clair contribue le plus aux variations de température, et que les autres quatre termes agissent donc comme modulateurs. De plus, le poids de chaque terme du SEB sur les variations horaires de température est analysé à l'aide de la méthode des forêts aléatoires, dont le principal avantage est sa capacité à traiter des milliers de variables d'entrée et identifier les plus significatives. Cette analyse a montré que quelle que soit la saison, les nuages sont le principal modulateur de l'effet du soleil sur les variations de température en une heure pendant la journée (même si c'est le terme radiatif ciel clair qui contribue le plus en termes de magnitude aux variations horaire de la température), et ils dominent complètement pendant la nuit. Le terme d'échange de chaleur avec l'atmosphère devient aussi important en fin d'après-midi quand la turbulence à la surface est maximale et la hauteur de la couche devient maximale.

Le deuxième objectif est l'étude du rôle spécifique des nuages dans les variations de température. Pour cela, d'autres observations incluant des profils lidar ont été utilisés, exclusivement dans des conditions nuageuses. Plusieurs cas ont été créés à partir de l'effet radiatif des nuages pendant le jour et la nuit pour (i) mieux comprendre comment ils affectent l'état de l'atmosphère et donc des autres variables à la surface,

et (ii) caractériser le type des nuages prédominants en fonction de leur effet radiatif. Une prédominance de nuages hauts est trouvée au SIRTA, et est en partie liée à la position géographique de l'observatoire puisqu'il est situé dans une zone où l'air chaud et humide provenant de l'océan Atlantique peut rencontrer les masses d'air froid et sec provenant de la région sibérienne. Cette rencontre des masses d'air (que ce soit l'air chaud qui rencontre l'air froid ou l'inverse) tend à former principalement des nuages de haute altitude.

Finalement, nous étudions la variabilité des résultats obtenus précédemment : (1) variabilité spatiale en appliquant la même méthode au site de la Météopole à Toulouse pour comprendre comment les conditions spécifiques locales affectent la contribution et l'importance de chacun des termes impliqués dans les variations de température de surface ; (2) variabilité en fonction des conditions de circulation de l'air à grande échelle en séparant nos résultats en fonction des régimes de temps nord-atlantique.

# Abstract

Large-scale dynamics dominate the surface temperature variations and atmospheric conditions in Western Europe at first order. However, this large-scale air mass circulation alone does not explain all the temperature and precipitation variability. At the second order, this variability depends on small-scale processes via the atmospheric boundary layer and the surface energy balance (SEB), which itself depends largely on radiation and thus on cloud properties.

The objective of this thesis is to better understand local processes and their influence on local climate variability, with a particular focus on the role of clouds.

To this end, the first objective is to quantify the specific local contribution of the main SEB and air advection terms acting on short-term (i.e. hourly) temperature variations in Ile de France, and to determine their importance and the conditions under which one or the other of these terms will be dominant. The five terms acting on temperature variations are radiation (which can be separated into clear sky and cloud contribution), heat exchange with the atmosphere, heat exchange with the ground, and temperature advection. We develop the OTEM model which allows us to estimate these terms almost exclusively from observations, using the SIRTA-ReOBS dataset. We show that the sum of these five terms gives a good estimate of the real hourly temperature variations. The contribution of each of the terms as a function of season and for day and night shows that the clear-sky radiative term contributes the most to the temperature variations, and that the other four terms therefore act as modulators. In addition, the weight of each SEB term on hourly temperature variations is analyzed using the random forest method, whose main advantage is its ability to handle thousands of input variables and identify the most significant ones. This analysis showed that whatever the season, clouds are the main modulator of the sun's effect on hourly temperature variations during the day (even though it is the clear sky radiative term that contributes most in terms of magnitude to hourly temperature variations), and they completely dominate during the night. The heat exchange term with the atmosphere also becomes important in the late afternoon when the surface turbulence is maximum and the layer height becomes maximum.

The second objective is to study the specific role of clouds in temperature variations. For this purpose, other observations including lidar profiles have been used, exclusively in cloudy conditions. Several cases have been created from the radiative effect of clouds during the day and night to (i) better understand how they affect the state of the atmosphere and thus other variables at the surface, and (ii) characterize the type of predominant clouds according to their radiative effect. A predominance of high clouds is found at SIRTA, and is partly related to the geographical position of the observatory since it is located in an area where warm and humid air from the Atlantic Ocean can meet cold and dry air masses from the Siberian region. This meeting of air masses (whether warm air meets cold air or the reverse) tends to form mainly high-

level clouds.

Finally, we study the variability of the results obtained previously: (1) spatial variability by applying the same method to the Météopole site in Toulouse to understand how local specific conditions affect the contribution and importance of each of the terms involved in the surface temperature variations; (2) variability as a function of large-scale air circulation conditions by separating our results as a function of North Atlantic weather patterns.



# TABLE OF CONTENTS

---

<b>Table of Contents</b> .....	<b>11</b>
<b>Introduction</b>	<b>14</b>
<b>1 Chapter 1: Description of the different components that influence the local temperature</b> .....	<b>18</b>
Introduction.....	18
1.1 European climate: regional-scale processes .....	18
1.2 Smaller local scale characteristics involving temperature variations .....	21
1.2.1 The Atmospheric Boundary Layer .....	21
1.2.2 Vertical structure of temperature and other variables in the ABL.....	24
1.2.3 Air advection.....	26
1.3 General description of the surface energy balance (SEB) .....	26
1.3.1 Radiative fluxes .....	28
1.3.2 Turbulent heat fluxes.....	30
1.3.3 Ground heat flux .....	32
1.3.4 Diurnal and annual cycle of the SEB terms.....	32
1.4 Clouds processes affecting SEB.....	34
1.4.1 Types of clouds and their composition .....	35
1.4.2 Radiative properties.....	36
1.4.3 General impact on SEB and 2 m temperature .....	37
Conclusion.....	38
<b>2 Chapter 2: Data and Methods</b> .....	<b>40</b>
Introduction.....	40
2.1 Data.....	40
2.1.1 The SIRTA observations.....	40
2.1.2 The Météopole observations.....	47
2.1.3 ERA5 and ERA5-Land Reanalysis .....	48
2.2 Methods.....	49
2.2.1 Observation-based temperature evolution model.....	49
2.2.2 The random forest method.....	69
Conclusion.....	74
<b>3 Chapter 3: Influence and contribution of each OTEM term on local 2 m temperature variations</b> .....	<b>76</b>
Introduction.....	76
3.1 Evaluation of the model .....	76
3.1.1 Statistical evaluation of OTEM.....	76
3.1.2 Hourly evolution of OTEM capability.....	81
3.1.3 Focus on specific cases that OTEM does not capture .....	84
3.2 Evolution of the different OTEM terms .....	88
3.2.1 Annual contribution of each term.....	88
3.2.2 Monthly-hourly contribution of each term .....	90
3.2.3 Advection dependence on wind direction.....	92
3.3 Importance estimation of each term: the random forest application.....	94
3.3.1 General estimation .....	94
3.3.2 Diurnal cycle of the weight estimates.....	94
3.3.3 Validation of the random forest method .....	97
3.3.4 Basic settings on random forest method to study the weight of each term .....	98
Conclusion.....	100

<b>4</b>	<b>Chapter 4: Specific influence of clouds on 2 m temperature variations under cloudy conditions.....</b>	<b>103</b>
	Introduction.....	103
4.1	Contribution of OTEM terms in cloudy conditions.....	103
4.2	Creation of clouds categories.....	106
4.2.1	Cloud categories based on their radiative effect.....	106
4.2.2	Specific lidar sampling for daytime cases.....	107
4.3	Daytime vertical low-atmosphere structure and characterization under different cloud effects.....	108
4.3.1	Seasonal cycle of cloud occurrence.....	109
4.3.2	Clouds with strong cooling effect.....	110
4.3.3	Clouds with weak radiative effect: cooling or warming.....	113
4.4	Nighttime low-atmosphere characterization under positive cloud effects.....	116
4.4.1	Clouds with weak warming effect.....	118
4.4.2	Clouds with strong warming effect.....	118
	Conclusion.....	119
<b>5</b>	<b>Chapter 5: Meteorological and geographical variability of the OTEM terms</b>	<b>122</b>
	Introduction.....	122
5.1	Dependence of the contribution of OTEM terms to weather regimes.....	122
5.1.1	Mean contribution of OTEM terms to weather regimes.....	122
5.1.2	Weight of OTEM terms for the weather regimes.....	124
5.2	Spatial variability of OTEM terms: comparison between Toulouse and SIRT A.....	127
5.2.1	Statistics of OTEM in the Météopole.....	128
5.2.2	Comparison between SIRT A and the Météopole.....	129
	Conclusion.....	134
	<b>Conclusions and perspectives.....</b>	<b>138</b>
	<b>References.....</b>	<b>146</b>
	<b>List of figures.....</b>	<b>166</b>
	<b>List of tables.....</b>	<b>174</b>
	<b>List of abbreviations.....</b>	<b>175</b>
	<b>Appendices.....</b>	<b>177</b>
	Appendix A.1: MLD at SIRT A and Toulouse.....	177
	Appendix A.2: Type of soils in France.....	177
	Appendix A.3: Sensible and latent heat fluxes at SIRT A and Toulouse.....	179
	Appendix A.4: Pressure at the MLD.....	179
	Appendix A.5: ERA5-Land.....	180
	Appendix B: Diurnal cycle of the OTEM terms.....	181
	Appendix C.1: Daytime and nighttime PDFs of OTEM terms at both sites.....	182
	Appendix C.2: Sensible and latent heat fluxes, and MLD at Toulouse.....	184



## INTRODUCTION

---

Regional climate variability is at first order driven by large-scale atmospheric conditions. In Western Europe, the North Atlantic Oscillation (NAO; Trigo et al., 2002), which is associated with the locations and intensities of the centers of the Iceland low and the Açores high pressures, controls the air mass advection over western Europe and explains a large part of weather variability. Extreme events and temperature anomalies are generally not exclusively explained by the presence of these large-scale air mass circulations (Vautard and Yiou, 2009). Indeed, synoptic and meso-scale atmospheric processes have been previously studied to explain interannual temperature changes in some parts of Europe (Efthymiadis et al., 2011; Xoplaki et al., 2003), or even precipitation occurrence (Xoplaki et al., 2004a; Bartolini et al., 2009). It is, therefore, necessary to consider small-scale processes such as surface-atmosphere interactions and cloud processes to better explain 2 m temperature variations (e.g. Chiriaco et al., 2014).

Temperature variations at 2m are related to the Surface Energy Balance (SEB) following surface-atmosphere interactions and solar radiation (Wang and Dickinson, 2013) that are separated into different components: latent and sensible heat fluxes, ground heat flux, and atmosphere radiation. Air advection also take part to this local variability. Studies have been made to parametrize these terms when direct measurements are not available (Miller et al. 2017; Arnold et al. 1996) but their exact contribution is uncertain. Depending on the time-scale considered, the importance of each SEB term on temperature variations will change (Bastin et al. 2018; Ionita et al. 2015).

The first objective of this thesis is to quantify the specific local contribution of the primary SEB terms (together with advection contribution) acting on short-term (i.e. hourly) temperature variations in a Western European location and to determine their importance and the conditions when a certain term predominates over the others, based on a simple linear model. The current study is inspired by Bennartz et al. (2013) who implemented a temperature variation model to study the influence of low-level liquid clouds on the ice surface melt period in July 2012 over the Arctic, by estimating all of these SEB terms for a case study. Here, the same approach is used for a different location near Paris and on a long-time period to get robust statistics. To do so, the study is based on direct measurements from the SIRTAReOBS dataset (Chiriaco et al., 2018), which includes many variables collected since 2002 at SIRTAR (Site Instrumental de Recherche en Télédetection Active; Haeffelin et al. 2005), an observatory located in a semi-urban area in the Southwest of Paris, France. This dataset is well suited to the current study objective because: (i) it allows the use of a multi-variable synergistic compilation to study and compare different characteristics of the atmosphere or the surface (e.g. Bastin et al. 2018; Dione et al. 2017; Cheruy et al. 2013, Chiriaco et al., 2014), (ii) it is located in Western Europe and allows access to the hourly time scale, i.e.

the scales of the local processes of the current study. The SIRTA-ReOBS dataset enables the development of a model (called OTEM in the following manuscript) to estimate all the terms involved in the 2 m temperature variation. Indeed, the use of OTEM in the current study considers all the variables acting within the ABL and controlling 2 m temperature variations, all of them estimated almost exclusively from surface-based observations. Thus, it allows to study separately the influence of each SEB term in a local scale. This allows to have a realistic and reliable estimation of the contribution of each term (radiative fluxes, turbulent heat fluxes, etc.) on hourly temperature variations, based among other things on a random forest analysis. Since the SEB terms and advection depend on local conditions, the spatial and temporal variability of these terms is analyzed, based on a weather regime analysis and another dataset from an observatory located in Toulouse.

Actually, clouds are well known to play a direct role and to modify directly 2 m temperatures and other near-surface variables in multiple time-scales (Parding et al., 2014; Broeke et al., 2006a; Kauppinen et al., 2014). Hence the second objective of this thesis is to understand the specific role of clouds and their associated characteristics on hourly temperature variations. Indeed, the influence of clouds on the temperature at the surface can vary depending on their physical properties and the altitude where they are formed (Hartmann et al., 1992; Chen et al., 2000). In general, low-level clouds tend to cool the surface, whereas high-level clouds, such as cirrus clouds, tend to warm it by absorbing a significant amount of Earth's outgoing radiation. The average contribution of clouds is to decrease the 2 m temperature, combined with the damping effects of soil moisture, by reflecting the solar radiation to space. But their damping effects vary depending on the season and by time of day. For instance, the decrease in 2 m temperature is the highest in fall in northern mid-latitudes for specific cases when precipitation is not significant (Dai et al. 1999). The local contribution of clouds to temperature variations is thus an important topic, to assess how this intake affects local climate variability and extreme local events, such as the extreme heatwave and drought during summer in EU in 2006 (Chiriaco et al., 2014; Rebetez et al., 2009) or the sudden melt of the ice sheet in Greenland on July 2012 (Bennartz et al., 2013). Several studies were made primarily focusing on the large scale effects of clouds on the radiative energy balance on a global scale, either at the top of the atmosphere (Arkin and Meisner, 1987; Raschke et al., 2005; Dewitte and Clerbaux, 2017; Willson et al., 1981; Allan et al., 2014; Cherviakov, 2016) or at the surface (Wild et al., 2015; Hakuba et al., 2013; Wild, 2017) or both at the same time (Hartmann, 1993; Kato et al., 2012; Li and Leighton, 1993). However, studies investigating their impact on a smaller local scale are still lacking, due to the limited availability of reliable field observations.

In addition, clouds are the main source of uncertainty for climate models (Dufresne and Bony, 2008; Vial et al., 2013; Nuijens et al., 2015; Sherwood et al., 2014; Watanabe et al., 2012; Yao and Cheng, 2012; Brient et al., 2016). Actually, there is a lack of data on clouds from the pre-industrial era, before pollution, and no comparison between these

and the current clouds could be done to analyze their temporal evolution on a larger scale. Also, models have difficulties to simulate correctly the clouds properties: the calculation grid in the models is usually too large (~km), yielding to a problem with relative humidity. In reality, the condensation appears at a much smaller scale than that of the calculations and clouds can form and dissipate in smaller scales (~m) and the models are not able to well simulate this pattern. Furthermore, the main issue relies on considering the atmospheric motions, precipitations and cloud radiation that interact simultaneously to form clouds, and at the same time conditionate their radiative effects. It is therefore important to study how clouds interact with other variables through processes, depending on their radiative effects on a local scale, and characterizing them using active instruments which is done in this thesis.

To summarize, the objective of this thesis is to study **the role of clouds and local processes on temperature variability by analyzing multi-variables observations.**

To answer this question, this manuscript consists in five chapters. The **first chapter** introduces the general scientific context of the study by firstly focusing on large scale circulation processes controlling at first order temperatures in Europe, and then a second part explaining why it is important to study smaller-scales processes and surface-atmosphere interactions to better explain temperature variability near the surface. A section is consecrated to present clouds properties and characteristics, and their influence on these small-scale processes. The **second chapter** presents the observational datasets used in this thesis to develop a ground-based observational model to estimate hourly temperature variations in a local scale, for two different observatories in France. Then, we describe how the different terms acting on hourly temperature variations (and thus involved in our model) are estimated, along with a brief presentation of the machine learning method (the random forest) used to evaluate their importance. The **third chapter** is consecrated on a first part to evaluate how well the model previously developed fits the hourly observations, based on statistics. On a second part, the contribution of each term is estimated from a monthly-hourly and annual cycle perspective, as well as their importance under different conditions by performing the random forest method. The **fourth chapter** focuses on a study about the specific role of clouds on 2 m temperature variations and on different standard meteorological variables. This evaluation is done partly based on lidar profiles analysis. In the **fifth chapter**, the variability of the terms controlling 2 m temperature variations are studied based on a weather regime analysis and on comparisons between two sites, in order to assess how specific local and weather conditions affect their contributions. Finally, a conclusion summarizes the main results obtained in this thesis and provides perspectives opened by this work.



# **1 CHAPTER 1: DESCRIPTION OF THE DIFFERENT COMPONENTS THAT INFLUENCE THE LOCAL TEMPERATURE**

---

## **INTRODUCTION**

The objective of this thesis is to study the role of clouds and surface processes on local temperature variability based on the analysis of multi-variables observations. This chapter introduces the scientific context and presents the different components that influence the local temperature variations. The first part consists of a brief explanation on European climate and how large-scale atmospheric processes control in first order surface<sup>1</sup> temperature but do not explain by themselves temperature variability and extreme events. Then, a special focus on small-scale atmospheric processes is addressed, and each of the terms acting on surface-atmosphere interactions is introduced. A focus on clouds and their physical properties is addressed, since we will show in this work that they are the main modulator of the sun effect on temperature control at small time scale, but also because they represent the main source of uncertainty in the future temperature forecast.

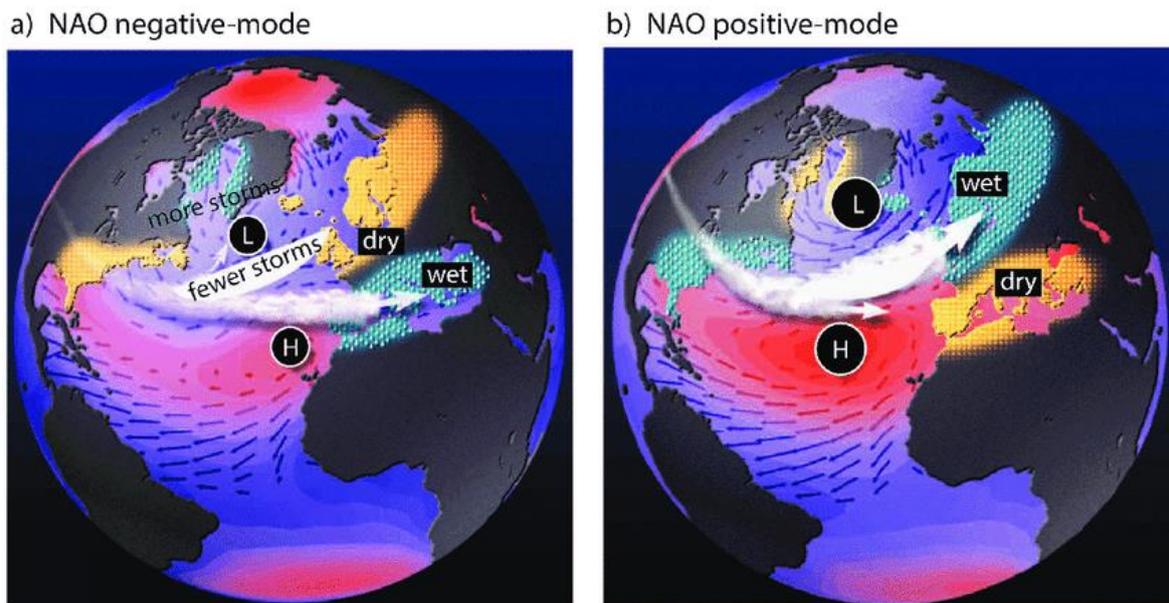
## **1.1 EUROPEAN CLIMATE: REGIONAL-SCALE PROCESSES**

Depending on the area of study, different atmospheric large-scale processes control at first order climate and surface temperature variations and anomalies in Western Europe, and these large-scale circulations are linked to pressure systems that are dominated by the Atlantic Ocean (Xoplaki et al., 2004a; Vihma et al., 2020; Tomczyk et al., 2019; Klavins and Rodinov, 2010). A marine climate is then associated in this area since it is mostly the Ocean that modulates temperatures and controls moisture availability in the troposphere (Palter, 2015). Indeed, Western Europe is located in a transition zone where the warmer air masses coming from the Sahara Desert meets the cooler and drier air masses from the Siberian Anticyclone along with the mild and humid westerly flows coming from the Atlantic Ocean. The encounter of these air masses creates low-pressure in the Scandinavian zone and high-pressure in the Azores islands, and the North Atlantic Oscillation (NAO; Trigo et al., 2002; Hurrell et al., 2003) weather phenomenon is established, controlled by the difference of atmospheric pressure level between these two zones. Therefore, two phases are defined based on an index function of this pressure difference: (1) a negative phase characterized by a low difference of pressure between the Arctic and the subtropical Atlantic basin, that brings moisture over low-latitudes and the Mediterranean and dry conditions over

---

<sup>1</sup> Surface temperature refers in this thesis to the air temperature at 2 m.

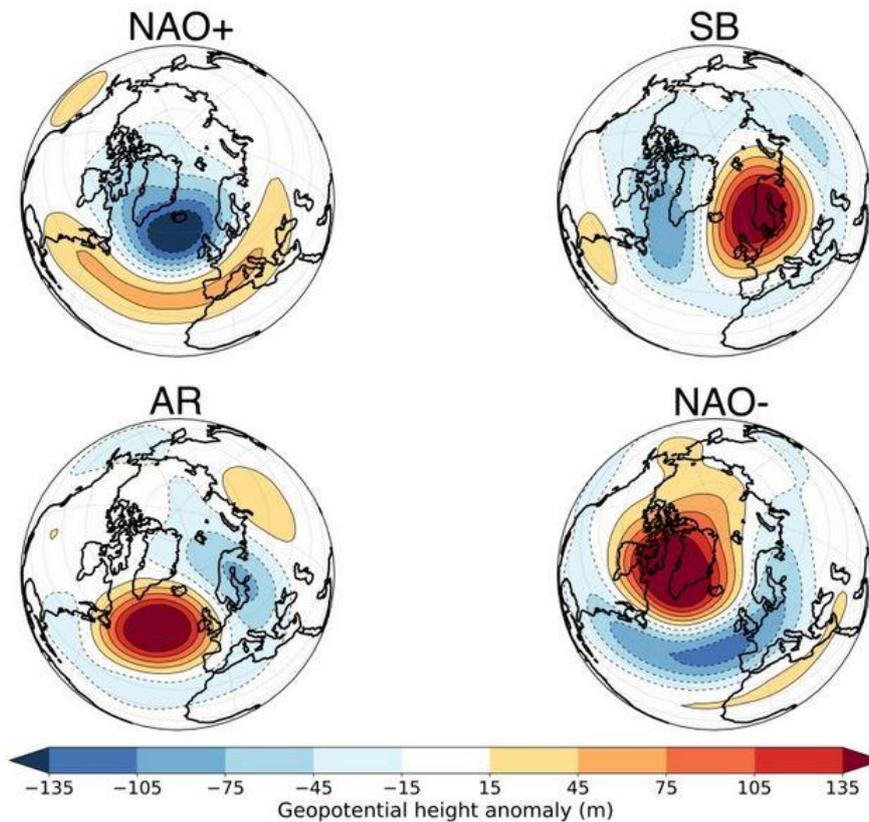
North Europe (Figure 1-1a). (2) A positive phase generated when the two pressure centers are intensified and thus enhancing storms and heavy rains in North Europe whereas a dry zone is established at the south (Figure 1-1b). Thus, the NAO is responsible in large changes in surface air temperature, winds, precipitations and moisture over the Atlantic region, in winter and summer seasons (e.g. Peings and Magnusdottir, 2014; Gastineau et al., 2013).



**Figure 1-1:** Explanatory diagram of the (a) negative and (b) positive phases of the North Atlantic Oscillation. This figures is extracted from Moura et al. (2017).

Since the upper air circulation is constantly changing between each season, the centers of actions of low- and high-pressure systems undergo significant spatial variations all over Western Europe. In order to anticipate the evolution of this atmospheric circulation and characterize its main features, four different patterns or regimes are established which will condition the weather over Europe in winter or summer. Each regime can last from 2-3 days up to 3 weeks (Matsueda and Palmer, 2018). Two of them are actually the two phases of the NAO, a third one is the Atlantic ridge which displays a strong anticyclonic ridge over North Europe, and the last one is an anticyclonic blocking between Greenland and Scandinavia (Figure 1-2). Once again, European climate weather is then mainly driven by the NAO and these two other weather regimes, which conditionate weather and climate conditions (e.g. Rojas et al., 2013; Raymond et al., 2018; Sanchez Gomez and Terray, 2005).

They are also related to some warming stratospheric events that influence low tropospheric and surface variables (e.g. Domeisen et al., 2020). These weather regimes affect also the production of energy since they control wind speed and solar power production, and their study remain crucial in order to guarantee a continuous energy supply, as shown by Wiel et al. (2019) who assessed the influence of these four regimes on the European energy sector and production during winter.



**Figure 1-2:** European weather regimes calculated from the daily potential height data at 500 mb isobaric level obtained from a combination of ERA40 and ERA-Interim for the period 1957-2014. The four regimes are the positive (upper left) and the negative (lower right) phases of the NAO, the Scandinavian blocking (upper right) and the Atlantic Ridge (lower left). This figure is extracted from Fabiano et al. (2020).

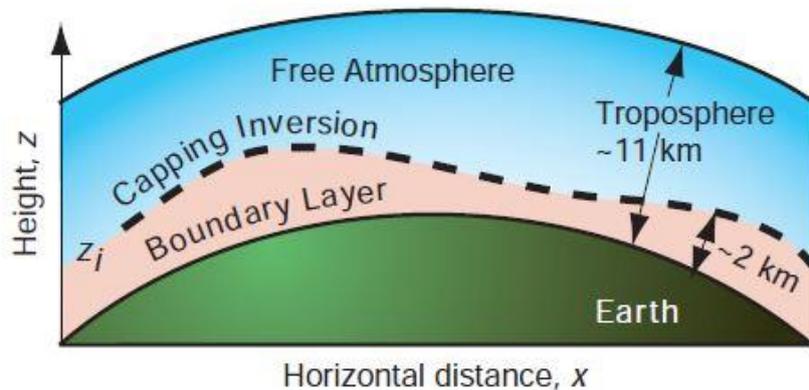
Winter and summer present then some defined weather regimes in Europe at mid-latitudes, but fall and spring are also important seasons with their weather tendencies. These last two seasons are known to be seasons of transition in mid-latitudes as they still present patterns of the season before while they also give a glimpse of the season to come. This is the reason why weather conditions vary faster than during summer and winter and thus weather regimes are not established for spring and fall. Furthermore, the rapid change of water vapor content in the troposphere (the lowest layer of Earth's atmosphere) during these two transition periods is a key element on determining weather and near-surface temperature extreme values, as shown by Betts (2011) in New England, Canada. Indeed, water vapor has an important greenhouse effect and can control the radiative cooling of the surface during the night, and its amount is mostly controlled by the vegetation, which happen to change drastically during spring (when the leave-out processes occur) and fall (when photosynthesis is "shut down").

## 1.2 SMALLER LOCAL SCALE CHARACTERISTICS INVOLVING TEMPERATURE VARIATIONS

The large-scale circulation of air masses cannot explain all the time by itself all the variability detected in temperatures and precipitations. In fact, surface temperatures in smaller time scales (less than a day) are highly influenced by other factors, such as clouds, turbulent heat fluxes and wind advection (Offerle et al., 2007; Fischer et al., 2012). Indeed, it has been shown that temperature extremes are more and more influenced by other variables such as soil moisture (Fischer and Schär, 2008), or aerosol presence (Barnett et al., 2006). Therefore, the study and characterization of such factors and the area where they have to be considered is essential to understand temperature variability. These different factors are presented in the following section.

### 1.2.1 The Atmospheric Boundary Layer

Processes that drive surface temperature occur within the atmospheric boundary layer (ABL). The Earth's surface then acts as the bottom boundary of the atmosphere and the portion most affected by this boundary is known as the ABL and thus it is the lowest portion of the atmosphere, being an important element to study climate, weather and air quality. This layer is responding to surface processes at short time-scale. The ABL upper limit is where a very stable layer known as the capping inversion (Deardorff, 1979) is established, where all the turbulence and moisture from the surface is trapped, limiting their intrusion in the free atmosphere (the layer that is not directly affected by these surface processes) (Figure 1-3 and Figure 1-4). Turbulent fluxes, heat exchange and friction occur thanks to the presence of this bottom boundary and to the presence of external forcings, and the fluxes developed are gradually dispersed horizontally and vertically as the day goes on, and they will govern the thickness of the ABL, which can range from some meters up to 2 km, with mean values ranging from 1 to 2 km for the mid-latitudes. Depending on the site, ABL thickness can reach greater values, as shown by Korhonen et al. (2014) who found values of the ABL above 3.5 km measured with radiosoundings in an eastern region in South Africa, a zone characterized with strong convective conditions.



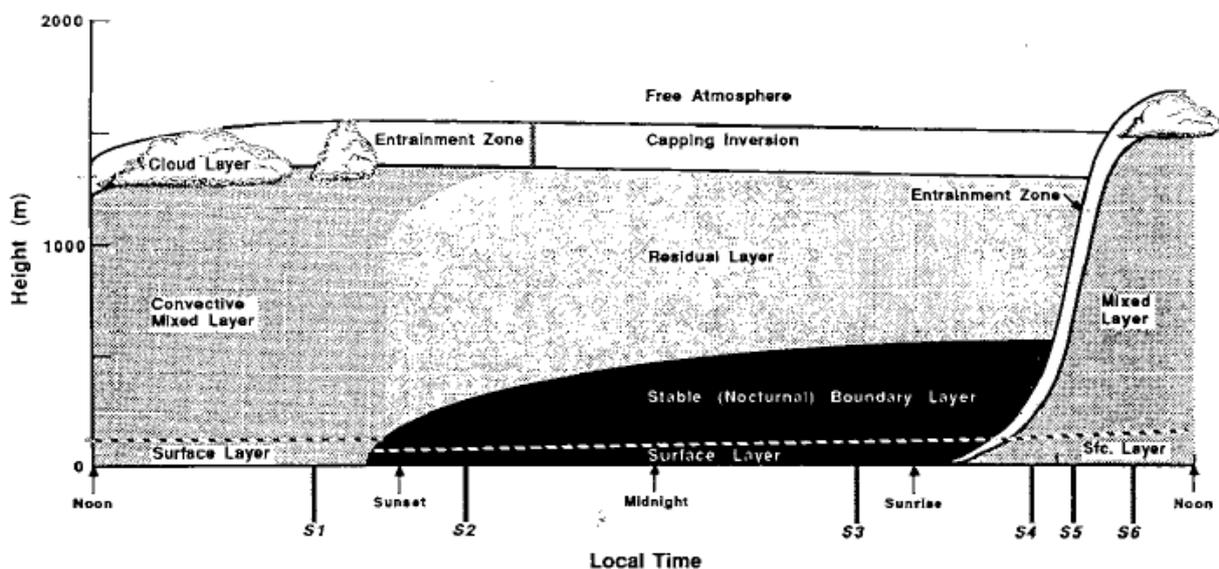
**Figure 1-3:** Vertical cross section of the Earth's atmosphere showing the two layers in the troposphere. This figure is extracted from Wallace and Hobbs (2006).

The main factor controlling ABL development is the solar radiation, which also controls the development of turbulent heat fluxes and transpiration from plants. Therefore, the boundary layer is not the same during day and nighttime. In high-pressure systems, thanks to subsidence of air, shallower boundary layers are established and the air is moved out towards lower pressure systems. In these systems, the upward motions and heat exchange carry this boundary layer air away from the surface to higher altitudes. Concerning its development, boundary layer height is quite difficult to estimate in these conditions, and thus the approximation of cloud-base (Stull, 1988) is used as the threshold of the ABL. Based on this criterion, the ABL thickness is actually thinner in low pressure systems than in high pressure ones. During daytime, the ABL has a very clear structure evolving – most of the time – with the diurnal cycle of solar radiation. On the contrary, the absence of solar radiation and the radiative cooling of the surface occurring during nighttime enhance the surface to be cooler than the air above and thus creating a so-known stable boundary layer (Caughey et al., 1979; Nieuwstadt, 1984).

Figure 1-4 presents a schematic – and idealized – boundary layer structure in a high-pressure region with fair-weather conditions. In such case, the ABL has a well-defined structure evolving with the diurnal cycle:

- During the day when solar radiation occurs, a *mixed layer* is developed and controlled by convection, which starts forming soon after the sunrise and reaching its maximum height in the late afternoon. At this time, the surface warms and it warms the air above it, creating thermals of warm air rising from the ground and transferring heat energy vertically.
- When sunset arrives, turbulence and thermals cease to form, and a so-called *residual layer* is set, which variables stay the same as those of the recently mixed layer. This residual layer is neutral, the turbulence in it is almost equally in all directions and does not have direct contact with the surface.

- At night, the bottom part of the residual layer that is in contact with the surface is transformed in a *stable layer*. The air becomes statically stable and it tends to suppress the turbulence, thus only some weak turbulence is presented at this time. The stable boundary layer is hard to measure since it has a poorly defined top which makes it hard to separate it from the residual layer.
- A *surface layer* appears in generally at the 10% of the boundary layer, either if it is a mixed or a stable one. Therefore, this layer is the lowest part of the ABL. In this layer, turbulence varies less than 10% of its magnitude.



**Figure 1-4:** Schematic atmospheric boundary layer structure over land in high pressure systems. This figure is extracted from Stull (1988).

When considering a weak horizontal wind during daytime, temperatures are warmer and weather conditions are more unstable than during nighttime, thanks to the changes of temperature, humidity and winds near the surface which are governed themselves by ABL physics. In this case, the ABL is unstable, the surface is warmer than the air aloft and turbulence is quite rapid between the surface and the air (time-scale of turbulence exchange about 30 min or less). On the other hand, nights are characterized by being cool and calm in non-windy conditions, with no (or just a few) turbulence developed, and the surface is colder than the air aloft, which the ABL is said to be stable at this time. These day and nighttime conditions can change and be affected by the increase in horizontal wind speed. Indeed, a so-called neutral boundary layer is rarely detected (Arya, 1988) but it is during overcast skies and windy conditions that it can be formed.

Different methods and approaches have been developed to estimate the thickness of the ABL from observations for both day and nighttime all over the world. These

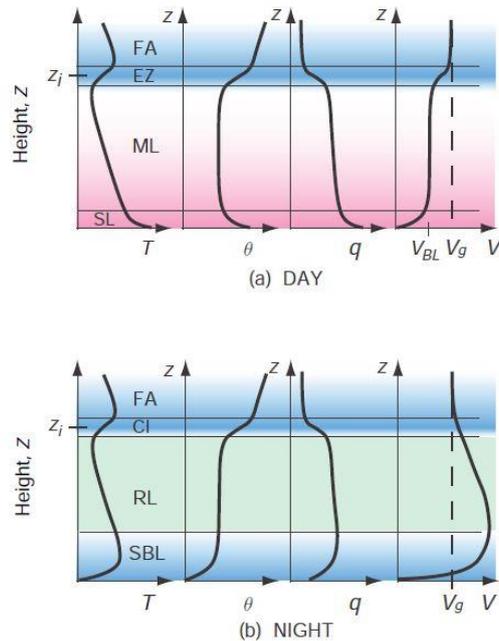
methods are applied to different type of measurements such as radiosondes (e.g. Pietroni et al., 2012), anemometers (e.g. Choi et al., 2011), or other remote sensing instruments such as radars (e.g. Lange et al., 2015) and lidars (e.g. Vivone et al., 2021; Kotthaus and Grimmond, 2018a, b; Haeffelin et al., 2012; Dang et al., 2019; Manninen et al., 2018; Toledo et al., 2014; Luo et al., 2014). A global estimation of the ABL thickness is very little available because these methods are very localized and realized over a short period of time (from some months to some years). Therefore, the scientific community is now focusing on larger-scale studies and conceiving new methods to the ABL height retrieval (e.g. Cimini et al., 2020; Barlow et al., 2017).

The difficulty of measuring the ABL relies sometimes on the vertical resolution of the instrument used for this purpose. For instance, Hennemuth and Lammert-Stockschlaeder, (2006) were not able to determine ABL height during nighttime because it was below the range of the lidar used for that study. Other cases are difficult to retrieve because the real ABL height is not distinguishable from the other aerosols top-layers heights. Therefore, many methodologies, including e.g. surface turbulent heat fluxes, vertical humidity and wind speed profiles, are combined with lidar measurements to have the closest and realistic estimation of the ABL height and its diurnal growth.

### 1.2.2 Vertical structure of temperature and other variables in the ABL

It remains important to study the variables in the troposphere since their vertical structure and interaction among themselves control both ABL diurnal cycle and growth rate. Vertical mean profiles of some classical meteorological variables within the ABL are illustrated in Figure 1-5 when no external forcing due to advection or large-scale circulations are considered. During daytime (Figure 1-5a) temperature decreases with altitude, with a similar gradient as the one for the potential temperature within the surface layer and in the entrainment zone (i.e. the capping inversion zone during daytime) where an inversion of temperature is spotted (thus some methods for determining ABL are based on this criteria). Specific humidity also tends to decrease with altitude in the ABL until it remains constant in the free atmosphere. As expected, wind velocity profile increases with height until reaching the geostrophic wind speed.

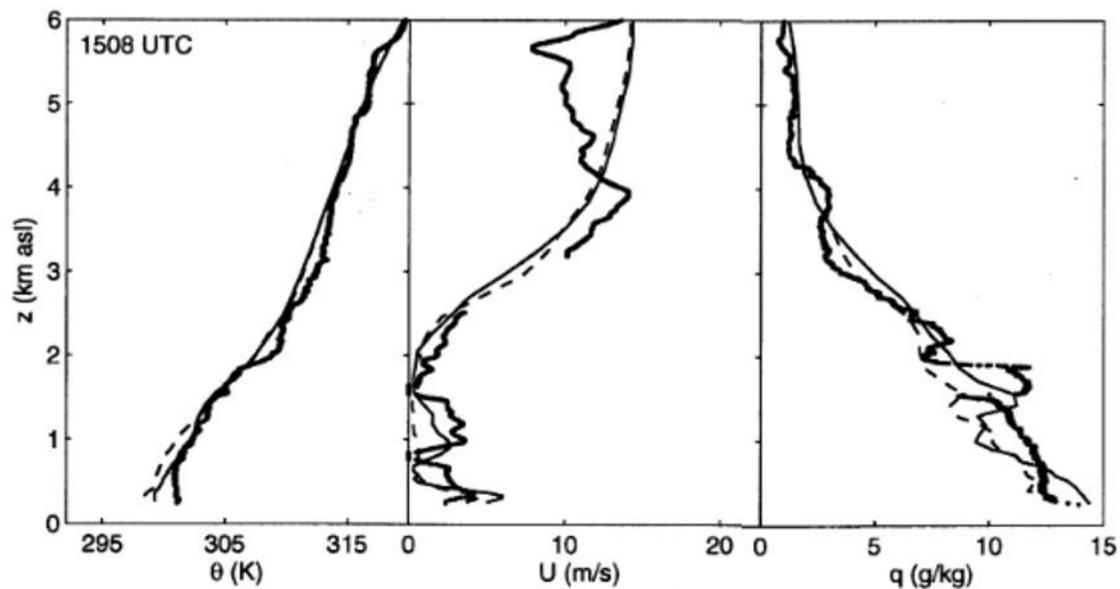
At night (Figure 1-5b), a temperature inversion near surface is generally observed within the so-called surface boundary layer, and above temperature decreases with height. Potential temperature and specific humidity both have very low values near the surface and increase on average all along the boundary layer with a stronger gradient in the residual layer. The wind vertical profile increases with height but with a lower gradient than during daytime due to the low turbulence that makes less effective the mixing of heat downward.



**Figure 1-5:** Vertical mean profiles of temperature ( $T$ ), potential temperature ( $\theta$ ), specific humidity ( $q$ ) and wind speed ( $V$ ) for (a) day and (b) nighttime. FA stands for free atmosphere, EZ for entrainment zone, ML for mixed layer, SL for surface layer, CI for capping inversion, RL for residual layer, SBL for stable boundary layer,  $z_i$  is the height of the capping inversion and  $V_g$  is the geostrophic wind. This figure is extracted from Wallace and Hobbs (2006).

Other examples of vertical profiles were done by Chow et al. (2006) shown in Figure 1-6, using radiosoundings launched at 15:08 UTC the 08/25/199 in the Riviera Valley in the Alps in Switzerland, with the objective to compare these measurements with two Large Eddy Simulations (LES) models. Note that despite the hour (13:08 LT), potential temperature follows a similar pattern as found during nighttime in Figure 1-5b. This pattern is linked to a stable atmosphere presented all that day, when a cold-air advection in the up-valley direction and subsidence of warm air from the free atmosphere prevents to a mixed-layer to form (Weigel et al., 2006; Weigel and Rotach, 2004). On the contrary, specific humidity follows the average daytime vertical behavior with a strong negative gradient near the surface (dots in Figure 1-6, right).

In this thesis, the ABL is a key parameter because it is within this layer that all the processes and variables controlling surface temperatures on smaller time scales occur, processes that are determined by the surface energy balance (SEB), which is defined in the next section.



**Figure 1-6:** Vertical potential temperature profile (left), wind speed (middle), and specific humidity (left) from radiosoundings (dots) launched at 15:08 UTC in the Riviera Valley in the Alps the 08/25/199, and two Large Eddy Simulations (dashed and solid line). This figure is extracted from Chow et al. (2006).

### 1.2.3 Air advection

Large-scale circulation conditions result locally in the advection of air. Thanks to the capacity of the air to hold and contain other substance properties, advection transports heat or other atmospheric physical quantities coming from larger-scale, such as turbulent heat fluxes (Sauter and Galos, 2016), humidity or aerosols (Belonenko et al., 2018), passing them from one area to another. Over land-surfaces, advection plays a major role because it is responsible of transport of pollutants and dust coming from far (Ningombam et al., 2015; Jia et al., 2015; Solomos et al., 2018).

Advection represents an important contribution to surface temperature, depending on the time-scale considered. Indeed, horizontal advection is responsible of both warm and cold temperature anomalies observed in Europe during blocking days in summer and winter from 1950 to 2012, as shown by Sousa et al. (2018). Furthermore, the transport of cold air could also trigger diabatic cooling and convective processes, especially in warm seasons, and thus reinforcing cold temperature anomalies. Therefore, air advection will play an important role in large-scale circulation processes, bringing warm or cold fronts, along with an impact on regional pressure systems, along with an important contribution to small-scale processes where this term will affect hourly temperatures and other surfaces processes.

## 1.3 GENERAL DESCRIPTION OF THE SURFACE ENERGY BALANCE (SEB)

As previously stated, some fluxes are developed in the bottom part of the ABL, with

temporal scale of associated processes ranges from some seconds up to several hours during the day. All the small-scale processes at the surface are related to solar radiation and advection of air.

The Surface Energy Balance (SEB) describes the portioning of energy fluxes arriving and leaving the surface, over a specifically interval of time. It represents the exchange of energy between the Earth's surface and the immediately overlaying atmosphere. These surface-atmosphere interactions within the ABL are essential elements of the climate system. For a given large-scale circulation condition, SEB controls the temperature at the surface and its diurnal cycle (Hartmann, 1994), depending on the amount of energy fluxes available. The advection of air that depends on both this large-scale circulation and local environment, will play a more or less important role depending on the time scale considered.

The amount of incoming solar radiation (the main source of energy to the Earth) varies as a function of Earth's rotation and the zone of study, therefore the SEB at the tropics (where deep convective zones can develop) presents completely different values than those at mid or high latitudes.

Over land, within the ABL, the fluxes are the radiation, the turbulent heat, and the ground heat. These fluxes and their intensity vary depending on the type of the vegetation, the type of the soil, cloud cover, and the presence of some aerosols such as dust. Figure 1-7 shows the different sources (i.e. fluxes) of contribution (over land) to the SEB, when an infinitesimally thin one-dimensional surface with no heat capacity is considered:

- The radiative fluxes ( $F_R$ ) that include the shortwave (SW, spectral interval:  $0.4 \mu\text{m}$  to  $4 \mu\text{m}$ ) solar radiation received at the surface and reflected by it (yellow arrows) and the longwave (LW, spectral interval:  $4 \mu\text{m}$  to  $30 \mu\text{m}$ ) radiation emitted from the surface and received from the atmosphere emission (red arrows).
- The turbulent fluxes ( $F_A$ ) defined as the sum of the sensible  $F_{sens}$  and latent  $F_{lat}$  heat fluxes (green arrows).
- The ground flux ( $F_G$ ) which is the heat flux into the ground measured from a certain depth to the top of the soil.

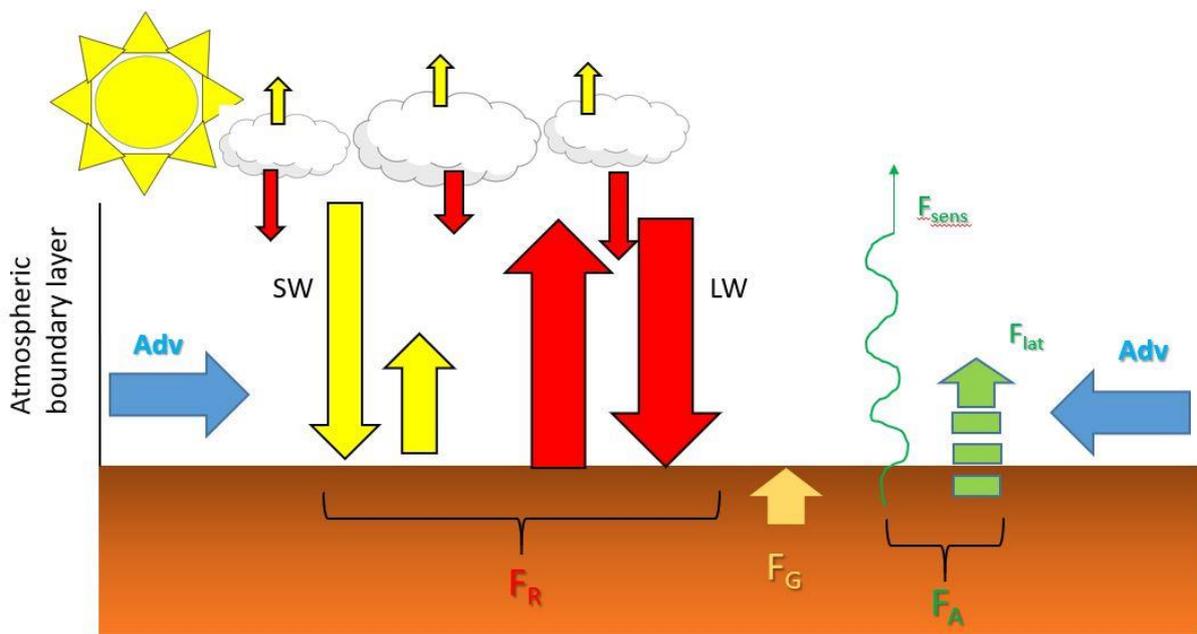
Because we are considering an horizontal homogeneity, the advection of air (blue horizontal arrows in Figure 1-7) does not make part of the surface energy balance but it has to be considered to determine the surface temperature variation (cf. Section 2.2.1.2.), hence it is added to the representation of Figure 1-7.

The SEB represents the equilibrium between the loss and gain of energy at the surface, represented by this equation:

$$F_R + F_A + F_G = 0 \quad (1.1)$$

Where  $F_R$ ,  $F_A$ , and  $F_G$  could be either negative or positive, whether they are energy gains or losses. Indeed, during daytime, on average, the radiative fluxes are positive whereas the other surface fluxes act as modulators with a negative contribution to the SEB. On the contrary, during nighttime, the radiative fluxes are negative, the sensible and ground heat fluxes negative, and no latent heat flux is developed at this time.

These fluxes are detailed in the following sub-sections.



**Figure 1-7:** Contributions to the surface energy balance over land.  $F_R$  is the radiative fluxes,  $F_G$  is the ground heat conduction,  $F_A$  is the surface heat fluxes with  $F_{sens}$  and  $F_{lat}$  (sensible and latent heat fluxes, respectively) and  $Adv$  is advection. The radiative fluxes (yellow and red arrows) are set to be positive when arriving to the surface and negative when leaving it. On the contrary, for the latent and sensible heat fluxes and the ground flux, upwards arrows represent a positive sign whereas downwards arrows are for negative signs.  $Adv$  is set as a horizontal arrow since it does not play part of the SEB but rather on estimating temperature variations (cf. Chapter 2).

### 1.3.1 Radiative fluxes

As previously mentioned, the incident solar radiation, which is primarily modulated by the rotation of the Earth, is the main contributor to surface fluxes. The incident solar radiation received at the surface corresponds to an electromagnetic SW radiative component, and along with the LW infrared radiative component emitted by the

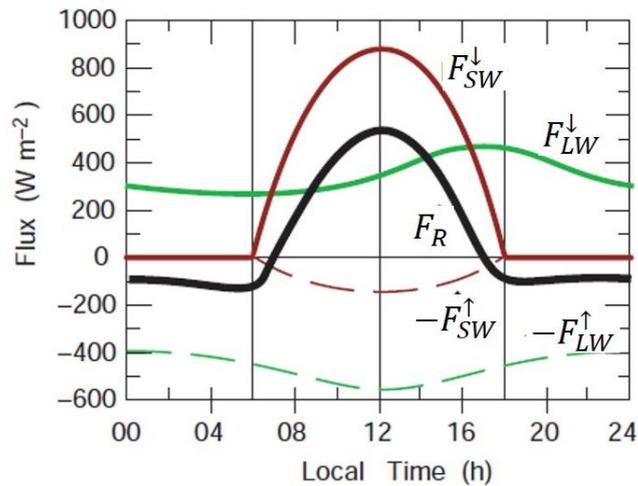
atmosphere and/or the Earth surface, they constitute the surface radiative fluxes. The net radiative flux at the surface is the difference between the radiative energy arriving and the one leaving the surface in these two wavelength bands:

$$F_R = F_{SW}^{\downarrow} - F_{SW}^{\uparrow} + F_{LW}^{\downarrow} - F_{LW}^{\uparrow} \quad (1.2)$$

Where the upwards and downwards arrows represent the upwelling and downwelling radiation. This flux is known to be the external forcing of the SEB. An example of diurnal cycle of these radiative fluxes in fair weather conditions is shown in Figure 1-8. As expected,  $F_{SW}^{\downarrow}$  is about 70% higher than  $F_{SW}^{\uparrow}$  which is approximately the albedo of Earth surface. The upwelling and downwelling LW fluxes are by the same order of magnitude because they depend on the temperature of emission (Stefan-Boltzmann law) which is close between the surface (upwelling emission) and the low-atmosphere (downwelling emission). They also do not vary that much in time during clear-sky conditions, and the  $F_{LW}^{\uparrow}$  presents a peak almost at the same time than  $F_{SW}^{\downarrow}$  because of the rapid response of the surface.  $F_{LW}^{\downarrow}$  depends more on the air temperature and thus its peak is delayed and appears in the late afternoon. Finally, the sum of all these fluxes (black curve) follows a similar behavior as  $F_{SW}^{\downarrow}$  with negative values during nighttime due to the absence of incoming solar radiation and LW cooling of the surface.

If clouds were present, sudden variations of this  $F_R$  flux would be observed because of both their albedo effect in the SW, and their LW emission.

These radiative fluxes can be affected by multiple factors such as weather conditions (cloudy atmosphere), and the presence of aerosols. Markowicz et al. (2021) studied how aerosols affect the radiative fluxes at the surface in Poland in clear-sky conditions, proving that when an episode of high aerosol concentration is presented, net radiative fluxes decrease due to the reduction of incoming solar radiation, affecting also the turbulent surface processes. Some other aerosols have been proved to clearly affected this fluxes, such as dust (Gkikas et al., 2018; Fawole et al., 2019; Mallet et al., 2009) or burning plumes (e.g. Lisok et al., 2018). In addition, the presence of certain aerosols could provoke extreme anomalies of temperature by modifying the radiative surface budget, inducing a radiative warming in the ABL by absorption of solar radiation and contributing to further warming the surface, as shown in Péré et al. (2011).



**Figure 1-8:** Mean diurnal cycle of the contribution of  $F_{SW}^{\downarrow}$  (red line),  $F_{SW}^{\uparrow}$  (red dashed line),  $F_{LW}^{\downarrow}$  (green line),  $F_{LW}^{\uparrow}$  (green dashed line) and  $F_R$  (black line) radiative fluxes in clear-sky conditions. This figure is extracted from Wallace and Hobbs (2006).

### 1.3.2 Turbulent heat fluxes

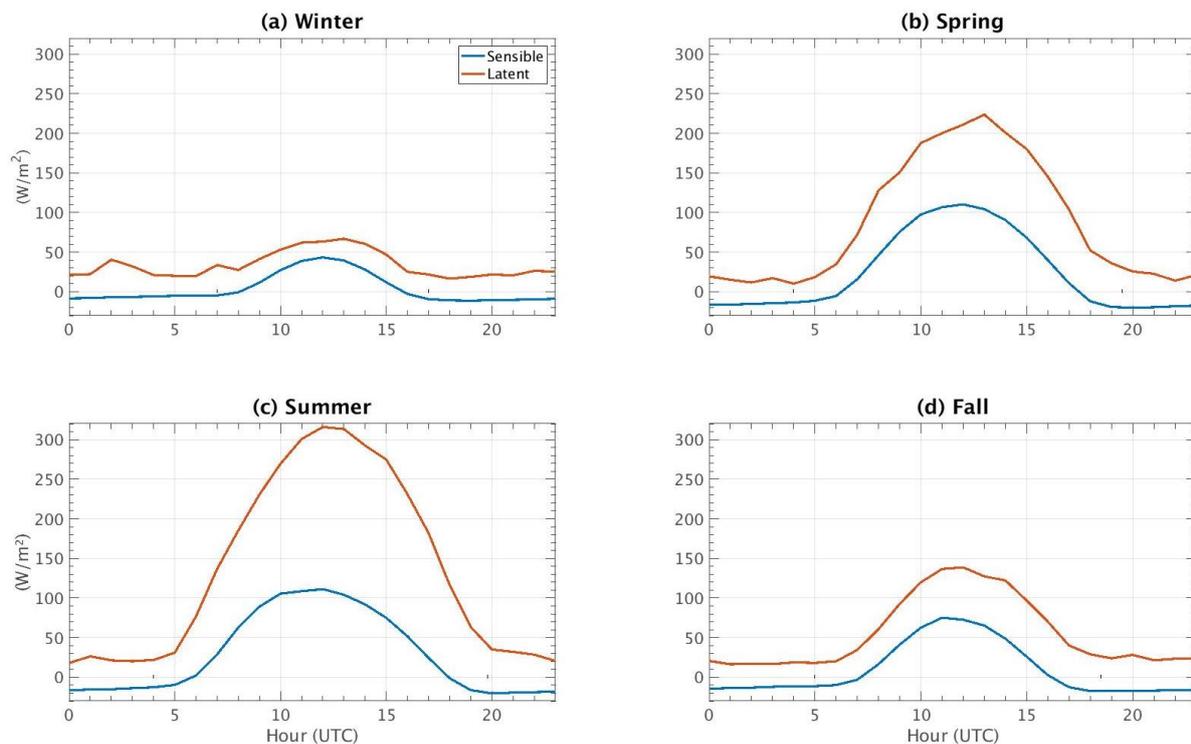
The turbulence heat fluxes at the surface are the sensible and latent heat fluxes. Considering again that we are in a fair-weather situation with weak or no winds, these fluxes tend to follow the solar diurnal cycle (Stull, 1988; Malardel, 2009).

#### 1.3.2.1 Sensible heat flux

The sensible heat flux is a response to surface warming due to incoming solar radiation. Indeed, as the surface warms, it tends to equilibrate its temperature with that in the air just above by removing the excess of heat and thus a sensible heat flux is developing. This turbulent flux can be directly measured with sonic anemometers that measure rapid fluctuations in temperature and vertical velocity, but these instruments are very vulnerable and expensive and thus this sensible heat flux is often estimated based on wind and temperature vertical profiles. Furthermore,  $F_{sens}$  diurnal values are affected in part by the soil properties, and its measurement with good accuracy in all conditions is challenging. In snow-covered surface for instance,  $F_{sens}$  remains low on average due to a stable atmosphere with very low surface temperatures along with a high heat capacity of snow (around 2100 J/kg/K). For such conditions, it remains quite difficult to measure  $F_{sens}$  and thus simulations considering turbulent transfer models previously developed which are adapted to specific snow conditions are used to determine this heat flux (Marks et al., 2008; Schlögl et al., 2018). On forest regions, the Bowen ratio (defined as the ratio of sensible and latent heat fluxes) is instead used to estimate this flux, on the hypothesis of knowing the other components of the SEB, as done by Maruyama and Segawa (2016). The latent heat flux is estimated by the difference between the radiative and ground fluxes, divides by the Bowen ration plus one. This method is widely used to estimate these turbulent fluxes because of its conceptual

simplicity, and therefore can be applied for different surface conditions, such as for Jegede et al. (2001) who used it in a tropical humid region. However, the most commonly method to measure this variable is the eddy covariance method (Foken et al., 2012), which consists on estimating the covariance between turbulent fluctuation of heat and moisture with respect to vertical wind velocity (Panofsky and Dutton, 1984).

For nighttime cases, the sensible heat flux is usually negative because the ground is cooling and hence an opposite effect is seeing compared to daytime hours. Figure 1-9 shows the mean diurnal cycle of the sensible (and latent) heat fluxes split into seasons, measured at the SIRTA observatory (refer to Chapter 2 for further details about these measurements). For all the four seasons, sensible heat flux has a clear diurnal cycle with a peak around noon, whose maximum values are found in spring and summer (Figure 1-9b and Figure 1-9c, respectively). At night, it becomes on average negative and its magnitude does not seem to vary between the seasons. For winter (Figure 1-9a),  $F_{sens}$  remains very low with a maximal mean value not even reaching 50 W/m<sup>2</sup>.



**Figure 1-9:** Mean diurnal cycle of the turbulent sensible and latent heat fluxes in the SIRTA observatory from 2003 to 2017, split into seasons.

### 1.3.2.2 Latent heat flux

This flux is defined as the sum of evaporation (bare soil) and transpiration (due to vegetation). If the surface is moist while it is warming during day, evaporation occurs removing the exceed of heat and a latent heat flux is then developed. Therefore, water vapor is transported from the surface to the atmosphere and thus this heat flux is

generally positive during day and negative during night when water vapor is condensing, but in Figure 1-9, we can see that at SIRTA this flux is on average positive also during nighttime.

The mean diurnal cycle of the latent heat flux is the red line in Figure 1-9, which for this zone is higher than  $F_{sens}$ , especially in, but both mean diurnal cycles present a similar behavior mostly dominated by the solar incoming diurnal cycle (not shown). At this site, this flux is on average positive during the night.

### 1.3.3 Ground heat flux

This  $F_G$  flux is small compared to the other fluxes and is measured from certain depth (between 5 and 50 cm) to the top of the ground. It is quite hard to measure because it includes conduction heat transfer processes.  $F_G$  is negative when the surface is warmer than the layers below (and thus cooling the surface) and positive when the opposite effect occurs (which generally occurs during the nights and contributes to warm the surface). Despite its really low value in magnitude (see Figure 1-11), this flux cannot be neglected in general circulation models (GCM). The type of soil and soil moisture will govern the magnitude of this flux since not all of them present the same properties and characteristics.

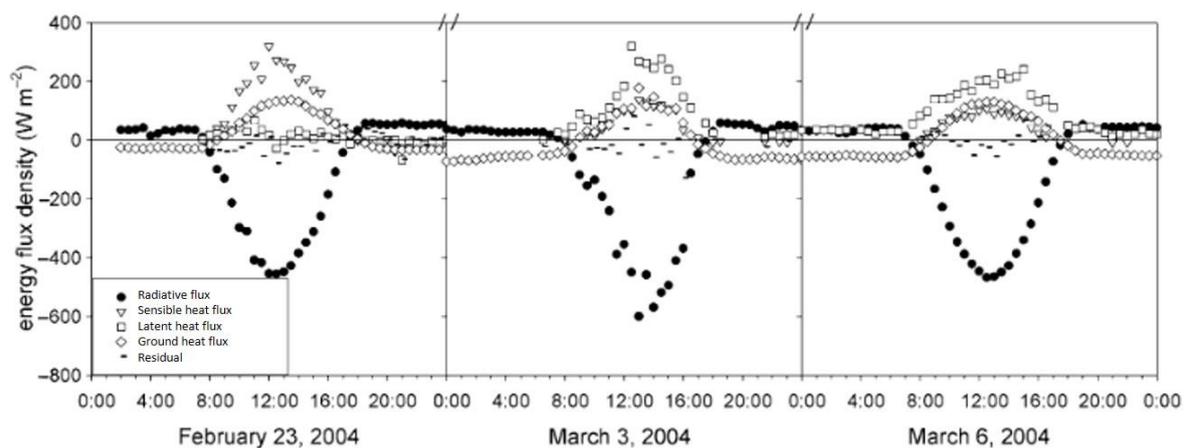
Stull (1988) proposed a simple parametrization to estimate  $F_G$  by assuming that it represents a certain percentage of the net radiative flux at the surface. He suggested that it represents a 10% of the total net radiative flux for daytime and a higher percentage of 50% for nighttime. Other simple parametrization is assuming  $F_G$  equivalent to 30% of the sensible heat flux. Other more complex and detailed methods have been elaborated in the last decades whose application depends on the area of study, the type of soil and the time of the day. As for the variables used as inputs, some of the approaches use the soil temperature (Alvalá et al., 2002; Jacobs et al., 2011), others use the thermal diffusivity of the soil (Wang and Bou-Zeid, 2012), and others the turbulent sensible heat flux (Liebethal and Foken, 2007). At SIRTA, this flux and its contribution to temperature variation is not directly measured but another parametrization is used based on the temperature difference between the near-surface temperature and the temperature at a specific depth (see Chapter 2, Section 2.2.1.2. for further details).

### 1.3.4 Diurnal and annual cycle of the SEB terms

All the aforementioned terms are the main source of energy within the ABL. Their estimation methods vary depending on the zone, and the availability of observations as previously stated. Nevertheless, when observations are available, they can be estimated using direct measurement of each of these fluxes, as done for example by

Mauder et al. (2007) in Nigeria, and shown in Figure 1-10. This figure presents the diurnal values of each SEB term for three different days having different soil conditions (details are given in the figure caption). A negative daily cycle is found for the flux radiation for the three days (in their study, they consider the net flux radiation to be negative in the SEB equation) where cloudy conditions predominate, with some fluctuations on 03/03/2004 due to variations on cloud cover along the day. High soil moisture conditions (middle and right figures) enhances higher latent heat fluxes values than when the soil is dry and then sensible heat flux is higher (left figure), and then ground heat flux becomes the same order of magnitude than the sensible heat flux.

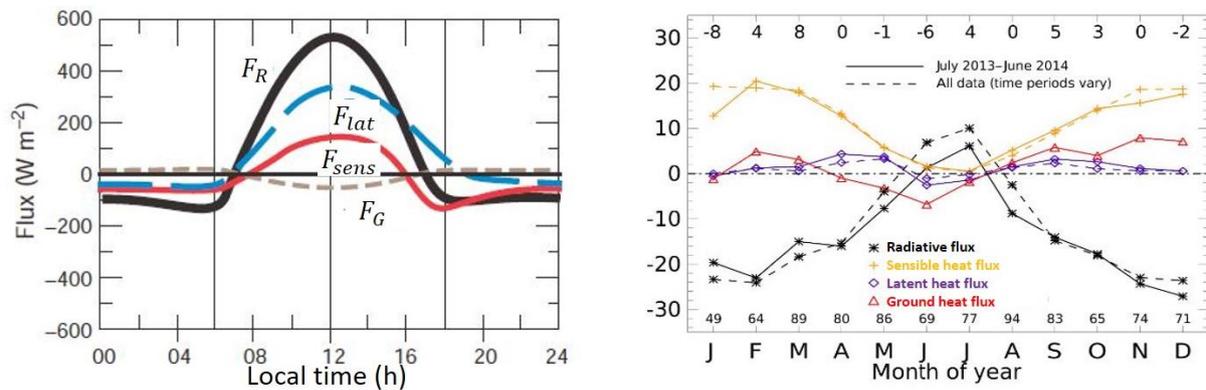
The residual in Figure 1-10 is due to the fact that the sum of  $F_{sens}$ ,  $F_{lat}$  and  $F_G$  is not exactly equal to  $F_R$  (cf. Eq. 1.1), is the highest for 03/03/2004 when cloudiness varied during day hours. Indeed, this residual is known as the non-closure of the surface energy balance and is defined as the failure of the sum of the turbulent heat and ground fluxes to account for the available energy provided by the radiative fluxes. This non-closure may be due to a failure of the instrument's measurements, the time-scale considered and the methods used to estimate each of these fluxes (e.g. Russell et al., 2015; Leuning et al., 2012; Oncley et al., 2007) or even the wrong estimation of the air advection (Gao et al., 2017). Nowadays understanding the causes of this non-closure and how measures and parametrization can be improved remain a very important topic on the research community.



**Figure 1-10:** Directly measured SEB terms at Ile-Ife, Nigeria, for the 02/23/2004 with low soil moisture and low nighttime water vapor (left), 03/03/2004 with high soil moisture and low nighttime water vapor (middle), and 03/06/2004 with high soil moisture and high nighttime water vapor (right). This figure is extracted and adapted from Mauder et al. (2007).

Other example of typical diurnal cycle of these SEB fluxes (cf. Eq. 1.1) is given in Figure 1-11 (left) when these fluxes represent a daily mean average on clear-sky conditions. This time the energy gain or loss is well partitioned among  $F_{sens}$ ,  $F_{lat}$  (the two being

positive), and  $F_G$  (that remains small in comparison to the other fluxes and negative). The annual cycle of this surface fluxes estimated in the Arctic by Miller et al. (2017) is presented in Figure 1-11 (right). Due to the lack of observations at this zone, they performed different parametrizations (e.g. using the Richardson number for the turbulent heat fluxes) to estimate each of these fluxes. One clear difference between the daily cycles from Figure 1-10 and the annual ones in Figure 1-11 (right) is the order of magnitude of the fluxes. Indeed, the zone of study of Figure 1-11 (right) being the Arctic, barely sensible and latent heat fluxes developed because of the weak solar radiation (especially latent heat flux where nearly not evaporation at all occurred in this zone during all the year) and vegetation, thus the low values for all the surface fluxes. For winter in Figure 1-11 (right), sensible heat fluxes are higher than the other ones throughout all the year thanks to the presence of snow which happens to be even warmer than the surface and the layers immediately above.



**Figure 1-11:** Diurnal cycle of the four SEB terms ( $F_R$  in black,  $F_{lat}$  in dashed blue,  $F_{sens}$  in red and  $F_G$  in dashed brown lines) under clear-sky conditions (left). Annual cycle of the four SEB terms ( $F_R$  in black,  $F_{lat}$  in purple,  $F_{sens}$  in yellow and  $F_G$  in red) (right). The figure on the left is adapted and extracted from Wallace and Hobbs (2006) and the one on the right from Miller et al. (2017).

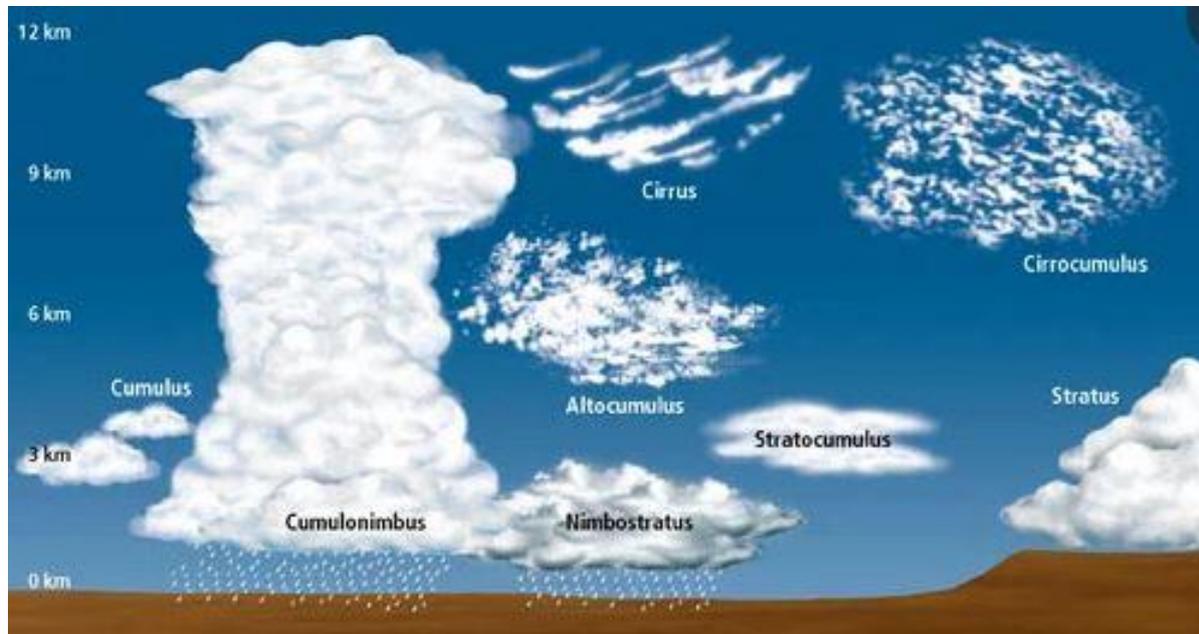
## 1.4 CLOUDS PROCESSES AFFECTING SEB

As it is widely known, one of the most important factors controlling the incoming solar radiation at the surface are clouds, and their study and comprehension are increasing with the objective to have better estimate of their effect on SEB and radiative fluxes and consequently, improve forecasting models. In this section, we present clouds general influence on the SEB terms and then on surface temperature.

### 1.4.1 Types of clouds and their composition

Clouds are defined as a set of either water droplets or ice crystals, or a mixture of the two of them. When this mass of air reaches the surface, it is not considered as cloud but as fog. Clouds form when water vapor condensates in the atmosphere. This condensation process occurs thanks to a decrease in pressure as the mass of moist air arises, letting it to expand and cooling adiabatically. The air therefore becomes supersaturated, and it condensates around soluble microscopic particles, the condensation nuclei.

Clouds are mostly classified by the height on which their base is found, and at the same time, this height conditionate their composition and radiative effects. Figure 1-12 presents a simple scheme of the different types of clouds that are found in the troposphere at mid-latitudes. Generally, low-level clouds correspond to heights between near-surface and 2 km, mid-level clouds between 2 and 7 km, and high-level clouds above 7 km. Because of their high altitude and then cold temperatures, cirrus clouds are composed by ice crystals, whereas low clouds such as stratocumulus or stratus contain water droplets. Clouds such as cirrocumulus can contain a mix of water droplets and ice crystal, but ice crystals tend to dominate their composition because of the evaporation of the water vapor presented in the droplets that later is deposited onto the ice crystals.



**Figure 1-12:** Scheme of the various types of clouds presented in the troposphere. This figure is extracted from <https://www.rochesterfirst.com/weather/weather-glossary/cloud-index-defining-types-of-clouds-and-their-unique-characteristics/>.

Because of their difference composition, hereby are listed some typical clouds and their effects on the SEB and temperatures at the surface depending on their macro and

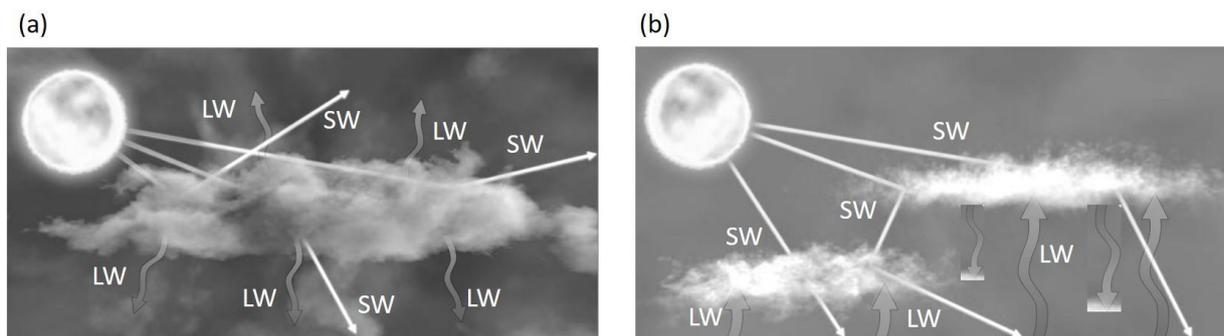
microphysical properties:

- Stratocumulus clouds (classified as low-level clouds) are convective cloud systems, whose formation and development imply radiation and turbulence processes as well as ABL physics occurring in large temporal scales (Wood, 2012), and thus they are coupled with the general circulation (Bretherton and Hartmann, 2009).
- Cumulus clouds are boundary-layer clouds (i.e. low-level clouds) because they form just below the top of the ABL (Wallace and Hobbs, 2006), as seen in Figure 1-4 in the afternoon, and therefore tend to cool the surface by reflecting solar incoming radiation back to the space. Indeed, their cloud cover varies rapidly in time, and diurnal variability of turbulent fluxes is more variable than when a stratocumulus convective cloud sets in the sky and do not allow these fluxes to well develop and they remain very low and constant along the day. Along with the stratocumulus, cumulus are the most frequent clouds for both ocean and land surfaces in the mid-latitudes (González et al., 2020), playing a major role also in temperature and the development of other types of clouds. The vertical extension of cumulus is of few meters, while the stratocumulus clouds penetrate beyond the capping inversion zone up to the free atmosphere and thus their vertical extensions is of some km.
- Cirrus clouds (high-level clouds) prevent the Earth's LW radiation to escape and combined with the telluric LW radiation they emit at very low temperatures, they tend to warm the lower atmosphere and the surface. Their radiative properties depend on their optical depth (e.g. Mauno et al., 2011) and the cloud top height, but they are the only type of clouds having (but not all of the times) a radiative warming effect (Kienast-Sjögren et al., 2016; Lolli et al., 2017). Moreover, the ice crystals composing cirrus clouds have different shapes and orientations affecting the extensions and scattering of light, which makes their radiative effect more difficult to study than liquid clouds one.

#### 1.4.2 Radiative properties

The main characteristic of clouds that is used in this thesis is their radiative effect because of their role on the SEB, as seen before. As previously stated, each type of cloud will have different radiative properties depending on their macrophysical (bottom and top cloud base, horizontal extension...) and microphysical (composition/particle size...) properties. Clouds produce two opposite radiative effects: they reflect a portion of the solar incoming SW radiation via their albedo effect (cooling effect), and they absorb and re-emit part of the Earth's LW radiation through a greenhouse effect (warming effect). High-level clouds have a weak albedo effect because they are optically thin, but they both absorb and re-emit a significant amount of LW radiation (Figure 1-13a) and overall, the difference between their albedo and greenhouse effect is weak, they could sometimes warm or sometimes cool the surface.

On the contrary, low-level clouds, because of their high amount of water droplets and hence important optical thickness, reflect a significant portion of the solar incoming SW radiation (Chen et al., 2000) and thus preventing the surface to warm, having on average a cooling effect on surface temperatures (Figure 1-13b). These radiative properties of clouds are quantified by the Cloud Radiative Effect variable (CRE), separated in the SW and LW domains. This CRE is defined as the difference between the radiative clear-sky and cloudy fluxes (cf. Section 2.2.1.2. – Eq. 2.23). Furthermore, clouds tend to cool during daytime and warm the lower layers of the atmosphere during nighttime.



**Figure 1-13:** Explanatory diagram of the radiative effect of (a) low- and (b) high-level clouds. This figure is extracted and adapted from [https://www.nasa.gov/pdf/135641main\\_clouds\\_trifold21.pdf](https://www.nasa.gov/pdf/135641main_clouds_trifold21.pdf).

The physical processes that create the high-clouds are not well resolved in the models, adding further uncertainty to the radiative estimations. Several parametrizations have been lately established to quantify these effects and compared them with observations (e.g. Zeng et al., 2021; Wang and Penner, 2010; Powell et al., 2012; Zhou et al., 2012), but still a higher land-instrument coverage is required to expand the observations and improve models. With respect to low-level and convective clouds, uncertainties lie for instance on the difficulties to clearly identify their cloud bases (and thus models may miscalculate their effects by considering wrong macrophysics).

### 1.4.3 General impact on SEB and 2 m temperature

The radiative fluxes on the SW and LW domain arriving at the surface are conditioned by the presence of clouds, and because these radiative fluxes are the main incoming source of energy of SEB, turbulence in lower layers of the ABL is also indirectly affected by clouds. In mid-latitudes zones, the role of clouds on the SEB, precipitations and other surface parameters such as soil moisture has been studied in order to better understand their relationship with these variables and to improve the different regional climate models (Cheruy et al., 2014; Boé, 2013; Santanello et al., 2011). Moreover, the

scientific community has addressed a special focus on the polar regions and the influence of clouds on the SEB due to the global warming and their relationship to the melting events of the last decades (e.g. Sedlar et al., 2010; Broeke et al., 2006; Bennartz et al., 2013; Miller et al., 2017; Van Tricht et al., 2016). These zones have indeed become a point of interest for studying the Earth's responses to a changing climate.

As for the effect of clouds on temperature, they could be involved in some extreme hot and cold events. For instance, the absence of low-level clouds in Europe was in part responsible of the extreme heat wave experienced on July 2006, by amplifying the daily surface temperature, as shown by Chiriaco et al. (2014). This absence of clouds was coupled with very dry soils that constraint the development of convective clouds and hence no precipitations occurred. Furthermore, McCaffrey et al. (2019) showed that cold pool events can last for several days in winter when a persisting cloud cover is set, which decreases the ABL growth rate and thus facilitating stable conditions to set when cold air is trapped near the surface. On the contrary, Betts et al. (2016) illustrated that the warmest temperatures during winter in Canada were found during cloudy conditions, when the LW warming by clouds dominates temperatures. An opposite effect is found for summer season when the coolest surface temperatures are found under cloudy skies due to a strong SW reflection back to space by clouds.

## **CONCLUSION**

The European climate, temperature variability and anomalies are affected by both the circulation of large-scale air masses and by the smaller-scale processes such as cloud radiation and surface-atmosphere interactions located within the atmospheric boundary layer. Observations of these small-scale variables are then necessary to further study their influence on local surface temperature variations. In the next chapter, the dataset, the temporal and spatial scales used in this study, along with a SEB prognostic model are presented in order to answer the following scientific questions:

- What are the characteristics (values and variability) of SEB terms and advection, as they drive surface temperature variations?
- Which term dominates surface temperature variations, depending on the location, the large-scale condition, and the temporal scale?
- What is the specific role of clouds in one-hour temperature variations?



## 2 CHAPTER 2: DATA AND METHODS

---

### INTRODUCTION

Thanks to the construction of several observatories in recent years, it is now easier for scientists to study all atmospheric processes based on direct measurements and observations, and thus to improve climate models to have better estimates of future extreme events and anomalies. In this chapter, we present a dataset in the first section that gathers all the necessary variables, allowing us to quantify the contribution and importance of local surface processes on hourly and local scale temperature variations, at an observatory located in the Paris region. Then, a second data set from an observatory in Toulouse is also presented in order to subsequently evaluate the spatial variability of the results obtained by comparing them with those of the first site.

In a second time, we present a method to estimate the contribution of each term involved in surface temperature variations in an hourly scale, using almost exclusively ground-based observations, with only few parametrizations and hypothesis. Two different models are applied to link these terms to the temperature variations: a simple sum, and a random forest method, which is also used to determine the importance of each term.

### 2.1 DATA

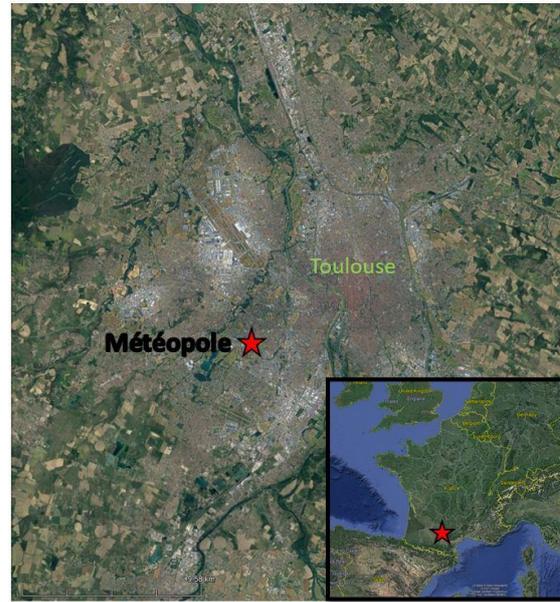
#### 2.1.1 The SIRTA observations

All over the world, several observatories have been built during the last decades to measure meteorological variables to study and characterize the atmosphere and better understand the processes involved in it. As seen in the first chapter, different surface, ground and atmospheric variables are necessary to analyze local temperature variations. The SIRTA (Site Instrumental de Recherche par Télédétection Atmosphérique) observatory allows retrieving all of these variables thanks to the wide range of passive and active instruments that are ground-based installed there and which are dedicated to the study of the atmosphere, clouds, fog, soil and other atmospheric variables that are used to address different scientific projects. This observatory is located in a semi-urban area 20 km Southwest of Paris (48.71° N, 2.20° E) with an altitude of 160 m above sea level as shown in Figure 2-1a, collecting long-term meteorological variables since 2002 for the oldest ones (Haeffelin et al., 2005). Thus, the SIRTA observatory is able to continuously measure dozens of variables on an hourly scale for more than a decade.

(a)



(b)



**Figure 2-1:** Satellite image of (a) the Paris region and (b) the Toulouse region. The yellow star in (a) indicates the position of the SIRTAs Supersite and the red one in (b) where the Météopole site is located. Bottom right in both figures: overview map. Figure extracted from the Google Earth tool.

Figure 2-2 presents photos of some of the instruments used to retrieve the variables we used in the study. For instance, solar and infrared radiometers are used to retrieve SW and LW radiation arriving at the surface, sonic anemometers along with hygrometers allow to obtain turbulent fluxes such as latent and sensible heat fluxes based on wind velocities, temperature and water vapor fluctuations, etc. Collecting these variables favors the development of several long-term monitoring studies of the atmosphere since all of them are collected in a single site. Moreover, the observatory is located in a geographical position where different large-scale atmospheric circulations meet, which makes it an important and strategic zone to assess each of their impacts on clouds, aerosols, radiation, relative humidity, surface temperature and other boundary layer variables controlling atmospheric processes. For instance, Cheruy et al. (2013) analyze atmospheric physics based on SIRTAs observations to compared them with some models outputs, Dupont et al. (2012) study fog formation and dissipation based on surface observed fluxes, etc.

### 2.1.1.1 The SIRTAs-ReOBS dataset

Many of the variables retrieved from the instruments available in the SIRTAs observatory are gathered in one single NetCDF file named SIRTAs-ReOBS. ReOBS project aims to synthesize all observations available at a single observatory at an hourly time scale with

an exhaustive data-quality control, calibration and rigorous treatment, in a single NetCDF file. Hence, the SIRTA-ReOBS dataset (Chiriaco et al., 2018) contains more than 60 variables. This dataset represents a great advantage in atmospheric studies since it allows to perform and carry out multi-year and multi-variables synergetic views of the atmosphere (e.g. Ringard et al., 2019; Bastin et al., 2019).

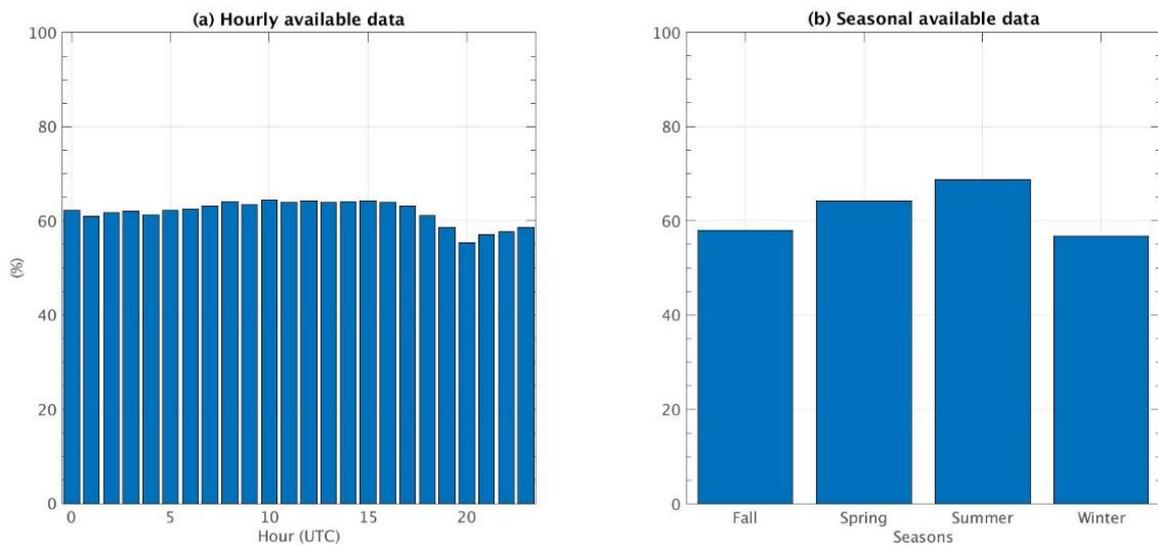


**Figure 2-2:** Photos of some of the instruments installed in the SIRTA observatory. This figure is extracted from Chiriaco et al., (2018).

*2.1.1.2 Meteorological variables to study 2 m temperature variations*

After a rigorous inventory, all the necessary hourly variables requested by our current study are available for the period going from January 2009 to February 2014 and are listed in Table 2-1 allowing to perform a multi-year analysis, variables that are more detailed in Section 2.2. Some variables that were retrieved from measurements at the SIRTA observatory sometimes present some gaps due to instrumental issues (Chiriaco et al. 2018; Pal and Haeffelin 2015), and thus their percentage of availability is also presented in Table 2-1. Indeed, Figure 2-3 presents the histograms of data availability when all the variables are simultaneously available for the period of interest. The availability of data is quasi-homogeneous and is around 60% for all hours (Figure 2-3a).

Gaps are most of the time due to the absence of the Mixing Layer Depth MLD, which is the height of the ABL) variable for some days, extended sometimes for more than two months (not shown) which reduces MLD availability to 71%. Other gaps are caused by the absence of radiative variables (e.g. 9% of missing data, see Table 2-1). From the point of view of seasons, summer is the period with the best data coverage (70%) and winter the less good (56%) (Figure 2-3b), whose absences are due precisely to gaps in the MLD variable. This absence is due to intermittent instrument failures from the lidar used (ALS-450, manufactured by Leosphere, France) that extended sometimes for a whole season, failures which did not allow to have measurements of some of the variables used to estimate the MLD (see Pal and Haeffelin, 2015 for further details).



**Figure 2-3:** Histograms of (a) hourly and (b) seasonal available data in the SIRTA ReOBS dataset, considering all the variables needed for the period of study 2009 – 2014.

Furthermore, some of the variables used are directly measured, such as  $T_{2m}$ ,  $T_G$  and the radiative surface upwelling and downwelling fluxes. Others such as the turbulent latent and sensible heat fluxes, and surface downwelling radiation for clear-sky conditions (e.g.  $F_{SW,CS}^{\downarrow}$  and  $F_{LW,CS}^{\downarrow}$ , defined as the clear-sky components of downward SW and LW radiation) are treated and derived from parametrizations using other classical meteorological variables (refer to Table 2 in Chiriaco et al., 2018 for further details). Figure 2-4 and Figure 2-5 show the hourly values of all the variables listed in Table 2-1 (except the temperature at the mixing layer depth which will be further explained in Section 2.1.3.1.). For this particular period, surface temperatures at SIRTA decreased up to  $-11\text{ }^{\circ}\text{C}$  in the winter of 2012 and reached a maximal temperature of  $37\text{ }^{\circ}\text{C}$  in summer at the same year (Figure 2-4a). As expected, both  $F_{SW}^{\downarrow}$  and  $F_{SW,CS}^{\downarrow}$  presents greater values than the LW radiation downward fluxes (Figure 2-4d and Figure 2-5b, respectively) reaching  $\sim 1000\text{ W/m}^2$  for the first ones whereas the LW downward radiation for all and clear-sky conditions range from 200 to  $500\text{ W/m}^2$  with a marked seasonal cycle. On the contrary,  $F_{LW}^{\uparrow}$  is the one with higher radiation values than  $F_{SW}^{\uparrow}$  since the surface

emits more radiation in the LW domain than the amount it can reflect in the SW domain, since  $F_{LW}^{\uparrow}$  depends on surface temperature and  $F_{SW}^{\uparrow}$  depends more on the amount of solar radiation that arrives to the surface and on its ability to reflect it, via its albedo effect, which depends itself on surface properties. The latent heat flux availability in the SIRTA-ReOBS file is very low for the period of study, as seen in Figure 2-5d, an issue that affects (but still allows) the estimation of one of the SEB terms as detailed in Section 2.2.1.2.b.

It is worth noting that the clear-sky downwards radiation fluxes (Figure 2-5a and Figure 2-5b) are not directly measured but rather derived using different parametrizations and radiative flux analysis techniques. Indeed, Long and Turner (2008) used LW radiation, surface temperature and humidity measurements (retrieved from based-ground instruments and radiosoundings launched 3 times/day) to detect periods of clear-sky conditions, and with these detected periods a methodology is applied to fit functions that allow to produce a continuous estimate of  $F_{LW,CS}^{\downarrow}$ . On the other hand, Dutton et al. (2004) developed a different methodology to estimate  $F_{SW,CS}^{\downarrow}$ , based on interpolated clear-sky observations using other meteorological variables such as the zenithal angle, surface albedo, and some cloud properties. In the SIRTA-ReOBS file, these clear-sky radiative fluxes are estimated based on Long and Turner (2008), where they used the variables previously mentioned to calculate a lapse coefficient in clear-sky days to later interpolated this coefficient to cloudy periods.

Furthermore, MLD measurements are calculated using an algorithm applied at SIRTA (STRAT+) and developed by Pal et al. (2013) using lidar observations based on aerosols stratification and cloud base height.

In addition, to get vertical information about clouds, the SIRTA-ReOBS dataset contains lidar profiles measured from a 532 nm lidar and calibrated with the STRAT algorithm (Morille et al., 2007). Then, to obtain the vertical description of the atmosphere, an hourly lidar Scattering Ratio altitude-intensity matrix is estimated as follows:

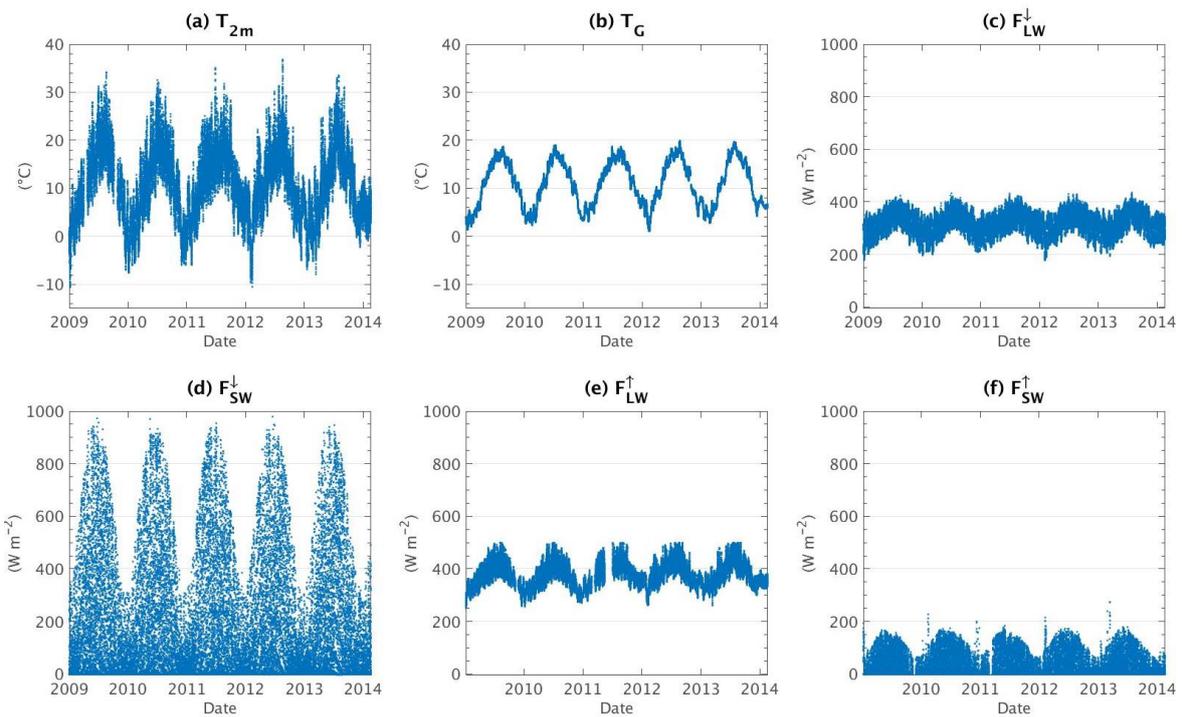
$$SR(z) = \frac{ATB(z)}{ATB(z)_{mol}} \quad (2.1)$$

Where ATB is the total attenuated backscatter lidar signal and  $ATB_{mol}$  is the signal in clear-sky conditions. The value of  $SR(z) = -999$  corresponds to non-normalized profiles, the value  $-777$  represents the profiles that cannot be normalized due to presence of a very low opaque cloud and the value of  $-666$  is set as invalid data. Then, bins located in the range  $0.01 < SR(z) < 1$  are for clear-sky conditions (in the case of fully clear-sky conditions,  $ATB = ATB_{mol}$  and thus  $SR(z) = 1$ ),  $1.2 < SR(z) < 5$  is defined as unclassified data, and for cloud detection, a threshold of  $SR(z) > 5$  is set. Details are given in Chiriaco et al. (2018) and Chepfer et al. (2010). Despite the fact that the instrument does not operate permanently (it does not operate when it is raining, when it's nighttime and

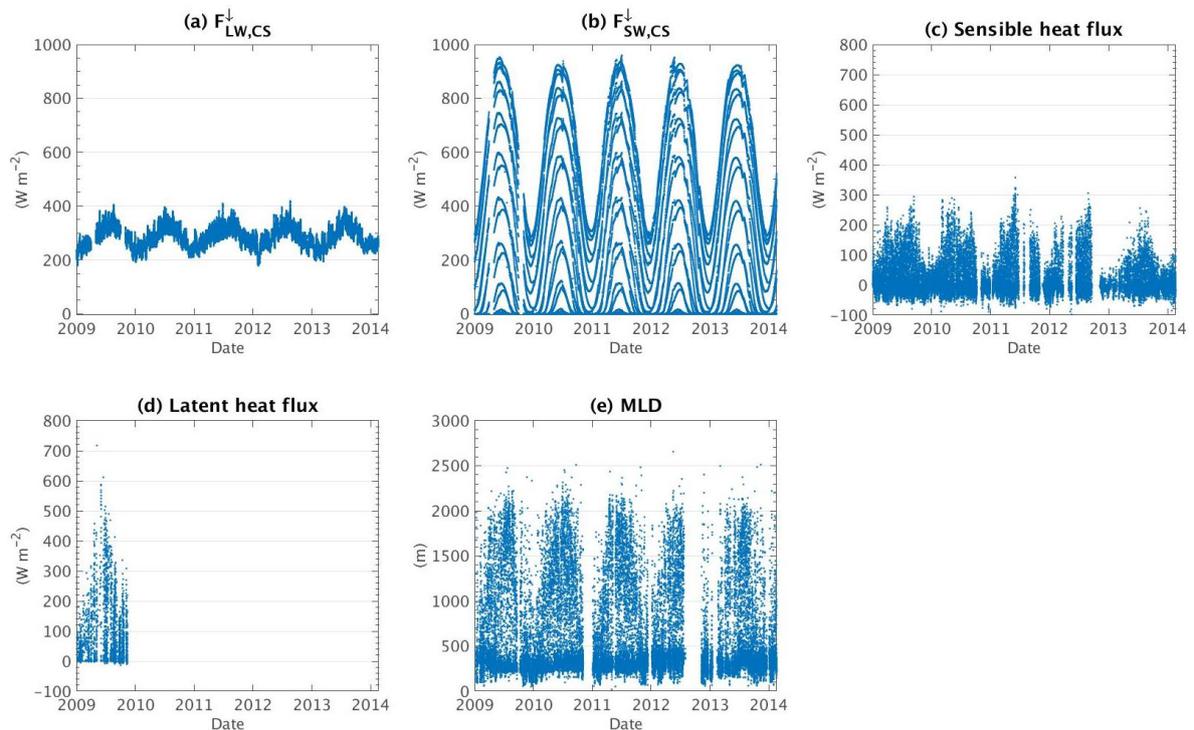
on the weekends), it is very powerful to get information on the vertical structure of the atmosphere (detection of aerosols and clouds) and so understand the category of clouds acting on the temperature variations.

**Table 2-1:** Variables available in the SIRTA-ReOBS dataset and the Météopole site used as inputs in the current study. The percentage of availability of SIRTA-ReOBS is from January 2009 to February 2014, whereas for the Météopole site is from January 2017 to December 2018. \* Variable not directly available in the dataset but estimated otherwise (see text for further information).

Variable, unit	Notation	SIRTA-ReOBS (%)	Météopole (%)
2 m air temperature, K	$T_{2m}$	97	100
Soil temperature 20 cm below the ground, K	$T_G$	99	100
Temperature at the mixing layer height, K	$T_{MLD}$	71*	100*
Surface downwelling LW radiation, $W/m^2$	$F_{LW}^{\downarrow}$	100	100
Surface downwelling SW radiation, $W/m^2$	$F_{SW}^{\downarrow}$	100	100
Surface upwelling LW radiation, $W/m^2$	$F_{LW}^{\uparrow}$	91	100
Surface upwelling SW radiation, $W/m^2$	$F_{SW}^{\uparrow}$	94	100
Surface downwelling LW radiation for clear-sky, $W/m^2$	$F_{LW,CS}^{\downarrow}$	97	100*
Surface downwelling SW radiation for clear-sky, $W/m^2$	$F_{SW,CS}^{\downarrow}$	97	100*
Surface upward sensible heat flux, $W/m^2$	$F_{sens}$	73	95
Surface upward latent heat flux, $W/m^2$	$F_{lat}$	5	74
Mixing layer depth, m	MLD	71	100



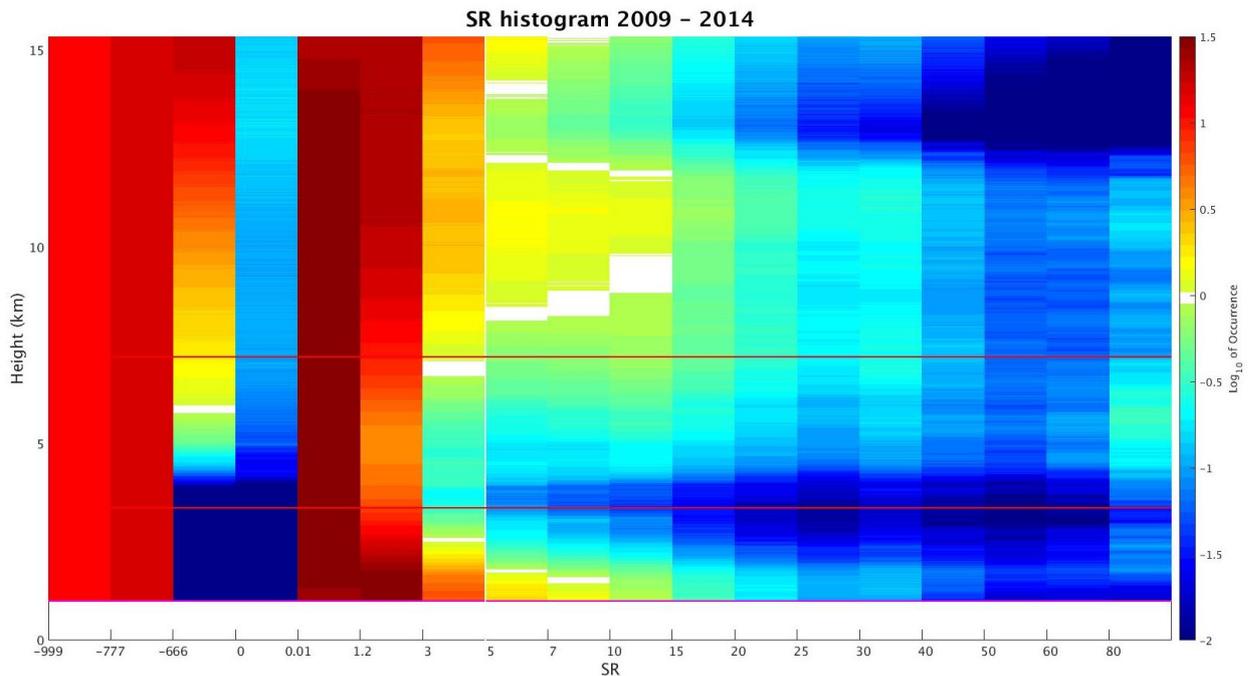
**Figure 2-4:** Hourly values from January 2009 to February 2014 of (a)  $T_{2m}$ , (b)  $T_G$ , (c)  $F_{LW}^{\downarrow}$ , (d)  $F_{LW}^{\downarrow}$ ,  $F_{LW}^{\uparrow}$  and (f)  $F_{SW}^{\uparrow}$ .



**Figure 2-5:** Same as Figure 2-4 but for (a)  $F_{LW,cs}^{\downarrow}$ , (b)  $F_{SW,cs}^{\downarrow}$ , (c) sensible and (d) latent heat fluxes, and (e) MLD.

A simple SR histogram built from the lidar data available in the SIRTA-ReOBS file is

shown in Figure 2-6 for the period of interest. Since the lidar does not operate while it is raining, the information corresponds to non-precipitant clouds. On average, a dominance of low- and high-level thin clouds is noted ( $5 < SR(z) < 25$ ), while mid-level thicker clouds also appear at the SIRTA observatory. Clouds that completely attenuate lidar photons are included in the three left columns: their amount is very important. Note that for this period, few high and low-level clouds that are optically thick are detected ( $SR > 80$ ).



**Figure 2-6:** Lidar scattering ratio ( $SR(z)$ ) histogram obtained by cumulating all lidar observations during daytime from 2009 to 2014. The color bar is the logarithm of the percentage of occurrence (the sum of one level is equal to  $\log_{10} 100\%$ ); lidar data showed in each subplot start above the instrument's recovery altitude ( $z = 1$  km); the red horizontal lines represent the limits of low-, mid- and high-level clouds; and the white vertical line shows the threshold of clouds detection ( $SR(z) = 5$ ).

### 2.1.2 The Météopole observations

Another observatory is here introduced to better understand the spatial distribution of the terms controlling surface temperature variations: the Météopole. This observatory is located in the METEO-FRANCE campus at the Southwest of Toulouse ( $43.57^\circ$  N,  $1.38^\circ$  E, 156 m above sea level, Figure 2-1b). This observatory is then located in a more urban zone than SIRTA (which is located in a semi-urban area), because of its proximity to the city center and it is widely surrounded by buildings and other installations. With respect to the SIRTA observatory, the Météopole site does not account with a single file where all the measures are already gathered. However, the data are well calibrated and have undergone rigorous treatments to their proper use in several technical and scientific

atmospheric projects (cf. the web site of the Centre National de Recherches Météorologiques – Meteo-France to see all the projects that are carrying out, <http://www.umr-cnrm.fr/spip.php?rubrique204>).

For the current study, the same variables used in the SIRTA observatory are requested for the Météopole and are listed also in Table 2-1 along with its availability from January 2017 to December 2018. The two downwards clear-sky radiative ( $F_{SW,CS}^{\downarrow}$  and  $F_{LW,CS}^{\downarrow}$ ), the same algorithm used in SIRTA is applied for the data (Long and Turner, 2008; Dutton et al., 2004). They have been specifically computed by J. Badosa for this thesis. As for the MLD at Toulouse, a different approach from the one used at SIRTA is applied to estimate this height, and thus at summer, a time when the ABL is higher at lower latitudes, in Toulouse MLD is lower than at SIRTA (see Appendix A.1, Figure A-1). The approach used at the Météopole to estimate this height consists on a machine learning method considering a classification problem and using a supervised method (further details are found in Rieutord et al., 2021).

### 2.1.3 ERA5 and ERA5-Land Reanalysis

The ERA5 dataset consists of an atmospheric reanalysis of the global climate, and is now on its 5<sup>th</sup> and latest generation release produced the European Centre for Medium-range Weather Forecasts (ECMWF) (ERA5: Fifth generation of ECMWF atmospheric Reanalysis of the global climate, 2020). It provides hourly data on different atmospheric, land-surface and sea-atmosphere parameters along with their uncertainties from 1950 to the present. One of its great advantages is that it covers the whole globe on latitude-longitude grids at  $0.25^{\circ} \times 0.25^{\circ}$  resolution (i.e. 31 km horizontal resolution), and with atmospheric parameters on 40 pressure levels from 1000 to 1 hPa (Hoffmann et al., 2019; Horányi, 2017), with a continuous daily updated frequency.

ERA5 reanalysis has been combined with several observations and models providing consistent and reliable analysis on widely atmospheric fields. Since this reanalysis allows to have a set of meteorological variables (previously assessed and calibrated) in an hourly scale over several decades thanks the addition of newly reprocessed datasets and instruments, their use ensure a good representation of the main Earth system processes. However, Reanalysis presents some problems in the proper management of capturing changes in the observing system, and several studies show that under certain conditions and in certain areas of the world, certain variables present some errors and unrealistic values for some locations at specific dates (Hersbach et al., 2020; Long et al., 2017; Dee and Uppala, 2009; Davis et al., 2017).

Table 2-2 presents the variables retrieved from both the ERA5 and ERA5-Land datasets and used as inputs in this study. Note that these variables are available in both SIRTA-

ReOBS and Météopole datasets, but due to the requirements of some terms acting on surface temperature variations, it is necessary to have other measurements next to the observatories make necessary the use of this reanalysis. Indeed, the temperature at 2 m and the two components of the wind speed at 10 m are used to estimate the advection term, whereas  $T_{MLD}$  is used to calculate the atmospheric heat exchange (see Section 2.2.1.2.).

The surface variables required for our study are retrieved from the ERA5-Land hourly Reanalysis (cf. following sections) since this dataset have a better horizontal resolution (horizontal resolution of  $0.1^\circ \times 0.1^\circ$ , Copernicus Climate Change Service, 2019).

Figure 2-7 shows the hourly differences between the temperature from each *in situ* dataset and the one retrieved from ERA5 ( $\Delta T = T_{2m,observatory} - T_{2m,ERA5}$ ). This difference is most of the time significant between the two datasets, an issue that must have been considered when evaluating the performance of the model further developed since it already adds some bias to the calculation.

**Table 2-2:** Variables retrieved from ERA5 dataset and used as inputs in the current study.

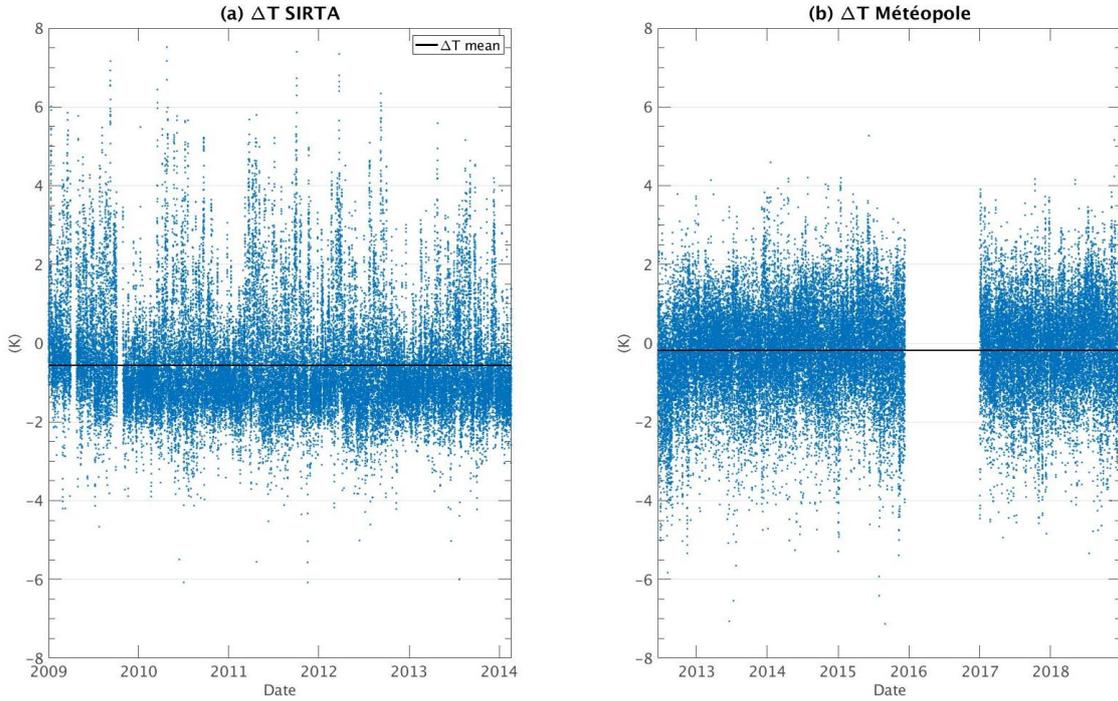
Variable, unit	Notation
2 m air temperature, K	$T_{2m}$
U - component of wind at 10 m, $m\ s^{-1}$	$u_{10m}$
V - component of wind at 10 m, $m\ s^{-1}$	$v_{10m}$
Temperature at the mixing layer height, K	$T_{MLD}$

## 2.2 METHODS

### 2.2.1 Observation-based temperature evolution model

Once all the datasets are organized and the variables needed are identified, the Observation-based Temperature Evolution Model (OTEM) that estimates one-hour temperature variations at the surface is developed (Rojas Muñoz et al., 2021). This model considers the surface variables mentioned in Chapter 1 that control temperature variations, in an hourly local scale. This approach used for this study is inspired from Bennartz et al. (2013) who implemented a temperature prognostic variation model to study the influence of low-level liquid clouds over the Arctic on the ice surface melt period in July 2012, by estimating all of these SEB terms for a case study. Here, we use the same approach for a different location (mid-latitude) and on a longer-time period

(5 years) to get robust statistics.



**Figure 2-7:** Hourly difference between  $T_{2m}$  and  $T_{2m,ERA5}$  for (a) SIRTA and (b) Météopole sites. The horizontal black line on each subfigure is the mean of this temperature difference and is equal to  $-0.55$  °C for SIRTA and  $-0.17$  For the Météopole site.

The temporal variability of local temperature can be expressed as the sum of four surface components, each characterizing a source of this variability:

$$\frac{\partial T_{2m}}{\partial t} = R + HG + HA + Adv \quad (2.2)$$

Where  $R$  is defined as the radiation term,  $HG$  as ground heat exchange term,  $HA$  as atmospheric heat exchange term and  $Adv$  as advection term.

Each of these terms represents a process involved in the variation of surface temperature  $\frac{\partial T_{2m}}{\partial t}$ , and can be calculated, using different measured meteorological variables as follows:

$$\frac{\partial T_{2m}}{\partial t} = \frac{\alpha + 1}{\rho c_p MLD} \Delta F_{NET} + \frac{T_G - T_{2m}}{\tau_G} + \frac{T_{MLD} - T_{2m}}{\tau_a} + \left( u_{10} \frac{\partial T_{2m}}{\partial x} + v_{10} \frac{\partial T_{2m}}{\partial y} \right) \quad (2.3)$$

Where  $T_{2m}$  is the surface temperature,  $t$  is the time,  $\alpha$  is a coefficient characterizing the form of the temperature profile in the boundary layer,  $\rho$  is the average air density of

the boundary layer,  $c_p$  is the specific heat of air, MLD is the height of the boundary layer,  $\Delta F_{NET}$  is the net radiative flux at the surface,  $T_G$  is the temperature in ground at 20 cm depth,  $T_{MLD}$  is the temperature at the top of the boundary layer,  $u_{10}$  and  $v_{10}$  are the zonal and meridional wind components, respectively, at 10m above the ground, and  $\tau_G$  and  $\tau_a$  are defined as relaxation timescales for heat exchange processes in the ground and the atmosphere, respectively.

This 2.3. expression is further explained in Sect. 2.2.1.2.

### 2.2.1.1 Useful theory elements

#### a. Energy conservation for a fluid particle

The first law of thermodynamics starts by the principle that a system has not only macroscopic kinetic and potential energy, but also it contains internal energy due in fact to the kinetic and potential energy of its molecules. Any change or alteration of one of these energies will either increases temperature system or produces some change in the relative position of its molecules.

The first law of thermodynamics indicates that the change in the sum of the macroscopic kinetic energy  $K$  and the internal energy  $U$  of the particle per unit time in a system is equal to the sum of the total power of forces from the external medium  $\delta\dot{W}_{ext}$  applying to the particle and the amount of heat exchanged per unit time with the exterior  $\delta\dot{Q}$ :

$$\frac{D(K + U)}{Dt} = \delta\dot{W}_{ext} + \delta\dot{Q} \quad (2.4)$$

Then, to calculate the external power exerted on the system, first it is calculated the work of the pressure forces exerted on the surface elements perpendicular to the axis Ox of the fluid particle, as shown in Figure 2-8. In a fluid particle, the pressure exerted at A is  $p_A ds$  and at B is  $-p_B ds$ , where  $ds$  is a surface element perpendicular to Ox. Now, if point A moves by  $dl_A$  and point B by  $dl_B$  for a time  $dt$ , the work balance of the pressure forces can be expressed as:

$$W_{pressure} = p_A ds dl_A + (-p_B ds dl_B)$$

Then, by multiplying and dividing by  $dx$  it is found:

$$W_{pressure} = - \left( \frac{p_B dl_B - p_A dl_A}{dx} \right) dv$$

Where  $dv = ds dx$  is the elementary volume of the fluid particle. Next, the work of pressure force is divided by  $dt$  in order to obtain the power of this force:

$$\frac{W_{pressure}}{dt} = \dot{W}_{pressure} = - \left( \frac{p_B u_B - p_A u_A}{dx} \right) dv$$

With  $u = dl/dt$  defined as the component of the velocity with respect to the x axis in

point A and B.

For an element  $dx$  that tends to 0, it is found:

$$\dot{W}_{pressure} = -\frac{\partial(pu)}{\partial x} dv$$

By doing the same analysis for the other two reference axes, the following equation is obtained:

$$\dot{W}_{pressure} = -\left[\frac{\partial(pu)}{\partial x} + \frac{\partial(pv)}{\partial y} + \frac{\partial(pw)}{\partial z}\right] dv = -div(p\vec{u}) dv$$

Also, the total work balance of the forces acting on the surface elements perpendicular to the Ox axis can be estimated as the sum of two terms. The first term is associated with the displacement of the particle's center of gravity (Figure 2-8b), which can be calculated as:

$$\dot{W}_{pressure}^K = p_A ds dl_M + (-p_B ds dl_M) = -\frac{(p_B - p_A)}{dx} dv dl_M$$

Where  $dl_M$  is the displacement of the particle's center of gravity during  $dt$ . By doing as before, when  $dx$  approaches 0 and dividing by  $dt$ , the following expression is found:

$$\dot{W}_{pressure}^K = -\frac{\partial p}{\partial x} u dv$$

With  $u$  the component of the velocity parallel to the Ox axis of the center of gravity  $M$  of the fluid particle. For the other two axes, by applying the same exact procedure, the first term in the total work balance of the forces can be expressed as:

$$\dot{W}_{pressure}^K = -\left[\frac{\partial p}{\partial x} u + \frac{\partial p}{\partial y} v + \frac{\partial p}{\partial z} w\right] dv = -\overrightarrow{grad}(p) \cdot \vec{u} dv$$

This term is associated with the pressure exerted by the macroscopic kinetic energy  $K$ .

The second term is associated with volume variations of the particle without any displacement of its gravity center (Figure 2-8c). This term is equal to:

$$\dot{W}_{pressure}^U = p_A ds dl'_A + (-p_B ds dl'_B)$$

Where  $dl'_A = dl_A - dl_M$  et  $dl'_B = dl_B - dl_M$ . The expression above can be simplified by neglecting the second order terms when  $dx$  approaches to zero:

$$\dot{W}_{pressure}^K = -p_M \frac{\partial u}{\partial x} dv$$

By adding the contributions of the y and z axes, it is obtained:

$$\dot{W}_{pressure}^U = -p \left[\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} + \frac{\partial w}{\partial z}\right] dv = -p div(\vec{u}) dv = -p \frac{D(dv)}{Dt}$$

This second term refers to the evolution of internal energy.

Returning to Eq. 2.4, the sum of internal  $e_i$  and kinetic  $e_c$  energies per unit mass can be

calculated as follows:

$$\frac{D \left( \overbrace{\mathcal{M}e_i}^U + \overbrace{\mathcal{M}e_c}^K \right)}{Dt} = \delta\dot{W} + \delta\dot{Q} \quad (2.5)$$

Where  $\mathcal{M} = \rho dv$ . After having carried out the balance of external working forces  $\delta\dot{W}$  acting on a fluid particle, it is found:

$$\delta\dot{W} = -div(p\vec{u})dv + \mathcal{M}\vec{g} \cdot \vec{u} = -pdiv(\vec{u})dv - \overrightarrow{grad}(p) \cdot \vec{u}dv + \mathcal{M}\vec{g} \cdot \vec{u}$$

Knowing that the variation in macroscopic kinetic energy can be expressed as:

$$\frac{De_c}{Dt} = \vec{g} \cdot \vec{u} - \frac{1}{\rho} \overrightarrow{grad}(p) \cdot \vec{u} \quad (2.6)$$

Eq. (2.6) comes from the assumption of being in a hydrostatic atmosphere where the energy available can be decomposed in the gravitational and pressure energies, and where the vertical acceleration has been neglected.

$De_c$  can therefore be subtracted in Eq. 2.5 and dividing by  $\mathcal{M}$ , it is deduced the first principle of thermodynamics to obtain an evolution equation for the internal energy  $U$  of a fluid particle:

$$\frac{De_i}{Dt} = -\frac{p}{\rho} div(\vec{u}) + \delta\dot{Q} \quad (2.7)$$

Where the term  $\delta\dot{Q}$  is the heat exchanged per unit masse and per unit time with the exterior.

To continue the simplifications, for an ideal gas, the internal energy  $U$  of a fluid particle as a function of temperature can be expressed as:

$$U = \frac{3}{2} NkT$$

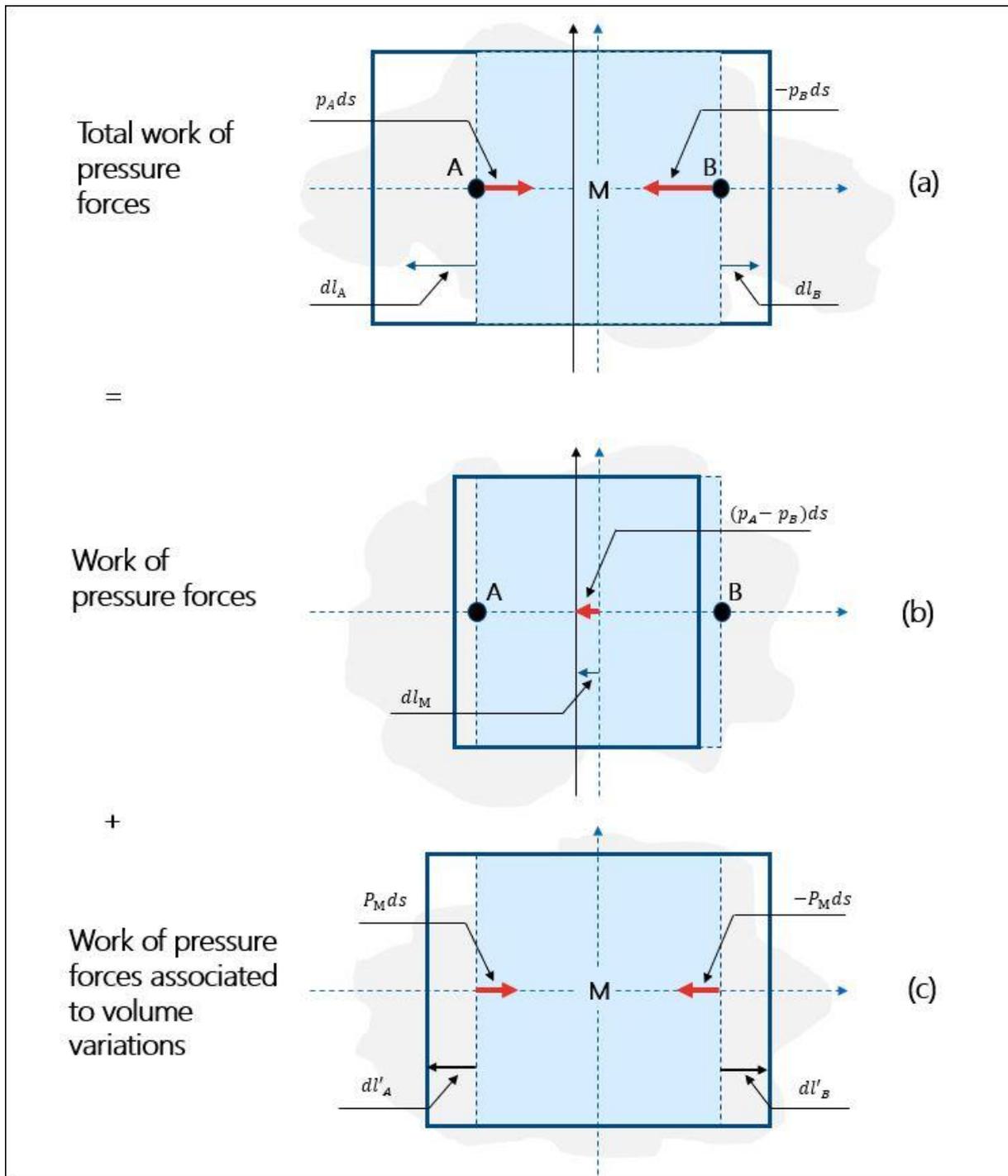
For a constant volume,  $C_v = 3/2Nk$  is defined as it is defined as the heat capacity at constant volume. Therefore, the internal energy  $U$  of a particle with masse  $\mathcal{M}$  and  $dv$  as its volume is equal to:

$$\frac{D(C_v T)}{Dt} = -pdiv(\vec{u})dv + \delta\dot{Q}$$

Then, dividing by  $\mathcal{M}$ , it is obtained:

$$\frac{D(c_v T)}{Dt} = -\frac{p}{\rho} div(\vec{u}) + \delta\dot{Q}$$

With  $C_v = \mathcal{M}c_v$ , where  $c_v$  is the specific heat at constant volume. This equation bears a strong resemblance to Eq. 2.7, the difference is that it was only done a replacement of  $e_i$  by  $c_v T$ .



**Figure 2-8:** Balance of the total work of the pressure forces exerted on the surface elements perpendicular to the axis  $Ox$  of a fluid particle. (a) The balance of the total work of the pressure forces breaks down into two terms: (b) the work of the balance of the pressure forces leading to the displacement of the center of gravity and (c) the balance of the work of the pressure forces resulting in a variation in volume of the particle without displacement of the center of gravity. Figure adapted from Malardel (2009).

By applying the law of conservation of mass, it is possible to eliminate the velocity in

order to obtain an equation based only on the density:

$$\text{div}(\vec{u}) = -\frac{1}{\rho} \frac{D\rho}{Dt}$$

Therefore, it is obtained:

$$\frac{D(c_v T)}{Dt} = \frac{p}{\rho^2} \frac{D\rho}{Dt} + \delta\dot{Q} = -p \frac{D\left(\frac{1}{\rho}\right)}{Dt} + \delta\dot{Q} \quad (2.8)$$

### b. Isobaric process

An isobaric transformation is one of the classical cases of thermodynamic evolution of a dry air particle (or a wet air particle without water phase change), and refers to a transformation during which the pressure does not vary. Note that for convective systems this approximation is no longer valid because we are not anymore in a hydrostatic system and vertical motions have to be considered, which mostly occur during storms. The assumption of taking the density of air to be approximately constant over the lower atmosphere allow us to assume its pressure also remains approximately constant. Thus, by taking up the first principle of thermodynamics (Eq. 2.8) and knowing that for an ideal gas  $\alpha_T = RT/p$ , another way to express this first principle is:

$$\frac{D(c_v T)}{Dt} = \frac{1}{\rho} \frac{Dp}{Dt} - R \frac{DT}{Dt} + \delta\dot{Q}$$

To simplify this expression, the  $\frac{DT}{Dt}$  are grouping in one side, which gives:

$$\frac{D(c_p T)}{Dt} = \frac{1}{\rho} \frac{Dp}{Dt} + \delta\dot{Q}$$

Where  $c_p = c_v + R$  is the mass heat at constant pressure. As the hypothesis of working on a constant pressure system is established, thus  $\frac{Dp}{Dt} = 0$  and then:

$$\frac{D(c_p T)}{Dt} = \delta\dot{Q} \quad (2.9)$$

For an isobaric transformation,  $c_p$  describes well the capacity of the gas to vary its temperature according to the heat exchanged with the outside.

### c. Enthalpy and vertical temperature profile

The assumption of working with an ideal gas in an isobaric environment, and applying the second law of thermodynamics, yield to an expression of the enthalpy as described in the following equation (Malardel, 2009; Wallace and Hobbs, 2006):

$$h = c_p T$$

With  $c_p$  the specific heat capacity. Now, the focus goes to study the enthalpy at a given

level. It is then defined as  $h(z) = c_p T(z)$  and if the assumption of having an air density to be constant in the surface layer, the total enthalpy in this layer from  $z=0$  to  $z=H$  can be written as:

$$h_{tot} = \int_0^H \rho c_p T(z) dz \quad (2.10)$$

To get the  $h_{tot}$  variations, it is then necessary to have the vertical temperature profile. In addition, the differentiation between day and night must be done because in the first case on average, an unstable situation is developed where the ground is warmer than the air flowing above, which causes increasing of turbulence and vertical energy transport, so the temperature profile has a negative gradient as going to upper levels from the ground. On the other hand, at night, the absence of solar radiation causes the ground to be colder than the air, and hence a fairly marked temperature inversion is developed, and a stable state is on average developed.

For the daytime analysis of temperature profile, the radiosoundings launched at Trappes Meteo-France station (48.77° N, 2.01° E) at 11:00 LT are now taken. Figure 2-9a shows a negative vertical temperature gradient at 11:00 LT for all the days of July 2011, which is also noticed in Figure 2-9c where each mean vertical temperature profile is estimated for each month of the year during the period of radiosoundings availability (from 2003 to 2017). These two figures suggest to work for daytime with a vertical temperature profile with a negative gradient, that is calculated as follows:

$$T(z) = T_{2m} + \beta * z \quad (2.11)$$

With  $\beta < 0$  and defined as the temperature gradient (i.e. the slope). At low layers, this temperature gradient is stronger than at higher altitudes. This strong low-level gradient is driven by intense solar heating of the surface, which heats up faster, especially during periods of strong sunshine (absence of clouds) and thus no precipitations at all (resulting in dry soils). This strong decrease rate is known as the super-adiabatic rate where the temperature gradient can reach up to  $-10$  °C/km (see Eq. 18.1 in Weiner and Matthews, 2003).

Figure 2-9b and Figure 2-9d are the same as Figure 2-9a and Figure 2-9c but for the radiosoundings launched at 23:00 LT. We noticed that for both cases a temperature inversion occurs at low layers (from the ground to  $\sim 300$  m) for almost every day of July 2011 and on average for all the months of the year (Figure 2-9b and Figure 2-9, respectively) when the air aloft the surface is warmer and the surface cools due to radiative cooling. Indeed, for a night stable case, Stull (1988) estimates the vertical temperature profile by a polynomial function having a zero gradient at the top of the ABL:

$$T(z) = T_{MLD} - \left(1 - \frac{z}{H}\right)^\alpha (T_{MLD} - T_{2m}) \quad (2.12)$$

With  $\alpha$  being a parameter that characterizes the shape of the temperature profile. A

value of 1 corresponds to a linear profile, while higher values denote a profile with larger gradients near the surface.

Figure 2-10 shows the averaged vertical temperature profile retrieved from the radiosoundings launched at 23:00 LT (blue line) and the vertical one calculated with Eq. 2.12 at 01:00 UTC (orange line), with  $H = 350$  m (see Section 2.2.1.2.) and  $\alpha = 1.5$ . Note that the appearance of the two temperature profiles is very similar (the difference between the two profiles is  $> 1$  °C) and thus the vertical temperature profile estimated by Eq. 2.12 can be used in our study, with a shape parameter equal to 1.5. Furthermore, it has been shown that this value of  $\alpha$  gives the best statistics between OTEM and the real observations of temperature variations (Section 3.1.).

The total enthalpy during daytime is calculated by replacing in Eq. 2.10 the temperature profile  $T(z)$  in Eq. 2.11:

$$h_{tot} = \rho c_p H \left[ \frac{\beta H}{2} + T_{2m} \right] \quad (2.13)$$

The equation above can be evaluated to estimate the change in surface temperature per unit change in enthalpy:

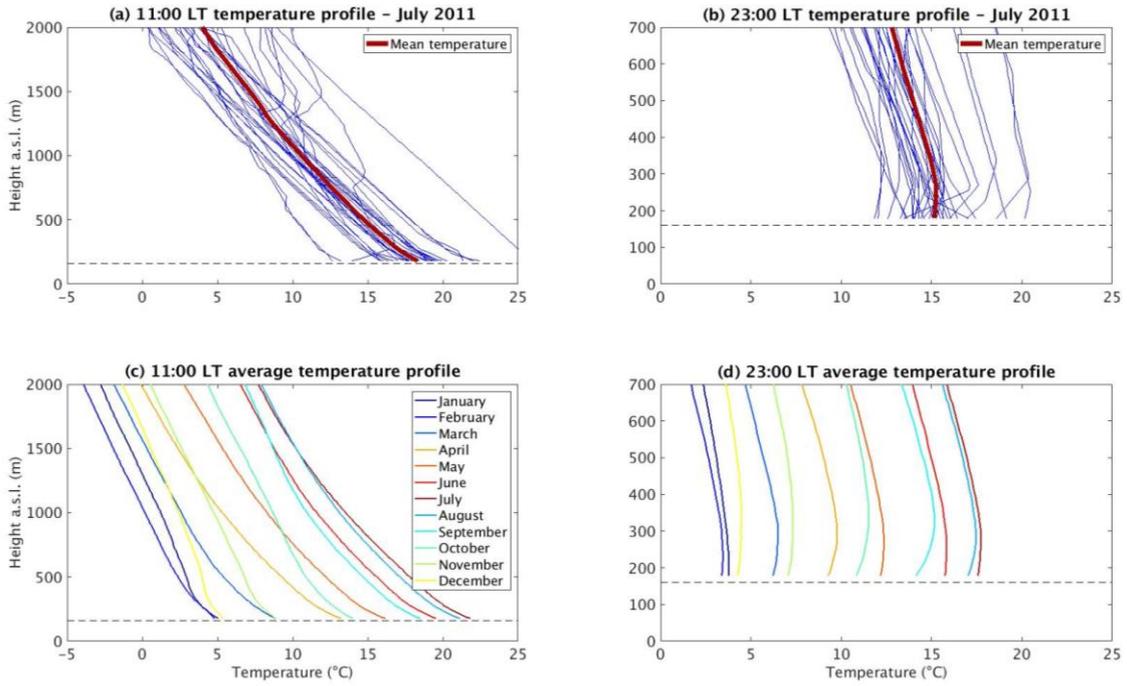
$$\frac{\partial T_{2m}}{\partial h_{tot}} = \frac{1}{\rho c_p H} \quad (2.14)$$

Proceeding as in the daytime case, the total enthalpy during nighttime knowing  $T(z)$  (Eq. 2.12) is equal to:

$$h_{tot} = \rho c_p H \left[ \frac{\alpha}{\alpha + 1} T_a + \frac{1}{\alpha + 1} T_{2m} \right] \quad (2.15)$$

For a given shape parameter, the change of surface temperature per unit change in enthalpy is estimated as:

$$\frac{\partial T_{2m}}{\partial h_{tot}} = \frac{\alpha + 1}{\rho c_p H} \quad (2.16)$$



**Figure 2-9:** Single non-averaged temperature profiles at (a) 11:00 and (b) 23:00 LT for a month of July of 2011. Each blue line in (a) and (b) corresponds to one day of the month whereas the thick dark red line is the average vertical temperature profile. Monthly averaged temperature profiles from 2003 to 2017 at (c) 11:00 LT and (d) 23:00 LT in Trappes. The horizontal dashed line represents the altitude above sea level where the Trappes observatory is located, which is 160 m. (a) and (c) have not the same vertical scale as (b) and (d).

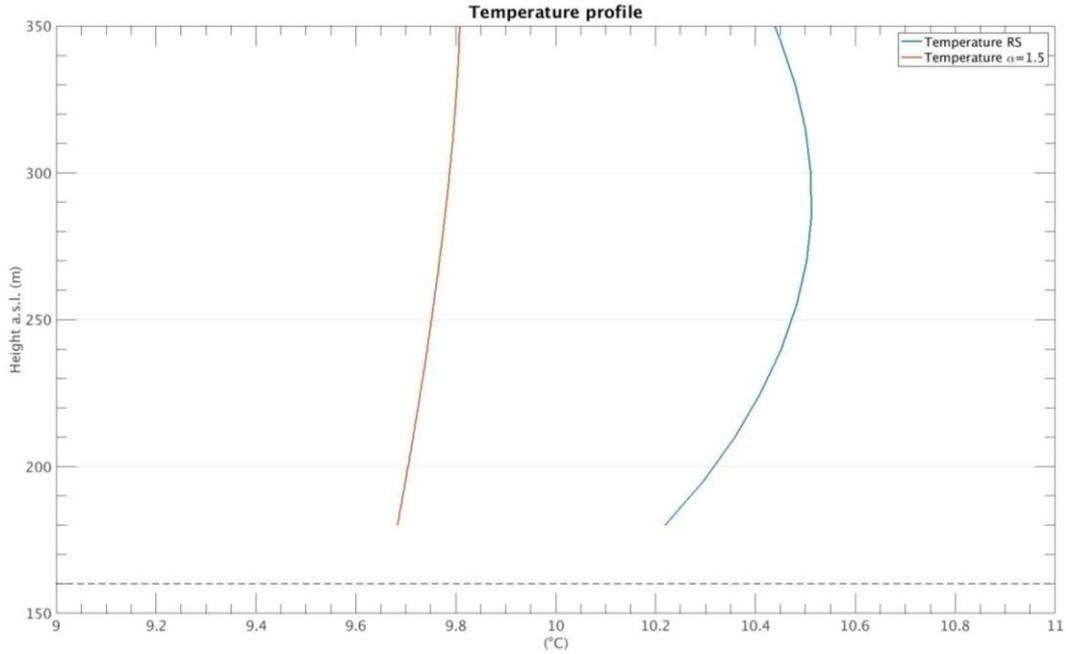
Note that the only difference between Eq. 2.14 and 2.16 is the shape parameter  $\alpha$  that for a daytime case is zero (i.e. no shape parameter) and during nighttime is equal 1.5.

### 2.2.1.2 Estimation of each OTEM term: SIRTA

#### a. Radiative term

The processes captured in Eq. 2.14 and 2.16 assume that the net radiative flux causes a forcing on the temperature profile such that the shape parameter remain constant and temperature is the only variable that responds to this radiative forcing. By taking Eq. 2.9 for an isobaric transformation and no phase change, an expression for the total enthalpy change due to the divergence of the radiative flux at the surface can be derived as follows:

$$\frac{\partial h_{tot}}{\partial t} = (F^\downarrow - F^\uparrow) \quad (2.17)$$



**Figure 2-10:** Averaged vertical temperature profile retrieved from the radiosoundings launched at 23:00 LT at Trappes (blue line) and averaged vertical temperature profile estimated with Eq. 2.12 (orange line) at 01:00 UTC. The averaged profiles are estimated from January 2009 to February 2014.

Replacing  $\partial h_{tot}$  in Eq. 2.14 and 2.16, for day and night, it is found respectively:

$$\frac{\partial T_{2m}}{\partial t} = \frac{1}{\rho c_p H} (F^\downarrow - F^\uparrow) \quad (2.18)$$

$$\frac{\partial T_{2m}}{\partial t} = \frac{\alpha + 1}{\rho c_p H} (F^\downarrow - F^\uparrow) \quad (2.19)$$

Finally, the divergence of the radiative fluxes at the surface hereafter is denoted as  $\Delta F_{NET}$  since it is the net radiative flux that reaches the surface. This divergence is simply defined as the differential of radiative flux entering to the surface compared to that leaving it. The parameter H is a scale of height corresponding to the vertical area where this surfaces processes occur. This scale is set as the MLD, a variable available in the SIRTa-ReOBS dataset (cf. Table 2-1). Therefore, the radiative term is calculated as:

$$R = \frac{\alpha + 1}{\rho c_p MLD} \Delta F_{NET} \quad (2.20)$$

The scale of height in Eq. 2.20 is taken as MLD because all the turbulent processes affecting the temperature variations are found within this layer. The ABL thickness varies depending, among other processes, on the amount of solar radiation, which enhances thermal and turbulent processes, making this thickness ranges from tens of meters to 2 km or more in the mid-latitudes. Thus, an average MLD thickness value in

an hourly and monthly scale is set during daytime. Since during nighttime the boundary layer depth is not well defined because of the weak turbulence and its very complex dynamical system (Shi et al., 2005; Walters et al., 2007; McNider et al., 2012), the average value of MLD at this time for each season is used and is presented in Table 2-3. These values also give the best statistics between the two datasets when a statistical evaluation is performed (see Section 3.1.).

**Table 2-3:** Nighttime averaged values of MLD for each season of the year.

Season	Averaged MLD (m)
Winter	314
Spring	368
Summer	417
Fall	334

The net radiative flux at the surface ( $\Delta F_{NET}$  in Eq. 2.20) is simply calculated as the difference between the radiative LW and SW fluxes leaving and arriving at the surface, which has been already defined in Eq. 1.2 (Chapter 1, Section 1.3.1.):

$$\Delta F_{NET} = F_{SW}^{\downarrow} - F_{SW}^{\uparrow} + F_{LW}^{\downarrow} - F_{LW}^{\uparrow} \quad (2.21)$$

Note that the only difference between Eq. 1.2 and Eq. 2.21 is that in the first the net radiative flux is called  $F_R$  whereas in this chapter is named as  $\Delta F_{NET}$ . It is possible to add and subtract the clear-sky (CS) downwelling radiation fluxes in Eq. 2.21 to obtain the flux radiative components in CS and cloudy (CL) conditions (i.e. the specific contribution of clouds), as follows:

$$\begin{aligned} \Delta F_{NET} &= F_{SW}^{\downarrow} - F_{SW}^{\uparrow} + F_{LW}^{\downarrow} - F_{LW}^{\uparrow} + F_{SW,CS}^{\downarrow} - F_{SW,CS}^{\downarrow} + F_{LW,CS}^{\downarrow} - F_{LW,CS}^{\downarrow} \\ \Delta F_{NET} &= \underbrace{F_{SW}^{\downarrow} - F_{SW,CS}^{\downarrow} + F_{LW}^{\downarrow} - F_{LW,CS}^{\downarrow}}_{CRE} + \underbrace{F_{SW,CS}^{\downarrow} - F_{SW}^{\uparrow} + F_{LW,CS}^{\downarrow} - F_{LW}^{\uparrow}}_{\Delta F_{NET,CS}} \end{aligned} \quad (2.22)$$

The first part of the right side in (2.22) is usually defined as the Cloud Radiative Effect (CRE).

The second part of the right side in (2.22) will be defined as  $\Delta F_{NET,CS}$ : admittedly, the upward radiative flux, whether SW or LW, depends indirectly on whether or not there is a cloud, but if we reason for a given downward radiative flux, the upward flux will not depend on whether there is a cloud or not posteriori. Nevertheless, in an annual global mean,  $F_{LW}^{\uparrow} > F_{LW,CS}^{\uparrow}$  (around  $0.5 \text{ W/m}^2$ ; Allan, 2011) due to the increase on longwave

radiation emitted to the surface by clouds, where a small proportion of this radiation is reflected back by the surface. As for the  $F_{SW}^\uparrow$ , it depends - relatively - only on the ability of the surface to reflect the solar radiation (surface albedo), and so this upward radiative flux is assumed to not vary if there are clouds or not.

This said, CRE and  $\Delta F_{NET,CS}$  are defined as:

$$CRE = F_{SW}^\downarrow - F_{SW,CS}^\downarrow + F_{LW}^\downarrow - F_{LW,CS}^\downarrow \quad (2.23)$$

$$\Delta F_{NET,CS} = F_{SW,CS}^\downarrow - F_{SW}^\uparrow + F_{LW,CS}^\downarrow - F_{LW}^\uparrow \quad (2.24)$$

With these two expressions, the radiative term in Eq. 2.20 can be separated in a radiative cloud ( $R_{CL}$ ) and clear-sky ( $R_{CS}$ ) terms, as follows:

$$R_{CL} = \frac{\alpha + 1}{\rho c_p MLD} CRE \quad (2.25)$$

$$R_{CS} = \frac{\alpha + 1}{\rho c_p MLD} \Delta F_{NET,CS} \quad (2.26)$$

Knowing that for daytime  $\alpha = 0$  and for nighttime  $\alpha = 1.5$ , and assuming  $\rho = 1 \text{ kg/m}^3$  and  $c_p = 1006 \text{ J/kg/K}$ , Figure 2-11b and Figure 2-11c present the behavior of these two terms.  $R_{CS}$  is able to warm (cool) the surface with a maximum (minimum) rate of  $4 \text{ }^\circ\text{C/h}$  ( $-1.5 \text{ }^\circ\text{C/h}$ ) where as  $R_{CL}$  has a more important cooling effect with values descending up to  $-3.7 \text{ }^\circ\text{C/h}$ , and warming the surface only with a maximum value of  $1.2 \text{ }^\circ\text{C/h}$ . Note that Eq. 2.25 and 2.26 capture well the seasonal (and expected) cycle for these two terms.

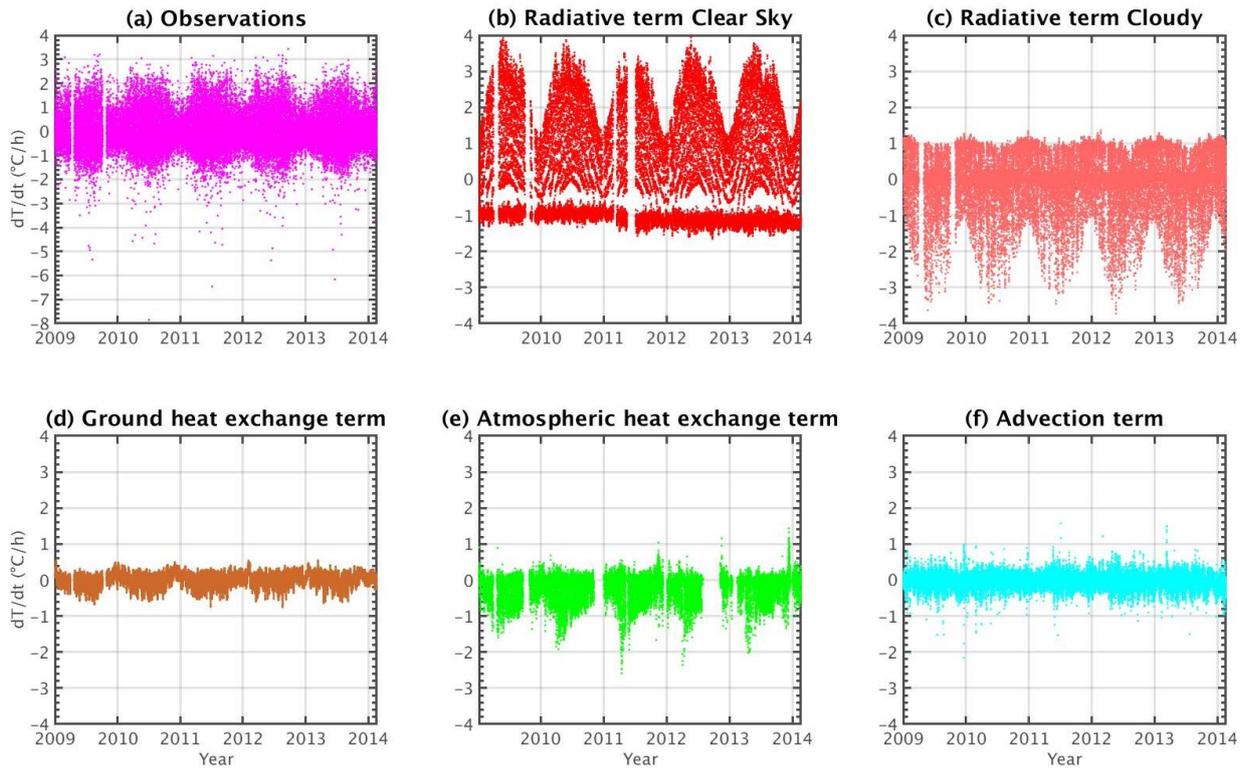
#### a. Atmospheric heat exchange

This term is estimated as:

$$HA = \frac{T_{MLD} - T_{2m}}{\tau_a} \quad (2.27)$$

In this equation,  $\tau_a$  can be calculated at equilibrium by assuming that the observed hourly temperature variations are zero and neglecting the ground heat exchange and advection terms (assuming their contributions are very minimal) in Eq. 2.3. This approach makes it possible to establish a balance between the radiative and atmospheric heat exchange terms so that the temperature does not vary, which yields to an estimation of  $\tau_a$  as follows:

$$\tau_a = \frac{\rho c_p MLD}{(\alpha + 1)} * \frac{(T_{MLD} - T_{2m})}{F_a} \quad (2.28)$$

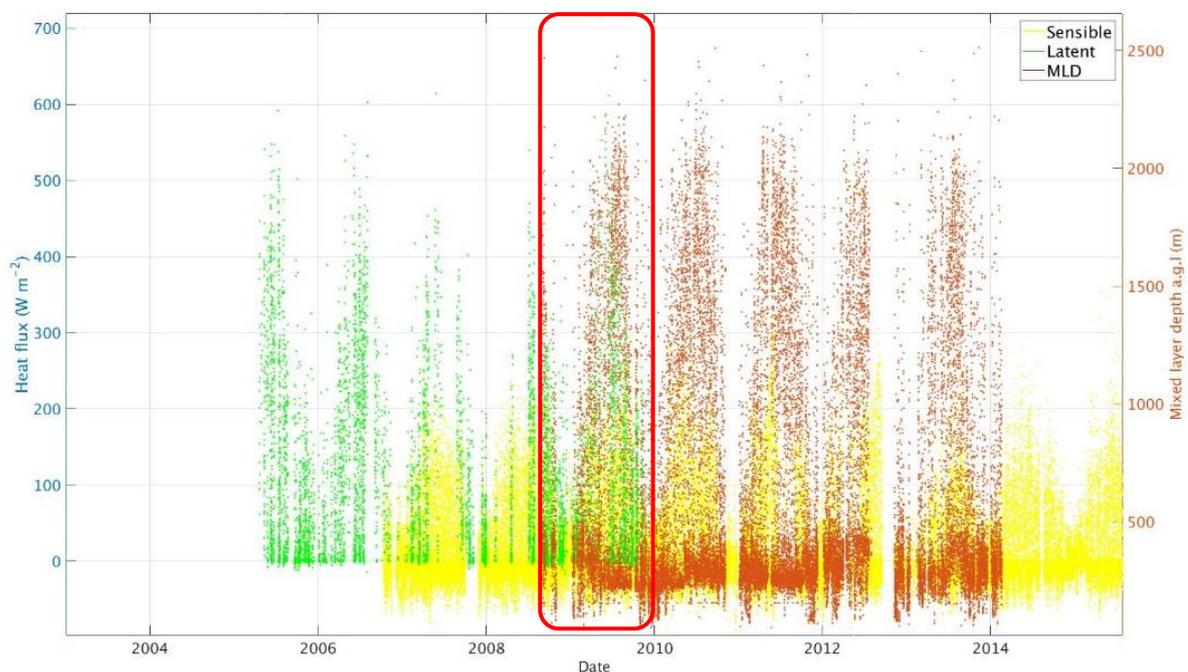


**Figure 2-11:** Hourly evolution of (a) observations, (b) the radiative clear-sky and (c) cloudy terms, (d) the ground heat exchange term, (e) atmospheric heat exchange term and (f) the Advection term.

For the atmospheric heat exchange,  $F_a = F_{lat} + F_{sens}$  (cf. Section 1.2.2.) since those are the fluxes controlling the heat exchange within lower atmosphere. The turbulence heat fluxes at the surface along with the MLD variable are necessary to estimate this relaxation time. However,  $F_{lat}$  has lots of missing values in SIRTAREOBS (material defection), and don't allow performing a complete analysis for the five years of study. For the Météopole dataset, this latent heat flux is more complete (Table 2-1), but the following method is also applied to the two sites since it gives better estimations of OTEM. As for the SIRTAREOBS dataset, Figure 2-12 presents the hourly data of  $F_{lat}$  and  $F_{sens}$  along with the MLD variable from 2003 to 2017.  $F_{lat}$  is available from mid-2005 to 2010 but still with a low percentage of availability (22%),  $F_{sens}$  starts at the end of 2006, and MLD measurements are available since the end of 2008. Thus, only one year and a couple of months are truly and commonly accessible for these three variables together (red box in Figure 2-12). This is the main reason why the study starts in January 2009. To handle the issue of availability of MLD,  $F_{lat}$  and  $F_{sens}$ , an hourly and monthly look up table (LUT) is created to have variable estimation of  $\tau_a$  calculated using Eq. 2.28.

First, a new variable is here estimated to calculate  $\tau_a$ :  $T_{MLD}$ , defined as the temperature at the MLD level. As seen in Table 2-1 the SIRTAREOBS dataset does not contain a vertical temperature profile at the hourly scale. Then to get  $T_{MLD}$  the following process is applied: (i) the pressure at the MLD level comes from the radiosoundings available

in SIRTA-ReOBS (MLD detection from the strongest gradient along the vertical with a wavelet method, which corresponds to the position of maximum turbulence), the availability of two profiles per day being sufficient for the pressure; (ii)  $T_{MLD}$  comes from the vertical temperature profile in ERA5 – which its hourly averaged on pressure levels with a vertical resolution of ERA5 of 25hPa for these low altitudes – at the nearest grid box from SIRTA Observatory and at this pressure level. Uncertainty remains in  $T_{MLD}$  because the pressure level is only available twice a day (radiosoundings launched at 11:00 and 23:00 LT), but values follow a physical behavior consistent with the 2-m temperature (Figure 2-13a).

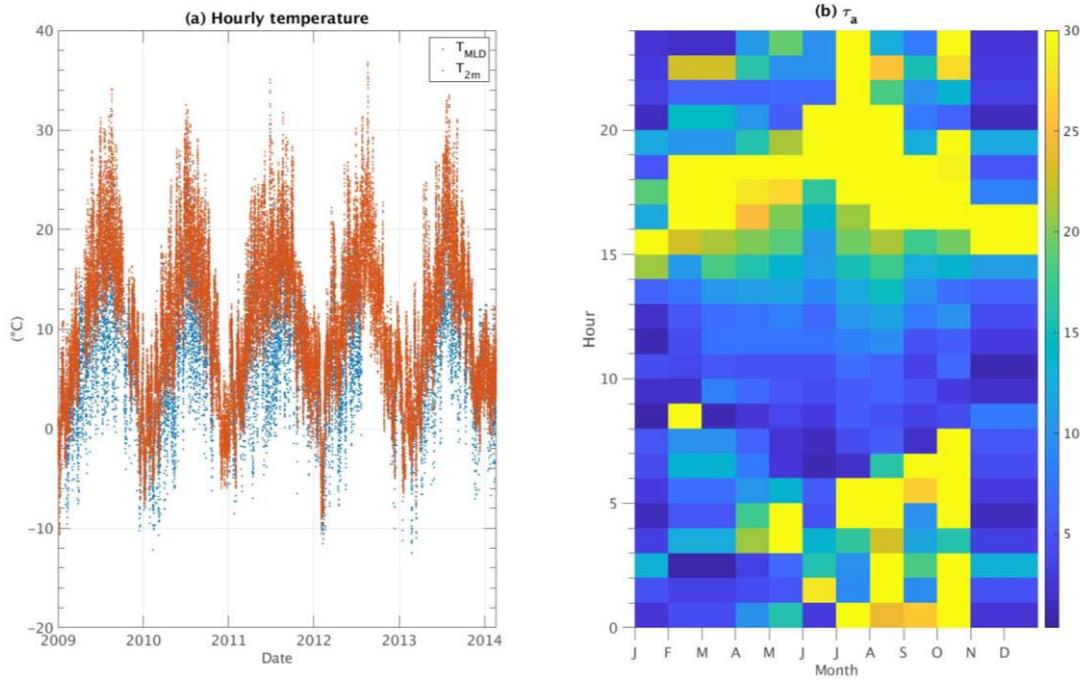


**Figure 2-12:** Hourly sensible (yellow dots), latent (green dots) heat fluxes at the surface and the mixed layer depth (brown dots) retrieved from the SIRTA-ReOBS dataset. The red square indicates the common period for the three variables (from November 2008 to December 2009).

All the variables required by Eq 2.28 are now available and then  $\tau_a$  can be estimated. Figure 2-13b presents a monthly-hourly average  $\tau_a$  for all the months of the year, these values constitute the LUT. During night, the relaxation time scale shows a higher month-to-month variability than during daytime mainly due to the gaps in the latent and sensible heat fluxes, whereas for daytime a more clearly behavior of  $\tau_a$  is spotted with low values during the morning that increase throughout the day but still remain lower than during nighttime. Sometimes during nighttime, this relaxation time exceeds 30 h, probably because the turbulence is so weak at that time that the heat flux does not reach completely the surface boundary layer (SBL). Stull (1988) suggested that for the night cases, values of  $\tau_a$  are on the order of 7 to 30 h, depending of the state of

the SBL. Then we force values to be limited to 30 h. For daytime,  $\tau_a$  remains smaller because the turbulence is stronger thanks to convection occurring near the surface, leading to a faster communication of surface information across the lower layers.

For some specific hour/month, the absence of the  $F_{lat}$  was very significant and thus it was not possible to get a value for  $\tau_a$ . It was the case for the month of December when none of the three variables in Figure 2-12 were available simultaneously. To handle this issue in the LUT, an averaged value between the previous and the following months for each hour is estimated.



**Figure 2-13:** (a) Hourly  $T_{MLD}$  (blue dots) and  $T_{2m}$  (orange dots) values from 2009 to 2014. The  $T_{MLD}$  variable is retrieved using both the SIRTA-ReOBS and ERA5 datasets. (b)  $\tau_a$  Monthly-hourly mean values. Units in the color bar are in hours.

The atmospheric heat exchange term variability is shown in Figure 2-11e. Most of the values are negative, indicating that this term will modulate the positive contribution made by the radiative forcing in clear-sky conditions.

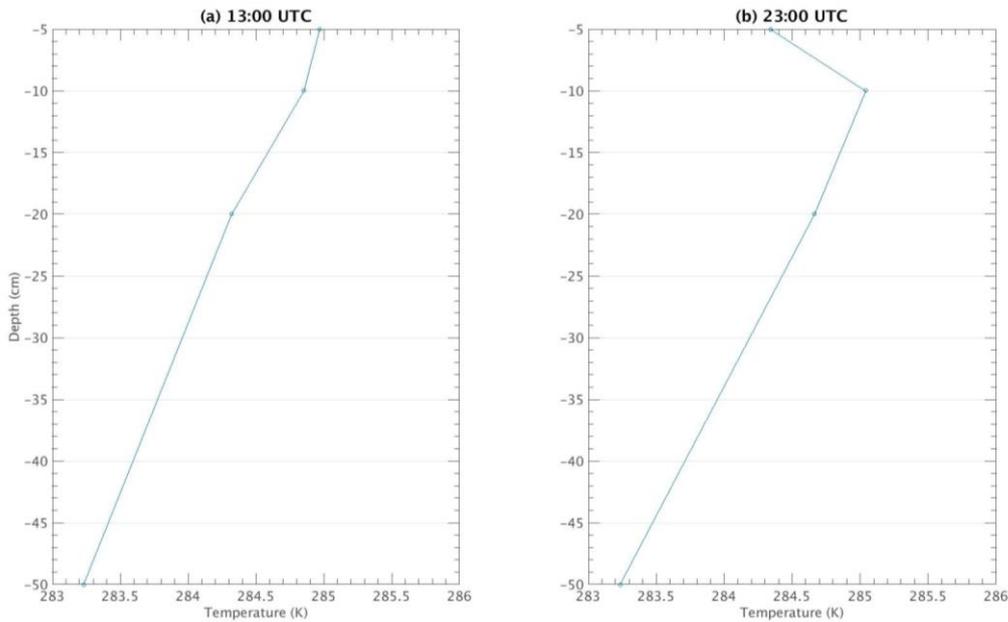
### b. Ground heat exchange

This term is estimated with the following equation:

$$HG = \frac{T_G - T_{2m}}{\tau_G} \quad (2.29)$$

On average, the temperature at the top of the ground during daytime (nighttime) is higher (lower) than the one at lower depth, as it's been shown previously (Al-Hinti et al., 2017; Popiel and Wojtkowiak, 2013). Hence, the assumption that the vertical profile of temperature within the ground at low depths follows approximately the same

behavior as the atmospheric temperature profile can be established, and therefore the approach adopted for the atmospheric heat exchange term can be here implemented, to evaluate the relaxation time scale for the ground heat exchange at a depth of  $H_s = 0.2$  m below the ground. It is considered the surface to be warmer than the temperature at 20 cm below the ground during the day thanks to the solar radiation arriving. The opposite happens during night, when the ground losses important longwave radiation and it gets cold faster than  $T_G$ . Figure 2-14 shows the vertical mean soil temperature profile for the four depths available in the SIRTA-ReOBS dataset (-5, -10, -20 and -50 cm) averaged at 13:00 UTC (a) and 23:00 UTC (b), and illustrates that indeed the ground behaves similar to the atmosphere in this aspect. The depth of 0.2 m is then chosen as the point where both atmosphere and ground vertical temperature profiles follow approximatively the same behavior.



**Figure 2-14:** Vertical mean soil temperature profile at SIRTA averaged for all the period of the study at (a) 13:00 UTC and (b) 23:00 UTC.

Knowing the temperatures at 2m above ground ( $T_{2m}$ ) and at 20cm below it ( $T_G$ ), the relaxation time scale can be calculated using a similar approach to the one used for  $HA$  (setting the left-hand side of Eq. 2.3. equal to zero and now neglecting  $HA$  and  $Adv$  terms), and results in:

$$\tau_G = \frac{\rho_G c_{p,G} H_G}{(\alpha + 1)} * \frac{(T_G - T_{2m})}{\Delta F_{NET}} \quad (2.30)$$

The equation to estimate  $\tau_G$  is very similar to the one to estimate  $\tau_a$  (Eq. 2.28), but since we are placed in the heat exchange at the ground, it is imperative to know its properties. The soil type at the SIRTA observatory is mostly a limestone type (obtainable from a regional predominant type of soils map in France added in Appendix A.2, Figure A-2 and extracted from Wulf et al. 2015). The approximatively

values for ground density  $\rho_G$  and heat capacity  $c_{p,G}$  for this type of soil are:

$$\rho_G = 1500 \text{ kg/m}^3$$

$$c_{p,G} = 830 \text{ J/kg/K}$$

A value of  $\tau_G = 19 \text{ h}$  in average for all the year is found for the day and night cases is found (with  $\alpha = 1.5$  during nighttime and  $\alpha = 0$  during daytime). Figure 2-11d presents the contribution of this term to temperature variations, which is very minor, with values oscillating between  $-0.8$  and  $0.5 \text{ }^\circ\text{C/h}$  with a more important negative contribution during summer than during winter. Despite its low contribution compared to the radiative and atmospheric heat exchange,  $HG$  remains important to have a better agreement between OTEM and the directly measured observations (cf. Chapter 3).

### c. Advection term

$$Adv = \left( u_{10} \frac{\partial T_{2m}}{\partial x} + v_{10} \frac{\partial T_{2m}}{\partial y} \right) \quad (2.31)$$

The advection term is estimated as presented in Eq. 2.31 using horizontal wind components at 10 m and temperature at 2 m above the ground from ERA5-Land Reanalysis. The purpose of using this dataset – and not the SIRTA-ReOBS file – is to have another point in each horizontal axis to estimate the transport of the air mass in zonal ( $\frac{\partial}{\partial x}$ ) and meridional ( $\frac{\partial}{\partial y}$ ) directions, between the SIRTA observatory and the immediately following ERA5-Land grid box. For the period and time scale considered, the advection term contributes to both warm and cool the surface with values going from  $-2.16$  up to  $1.57 \text{ }^\circ\text{C/h}$ , as shown in Figure 2-11f.

### d. Observed 2 m temperature variations term

The real values of the temperature variations are calculated as follows:

$$\frac{\partial T_{2m}}{\partial t} = \frac{T_h - T_{h-1}}{h} \quad (2.32)$$

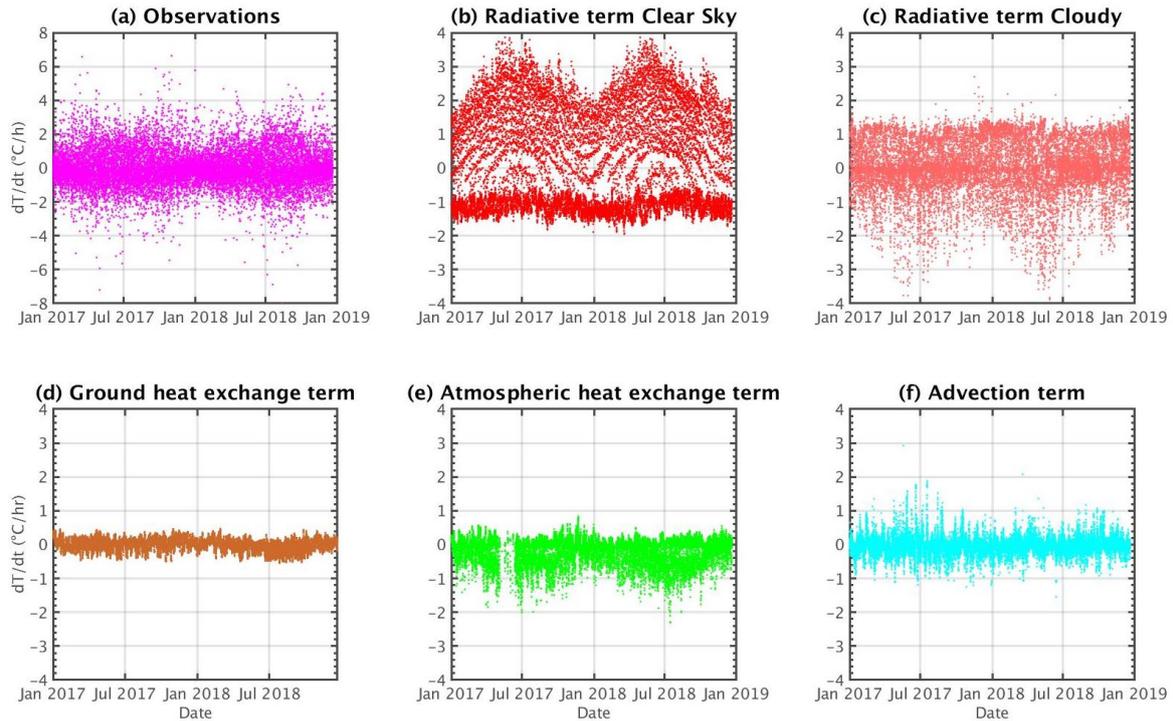
Where  $T_h$  is the temperature at 2m above the surface at hour  $h$  and  $T_{h-1}$  is the temperature at the previous considered hour.

Figure 2-11a presents the evolution of this term, denoting, as expected, a seasonal cycle. Temperature variations have stronger negative than positive values, reaching few times a rate of  $-6 \text{ }^\circ\text{C/h}$  or even  $-8 \text{ }^\circ\text{C/h}$  for some day in summer 2010.

The sum of all terms in the right-hand side of Eq. 2.3 is hereafter called  $\frac{\partial T_{2m}}{\partial t}_{OTEM}$ , and the real value of temperature variation is called  $\frac{\partial T_{2m}}{\partial t}_{obs}$ .

### 2.2.1.3 Estimation of each OTEM term: Toulouse

A similar procedure to estimate each SEB term done in the previous section in SIRTa is here performed but for the Météopole site in Toulouse. The period when all the data to study temperature variations is available goes from January 2017 to December 2018. Figure 2-15 shows the hourly contribution of the observed temperatures, each of the terms acting on SEB, as well as the advection term.



**Figure 2-15:** Same as Figure 2-11 but for the Météopole site in Toulouse.

All the terms shown in Figure 2-15 are estimated using the same equations with the same hypothesis on atmosphere stability for day and nighttime as for SIRTa. Nevertheless, due to the different location zone of the Météopole and weather conditions, the hourly contributions of each term vary and the following variables changed with respect to SIRTa:

- For the radiative terms, an average MLD thickness value in an hourly and monthly scale is set during daytime, whereas for nighttime, mean values for each season are used, and are presented in Table 2-4.
- For the ground heat exchange term, the area where the Météopole site is located presents a clay soil (cf. Appendix A.2, Figure A-2). The heat capacity of this type of soil is  $c_{p,G} = 900 \text{ J/kg/K}$ . The density of this soil is set to be  $\rho_G = 1280 \text{ kg/m}^3$  (both of these values are extracted from tables specifying these characteristics for different materials and types of soils). Since these values do not vary that much with those found for SIRTa, a  $\tau_G = 20 \text{ h}$  on average is found.

- For the atmospheric heat exchange term, sensible and latent turbulent heat fluxes are different than those at SIRTA (cf. Appendix A.3, Figure A-3), the latter observatory having lower values of sensible heat flux throughout the year (except in spring). Therefore,  $\tau_a$  values are different, but the same principle of creating a LUT as for SIRTA is applied in order to estimate HA. The mean monthly-hourly values of  $\tau_a$  are presented in Figure 2-16 with values close to 30 h during nighttime and at sunset hours, and with values close to 10 h during daytime. Furthermore, radiosoundings are not available for the Météopole, thus a direct measure of the vertical pressure profile is not retrieved (as it is done for SIRTA). The following equation is then used to estimate the approximative pressure value at the MLD ( $p_{MLD}$ ):

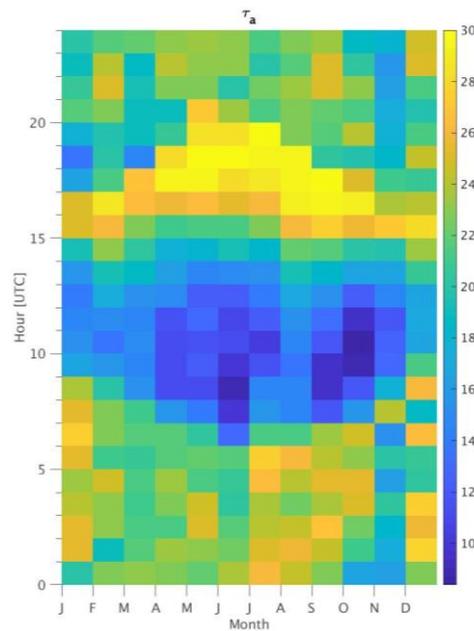
$$p_{MLD} = p_{2m} \cdot \left[ 1 + \frac{L_b}{T_{2m}} * MLD \right]^{\frac{-g * M}{R_0 * L_b}} \quad (2.33)$$

Where  $p_{2m}$  is the pressure above 2m of the ground (available in the Météopole dataset),  $L_b$  is the standard temperature lapse rate (-0.0065 K/m),  $g$  is the Earth-surface gravitational acceleration (9.81 m/s<sup>2</sup>),  $M$  is the molar mass of dry air (0.029kg/mol), and  $R_0$  is the universal gas constant (8.31 J/mol/K). Eq. 2.33 is generally used to estimate pressures for an altitude up to 11 km. The pressure at MLD (Appendix A.4, Figure A-4 for details on numeric values of this pressure) is necessary to estimate the temperature at MLD using once again the ERA5 reanalysis (as done for SIRTA for the HA term estimation).

**Table 2-4:** Same as Table 2-3 but for nighttime MLD mean values at the Météopole site.

Season	Averaged MLD (m)
Winter	354
Spring	286
Summer	261
Fall	273

A further analysis of the difference between the terms in SIRTA and the Météopole is carried out in Chapter 5 in order to evaluate how specific local and weather conditions affects their contribution and importance.



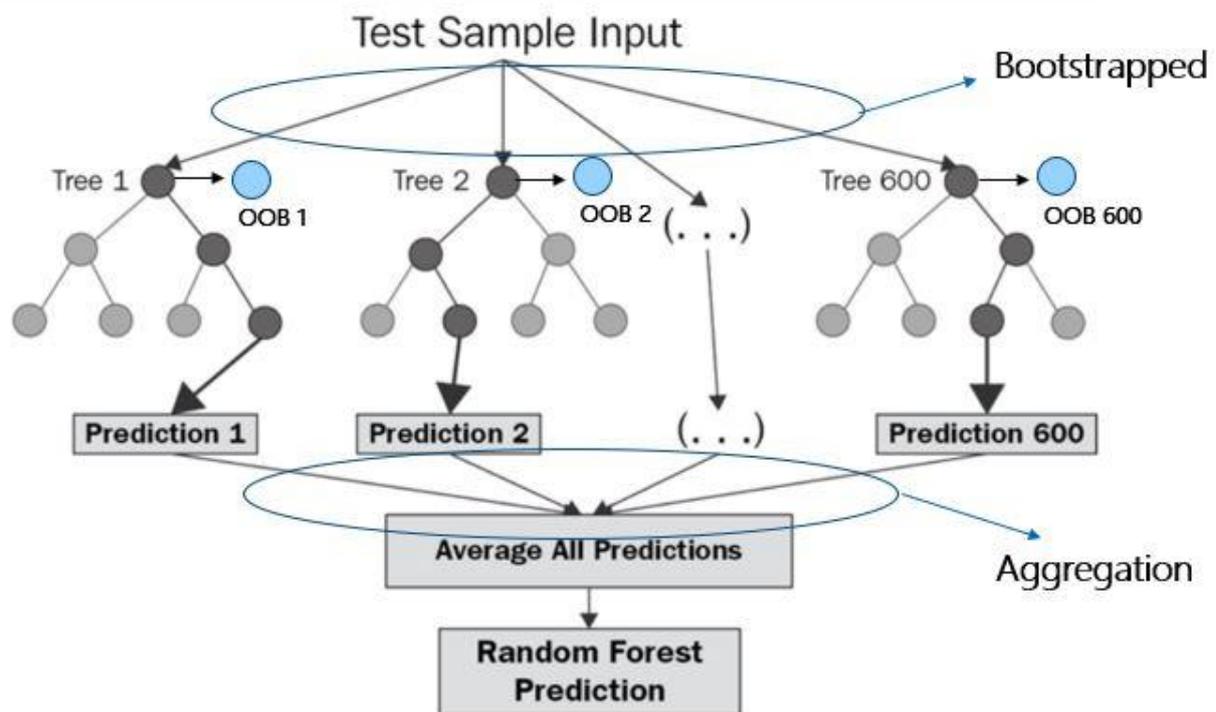
**Figure 2-16:** Same as Figure 2-13b but for the Météopole site.

## 2.2.2 The random forest method

Further on this study, new predicted values of surface temperature variations and the importance of each of the terms acting on it and previously presented in Section 2.2.2. are estimated based on a machine learning method: the random forest. This supervised machine learning method can be used for both classification and regression issues, which form the majority of current machine learning systems. Since OTEM is considered as a simple one-degree regression model which is built as the sum of five terms to estimate hourly surface temperature variations, the random forest method can be here implemented. Furthermore, each term evaluates and contributes differently throughout day and for each season of the year, and thus it is crucial to know the moments when one term will predominate over the others, a feature that can be estimated thanks of one of the random forest's attributes.

### 2.2.2.1 Description of the method: new estimation of 2 m temperature variations

This random forest method consists of bootstrapped-aggregated decision trees, a method that combines and gathers together the results of these trees to construct more powerful prediction models.



**Figure 2-17:** Basic random forest structure for a regression problem. This figure is extracted and adapted from <https://medium.com/swlh/random-forest-and-its-implementation-71824ced454f>.

Figure 2-17 illustrates a basic diagram on how this method works. Basically, the training algorithm for random forests applies the general technique of bagging, where the key to bagging is that trees are repeatedly fit to bootstrapped subsets of the observations. The bootstrapped term refers here to the fact of choosing randomly data that can be chosen several times to build decision trees (no selection restrictions). The random forest method works as follows:

1. It builds a number of decision trees on bootstrapped training samples from the test sample.
2. In each tree, an attribute selection is set: depending on the samples chosen (randomly) for training, the decision tree algorithm looks at all features and its values to determine which feature value would lead to the best node split. This is done by looking the mean squared error (MSE) which needs to be minimized.
3. Once the decision tree finds the best split point selection, it starts splitting the dataset at that value (called the root node). This principle of attribute selection continues to be applied at each split node.
4. The decision tree continues to split until either it's grown enough leaf nodes so that they reach each sample (i.e. there are no more data to continue the split), or a stopping criterion has been previously set (prune criterion). The last value found at the last node is the best estimation of the decision tree

5. The other trees apply the same algorithm to find their best estimations (prediction response) from the bootstrapped sample on which they were created, and then their estimations are averaged (i.e. aggregation) (Figure 2-17). All the decision trees are trained in parallel
6. The averaged result is the best random forest estimation/prediction

As previously explained in Section 2.2., hourly temperature variations at the surface are firstly estimated as the sum of five terms. Since the random forest method also allows to have predictions values from the trained input data (i.e. the five terms), it is used to determine new predictions of temperature variations values. These predicted values will be more accurate than those obtained from the simply prognostic model (cf. Eq. 2.2), and they are used to compare observed temperature variations for SIRTa in Chapter 3. This time, the random forest trains itself and the predictions are no longer the sum of the five terms, but a function that the method has found to give the best results.

The main advantages of this method are listed below:

- Low variance error thanks to bagging method (decorrelated trees). It has been shown that the method is not sensitive to small and medium changes in the input data
- Improves prediction accuracy
- **Presents estimates for variable importance (Out-of-Bag error estimation)**
- Among all the available classification methods, random forests provide the highest accuracy.
- Handle very easily big data with numerous variables
- Handle the missing values and maintains accuracy for missing data
- The method won't overfit the model
- Matlab toolbox is available

Despite all the advantages above, this method presents one big limitation: gaining a full understanding of the decision process for each decision tree is infeasible. Therefore, there is little control over what the model does. In fact, the random forest method is known as a black box for statistical modelers: no one knows what happens inside each tree but the prediction and the importance estimations obtained as outputs are strongly reliable.

### *2.2.2.2 Estimation of the relative importance of the different terms*

From the advantages listed above, the one in bold is the most interested for the current work as it will allow estimating the importance of each term (or predictor) acting on the temperature variation, in a nonparametric way. One of the main advantages of this method is that it allows covering not only the impact of each term individually in OTEM

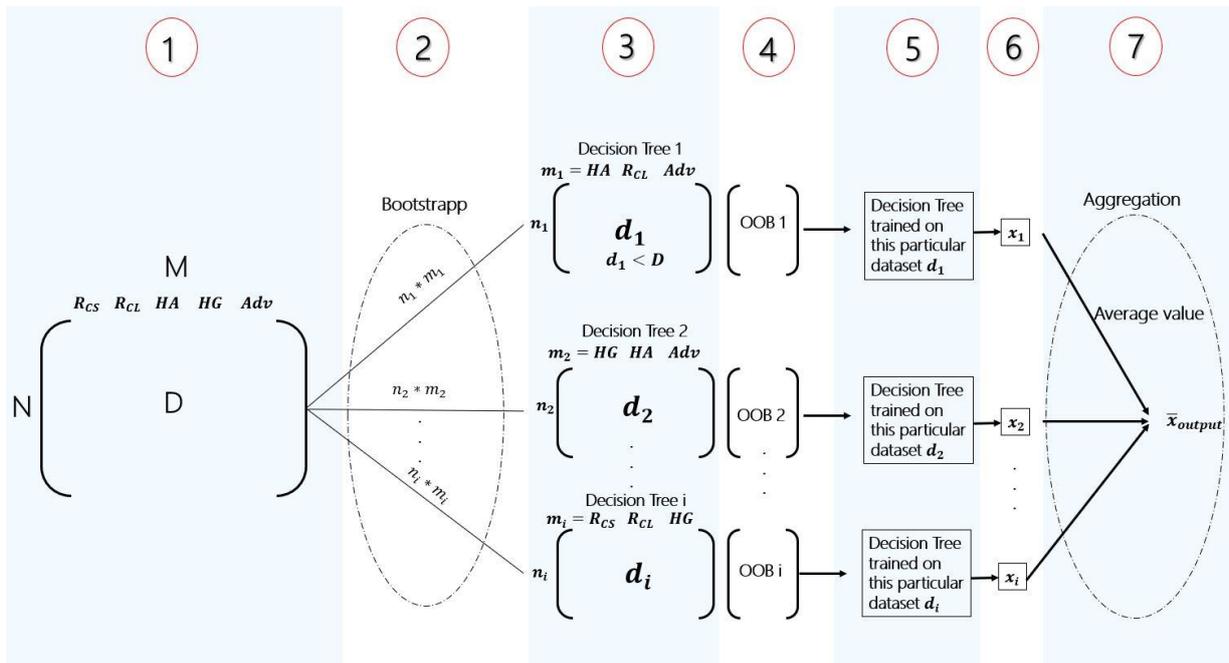
but also the multivariate interactions with other predictors. Indeed, other approaches to estimation of predictor importance (e.g. simple squared marginal correlations method, squared standardized coefficients) do not give reliable results when the problem involves correlated predictors (Grömping, 2007), which is the case here.

The method creates several **bootstrapped-aggregated** decision trees with row sampling and feature sampling from a specific dataset. Figure 2-18 shows a simple step-by-step diagram of the method. Hereby each step (numbers enclosed in circles at the top of the figure) is explained in detail:

1. A dataset called  $D$  is set, which in our case is our model (OTEM) for the period of interest. This dataset has  $N$  rows (i.e.  $N$  hours) and  $M$  features (i.e. the five terms of OTEM).
2. From the particular dataset  $D$ , we pick up randomly some sample of rows  $n_1$  and features  $m_1$  with *replacement*<sup>2</sup> to build a first decision tree (this process is called **bootstrapping**). Note that  $n_1 < N$  and  $m_1 < M$  ( $n \approx \frac{2}{3}N$  and  $m \approx \frac{2}{3}M$ ) (James et al., 2013). The same procedure is repeated to create a second decision tree with  $n_2$  samples and  $m_2$  features. This bootstrap sampling continues until the number of trees  $i$ .
3. A first decision tree is then created with  $n_1$  samples and  $m_1$  features, and it. In this case, consider that the features  $m_1$  (i.e. terms) are  $HA$ ,  $R_{CL}$  and  $Adv$  and then our model to estimate surface temperature variations will be the sum of these three features. A second decision tree is similarly created, then the process continues until the  $i$ th decision tree is created with  $n_i$  samples and  $m_i$  features.
4. Since for each decision tree, approximatively one third of the samples and features are not used to trained it, they are called *Out-Of-Bag* (OOB) data. They are later used to estimate the importance of a specific feature. For the first decision tree, a first  $OOB_1$  observations are not used to trained it.
5. The first decision tree is be trained on  $d_1$ , the second one on  $d_2$ , and so on until decision tree  $i$ th is trained on  $d_i$
6. A first prediction  $x_1$  is given by the first decision tree after trained, later a second prediction  $x_2$  is obtained from the decision tree number 2, and it continues until a last prediction  $x_i$  is given by the  $i$ th decision tree.
7. All the predictions from all the decision trees are averaged (which in random forest is called **aggregation**) to get a final prediction  $\bar{x}_{output}$ . For our case, the final output is improved prediction of  $\frac{\partial T_{2m}}{\partial t}$ .

---

<sup>2</sup> In random forest, replacement refers to the fact that some of the samples and features previously taken to build a decision tree may be repeated to build later other decision trees. The data is selected randomly but there is no restriction concerning data repeatability.



**Figure 2-18:** Simplified step-by-step explanatory diagram of the random forest method. This diagram is an example on how random forest method works, therefore it does not reproduce and takes exactly the same features shown e.g. in step 3. Refer to the text for further details on the syllabus.

To determine the importance estimation of a particular term, the OOB dataset introduced in the step 4 is further analyzed. As explained in Section 2.2.2.1., the MSE is used to calculate this importance. To further explain its calculation, we consider the first decision tree in Figure 2-18. This tree is built and trained with only three of the five terms of OTEM ( $HA$ ,  $R_{CL}$  and  $Adv$  for instance) and with some hours of the period of interest. The other two terms ( $R_{CS}$  and  $HG$ ) are left OOB. Because in this tree  $R_{CS}$  was not used for training, the random forest method takes this tree to estimate its importance. The following procedure is made by the random forest method:

1. The method adds this term with its hourly values to the first trained decision tree in order to find a new response. We called this new response  $\frac{\partial T_{2m}}{\partial t_1}^*$
2. Since this first tree is already trained, we know its predicted value before the values of  $R_{CS}$  were added ( $x_1$ ). Then, we estimate a first MSE as follows:
  - $MSE_1 = \frac{1}{n_1} \left( x_1 - \frac{\partial T_{2m}}{\partial t_1}^* \right)^2$
3. Then, the rows in  $R_{CS}$  (i.e. the hours) are permuted (i.e. swapped) and the random forest method calculate another prediction after this permutation, called  $\frac{\partial T_{2m}}{\partial t_1}^P$
4. A new  $MSE^P$  is estimated, but considering this time new predicted value and the trained one as follows:
  - $MSE_1^P = \frac{1}{n_1} \left( x_1 - \frac{\partial T_{2m}}{\partial t_1}^P \right)^2$

5. The difference between the two errors is estimated ( $MSE_1 - MSE_1^P$ )
6. Finally, the same procedure is repeated (1 to 4 step) for the other trees having excluded  $R_{CS}$  and then it adds up all these differences.

Based on this principle, the larger the error (the sum of all the differences), the more important the term. Indeed, this importance estimation feature from the random forest method has been previously used to calculate the variable importance in different datasets for many applied problems in sciences or health fields for both regression and classification studies (e.g. Archer et Kimes 2008; Genuer, Poggi, et Tuleau-Malot 2012; Strobl et al. 2007). The use of this method for the current study is detailed in Chapter 3, Sect. 3.3.1.

## CONCLUSION

A simple – but complete – model based almost exclusively on ground-based observations is here developed in order to study hourly temperature variations at the surface (OTEM). The hourly values for all the variables requested by OTEM are retrieved from the SIRTA-ReOBS file which is a powerful dataset allowing to perform a precise calculation and estimation of all the five terms acting on OTEM, with a wide range of available data to do a multi-year analysis. OTEM is simply estimated as the sum of the five terms. To analyze the spatial evolution of the terms and temperature variations, a second dataset from Toulouse is also used and OTEM is also there applied. After each term is estimated and its contribution is briefly presented, the random forest method is introduced, a machine learning method that offers improved predictions of hourly temperature variations, and gives estimates of the importance of each term that controls these variations.

The estimation of each term allows us in Chapter 3 to assess how well its sum follow the real observations, and to estimate their contribution and importance under different conditions and for different temporal scales.



### 3 CHAPTER 3: INFLUENCE AND CONTRIBUTION OF EACH OTEM TERM ON LOCAL 2 M TEMPERATURE VARIATIONS

---

#### INTRODUCTION

The main objective of this chapter is to evaluate the contribution of the terms (previously described in Chapter 2) acting on local short-term 2-m temperature variations under different conditions. Having a multi-year database allows to estimate the reliability of our model (see Sect. 2) by statistically comparing it with the observed temperature variation (Sect. 3.1.1). Specific cases that are not well captured by our model, corresponding to extreme temperature variations are explained in Sect. 3.1.2. The averaged contribution of each term at monthly-hourly and annual scale is also presented, in order to identify the conditions when one term contribute the most over the others, including the use of a powerful machine learning method also described in the previous chapter, presented in Sect. 3.2.

#### 3.1 EVALUATION OF THE MODEL

In this first section, a statistical evaluation of OTEM is initially performed, in order to evaluate how well it reproduces the hourly temperature variations (i.e. comparing  $\frac{\partial T_{2m}}{\partial t}_{OTEM}$  to  $\frac{\partial T_{2m}}{\partial t}_{obs}$ ). Since some assumptions and hypothesis are necessary (on atmosphere stability and ground behavior, and in data treatment due to gaps in the latent heat flux variable in the SIRTAReOBS dataset - cf. Chapter 2), it is important to assess the reliability of the results obtained. This evaluation allows later to study individually each term of OTEM and their linear sum under different situations and conditions, in order to have confidence on their contributions.

##### 3.1.1 Statistical evaluation of OTEM

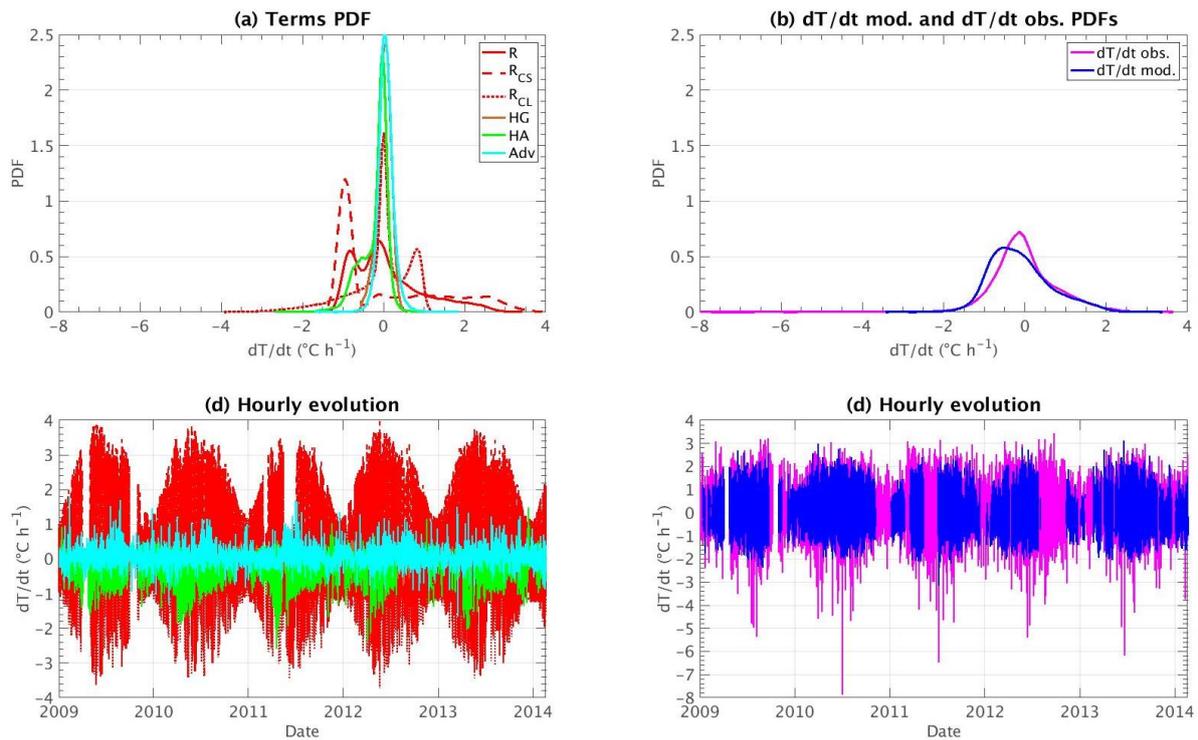
$$\frac{\partial T_{2m}}{\partial t} = \frac{\alpha + 1}{\rho c_p MLD} \Delta F_{NET} + \frac{T_G - T_{2m}}{\tau_G} + \frac{T_{MLD} - T_{2m}}{\tau_a} + \left( u_{10} \frac{\partial T_{2m}}{\partial x} + v_{10} \frac{\partial T_{2m}}{\partial y} \right) \quad (3.1)$$

Statistics are considered to assess how well OTEM (the sum of the four terms) reproduces the observations (i.e. comparing the observed temperature variations values of Eq. 3.1 to the OTEM ones). The probability density functions (PDF) of each of the terms of Eq. 3.1 are shown in Figure 3-1a, indicating that:

- The radiative term  $R$  (red line) has contribution values between  $-1.7$  and  $3.4$  °C/h, with one negative peak at  $-0.85$  °C/h, another at  $-0.20$  °C/h and a positive peak at  $0.85$  °C/h. This term can be separated onto:
  - o A radiative clear-sky term  $R_{CS}$  that presents a main contribution in positive values going up to  $4$  °C/h, despite an important negative peak at  $-1.0$  °C/h.
  - o A radiative cloud term  $R_{CL}$  that contributes to  $\frac{\partial T_{2m}}{\partial t}$  from  $-3.72$  to  $1.21$  °C/h with most of its values between  $-0.20$  and  $1.0$  °C/h.
- The atmospheric heat exchange term  $HA$  can decrease the temperature by  $-2.56$  °C/h and can increase it by  $1.16$  °C/h, and presents a peak around  $-0.60$  °C/h
- The ground heat exchange  $HG$  has a Gaussian distribution with its center at  $0$  °C/h with contribution values between  $-0.76$  and  $0.54$  °C/h
- The advection term  $Adv$  follows a Gaussian distribution centered at  $0$  °C/h and can lower the temperature by  $-1.64$  °C/h and increasing it by  $1.70$  °C/h.

Figure 3-1b presents the PDF of  $\frac{\partial T_{2m}}{\partial t}_{obs}$  (pink line) with temperature variations values ranging from  $-8.0$  °C/h to  $3.60$  °C/h but the high negative values represent very few points. The  $\frac{\partial T_{2m}}{\partial t}_{OTEM}$  PDF (blue line) has values between  $-3$  and  $3.40$  °C/h. When comparing the two PDFs, OTEM follows quite well the positive temperature variations, and low differences occur for cases when temperature decreases in the hour (i.e.  $\frac{\partial T_{2m}}{\partial t} < 0$  °C/h). In particular high negative values, even if they represent few points, are not reproduced by OTEM.

Figure 3-1c shows the hourly evolution of all the OTEM terms and Figure 3-1d the hourly evolution of the observed temperature variations and the sum of all the terms, for the period of study. If OTEM were perfect, the pink and blue lines would be superimposed. We can notice that all  $\frac{\partial T_{2m}}{\partial t}_{obs}$  high negative values that OTEM does not capture are in summer. This issue is addressed in Section 3.1.3.

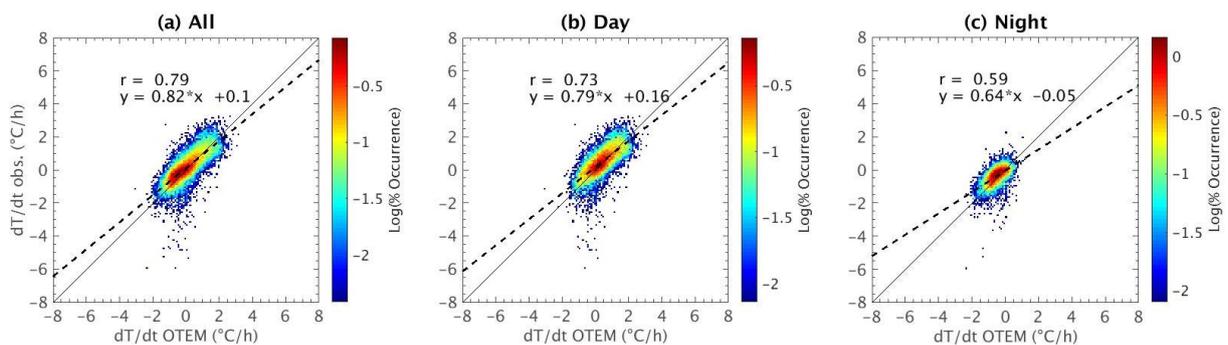


**Figure 3-1:** PDFs of (a) radiation (red line), radiative clear-sky (red dashed line), radiative cloud (red dotted line), heat ground (brown line) and heat atmospheric (green line) exchange, and advection (cyan line) terms, and (b)  $\frac{\partial T_{2m}}{\partial t}_{obs}$  (pink line) and the  $\frac{\partial T_{2m}}{\partial t}_{OTEM}$  (blue line). (c) Hourly evolution of all OTEM terms and of (d) observed temperature variations and OTEM for the period of study (same colors as in (a) and (b)).

To evaluate how well the temperature variations found by OTEM are reliable, we do a scatter plot of the observations versus OTEM for all the period of study, as shown in Figure 3-2a, along with its correlation coefficient ( $r$ ) and the best fit equation (in a least-square sense) between the two datasets. The correlation is good (0.79) and the bias remains small ( $-0.20$  °C/h, not shown), the latter corresponding just to 6% of the maximum positive value of  $\frac{\partial T_{2m}}{\partial t}_{obs}$ . Concerning the slope of the best fit linear regression (0.82) in Figure 3-2a, OTEM overestimates the observations, especially for negative values rather than positive ones. However, OTEM has difficulties in reproducing the negative extremes of temperature change within an hour, e.g.  $-6$  °C/h which corresponds to the 5<sup>th</sup> of July 2011 at 20:00 UTC, or  $-8$  °C/h which corresponds to the 2<sup>nd</sup> of July 2010 at 15:00 UTC, as presented in Figure 3-1c where the hourly evolution of all the terms,  $\frac{\partial T_{2m}}{\partial t}_{obs}$  and  $\frac{\partial T_{2m}}{\partial t}_{OTEM}$  are plotted for the period of study. There are other days when  $\frac{\partial T_{2m}}{\partial t}_{obs}$  presents an important negative value not reproduced by OTEM, mostly during summer. A further analysis of these extreme one-hour temperature gradients is carried out in Section 3.1.2.1.

The same scatter plots separated into day and nighttime cases are presented in Figure

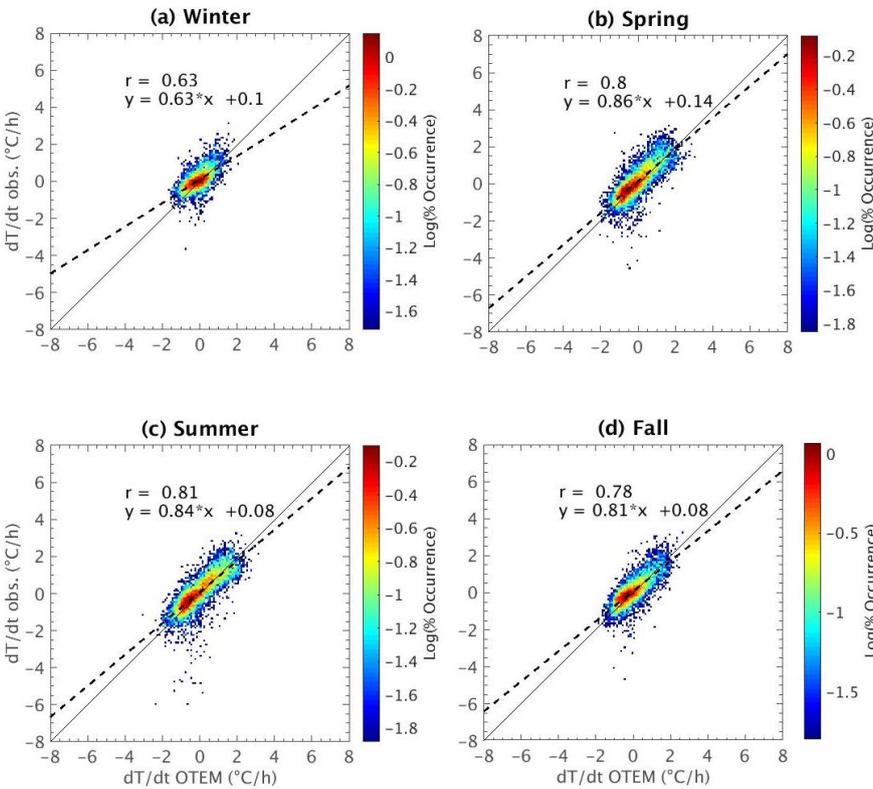
3-2b and Figure 3-2c, respectively (day and nighttime cases are identified based on the solar zenith angle variable, SZA). The correlation coefficient is much better for day (0.73) than nighttime (0.59), with lower values of observed temperature variations during night. These correlations are both lower than the one of the entire dataset because of the lower covariance values found between OTEM and observations. Indeed, for the entire dataset (i.e. Figure 3-2a) the covariance increases and the degree of association between OTEM and observations increases. Furthermore, the accuracy of OTEM during nighttime may have been affected by the hypothesis made concerning the MLD, since its value is not well defined during that time due to the weak surface turbulence and an averaged one is used for each season based on the radiosoundings launched at Trappes (cf. Section 2.2.1.2.), accuracy that could be improved using radiosoundings launched all the hours during nighttime. In addition, the temperature variations at night might be hard to quantify accurately due also to the minor contribution of each of the non-radiative terms, for which contributions are close to zero, especially for the  $HA$  term, where  $F_{lat}$  and  $F_{sens}$  are most of the time very low (Section 3.2.1.).



**Figure 3-2:** Scatter plot of  $\frac{\partial T_{2m}}{\partial t}_{obs}$  vs  $\frac{\partial T_{2m}}{\partial t}_{OTEM}$  for (a) all the hourly data, (b) only daytime cases and (c) only nighttime cases. The correlation coefficient “r” and the linear equation fitting the best for the two datasets are also shown and correspond to the diagonal dashed line, whereas the solid one represents the  $y=x$  equation.

Figure 3-3 shows the scatter plots as in Figure 3-2 but for all the four seasons of the year (day and night together). In addition, Table 3-1 presents some statistics calculated for each season. Correlation is high (0.81) and bias is low in summer. Nevertheless, the standard deviation of the difference between OTEM and observations remains high for this season (0.55 °C/h), mostly due to the higher values for both observed and modeled temperature variations as seen in Figure 3-3c in comparison to other seasons. For the spring and fall seasons, the correlation coefficient is still high (0.79 on average) and it is the lowest in winter (0.63) with a higher bias (-0.31 °C/h) but low standard deviation (0.46 °C/h). This low correlation is partly due to the fact that during winter nights are much longer than for the other seasons and thus the bias increases due to the hypothesis previously made that explain low correlation during night. Indeed, the correlation coefficient for night/Winter is 0.50. Separating into seasons when day and

nighttime cases are all simultaneously considered (second column in Table 3-1) yields to a better linear relationship (except in winter) than when all the dataset is considered (Figure 3-2a). This is related once again to the covariance which happen to be this time better when splitting into seasons (i.e. the strength of the linear relationship between these two datasets is high). Moreover, when we remove the extreme values of the observations and OTEM (i.e. considering the values that are within the 5th and 95th percentiles), statistics for all the seasons improves, increasing their correlation coefficient and decreasing both bias and standard deviation, especially for summer season (values in brackets in Table 3-1).



**Figure 3-3:** Same as Figure 3-2 but for (a) winter, (b) spring, (c) summer and (d) fall.

Finally, Table 3-2 shows the relative bias split into seasons for day and nighttime cases. For daytime cases only, this bias is more important for spring and winter whereas for summer and fall remains low. On the other hand, during nighttime, winter is the season with the highest values (related to the hypothesis done concerning the stability of the atmosphere), and summer presents the best statistic.

**Table 3-1:** Statistics values between  $\frac{\partial T_{2m}}{\partial t}_{obs}$  and  $\frac{\partial T_{2m}}{\partial t}_{OTEM}$  for the four seasons of the year. The values in brackets correspond to the statistics within the 5th and 95th percentiles.

Season	Correlation coefficient	Corr. coeff. Day	Corr. coeff. Night	Bias (°C/h)	Standard deviation (°C/h)
Winter	0.63 (0.68)	0.62 (0.63)	0.50 (0.51)	-0.31 (-0.23)	0.43 (0.40)
Spring	0.80 (0.81)	0.74 (0.75)	0.57 (0.59)	-0.15 (-0.11)	0.52 (0.45)
Summer	0.81 (0.84)	0.77 (0.80)	0.59 (0.61)	-0.18 (-0.10)	0.55 (0.46)
Fall	0.78 (0.80)	0.74 (0.76)	0.59 (0.60)	-0.21 (-0.10)	0.46 (0.41)

**Table 3-2:** Mean values of the observations (first line), and OTEM (second line), and the relative bias (in %, third line) split into day and nighttime for every season of the year.

	Day				Night			
	Winter	Spring	Summer	Fall	Winter	Spring	Summer	Fall
Obs (°C/h)	0,22	0,29	0,35	0,37	-0,35	-0,69	-0,74	-0,48
OTEM (°C/h)	0,29	0,39	0,42	0,43	-0,18	-0,52	-0,65	-0,34
Relatif bias (%)	31	34	20	16	-48	-24	-12	-29

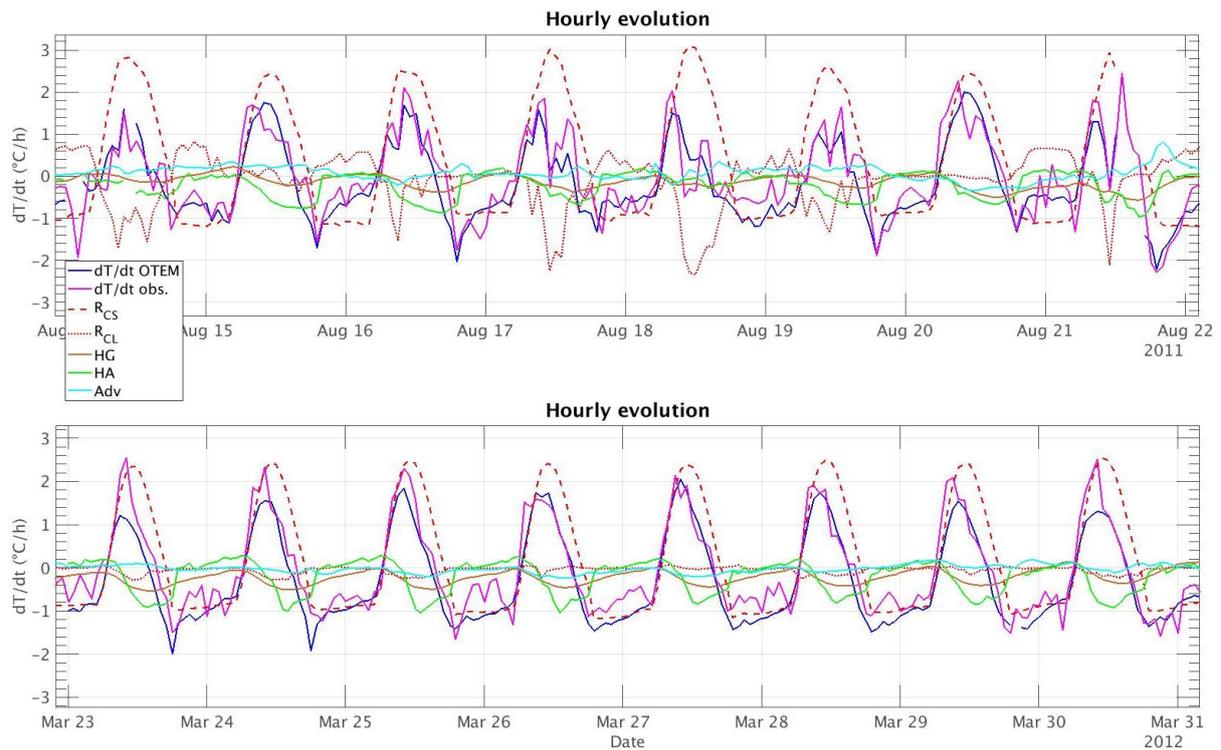
### 3.1.2 Hourly evolution of OTEM capability

Figure 3-4 is a zoom of figure 3.1.c for 8 days of mid-August 2011 (top panel) and the last 8 days of March 2012 (bottom panel). More the pink and the blue lines are closed, more OTEM is efficient. From this figure, we can see that:

- On average, OTEM follows quite well the observed one-hour temperature variations for the two cases shown in Figure 3-4. Indeed, for the days of August (Figure 3-4a),  $\frac{\partial T_{2m}}{\partial t}_{OTEM}$  is able to reproduce the sudden decreasing and increasing of  $\frac{\partial T_{2m}}{\partial t}_{obs}$  during daytime, which is particular for this period and probably due to the great variability of cloud cover which affects temperature variability, as previously shown by Bastin et al. (2018) at the SIRTA observatory
- OTEM is better for day than for nighttime when a greater difference of values between  $\frac{\partial T_{2m}}{\partial t}_{obs}$  and  $\frac{\partial T_{2m}}{\partial t}_{OTEM}$  is found for the latter, as shown in Figure 3-4. At

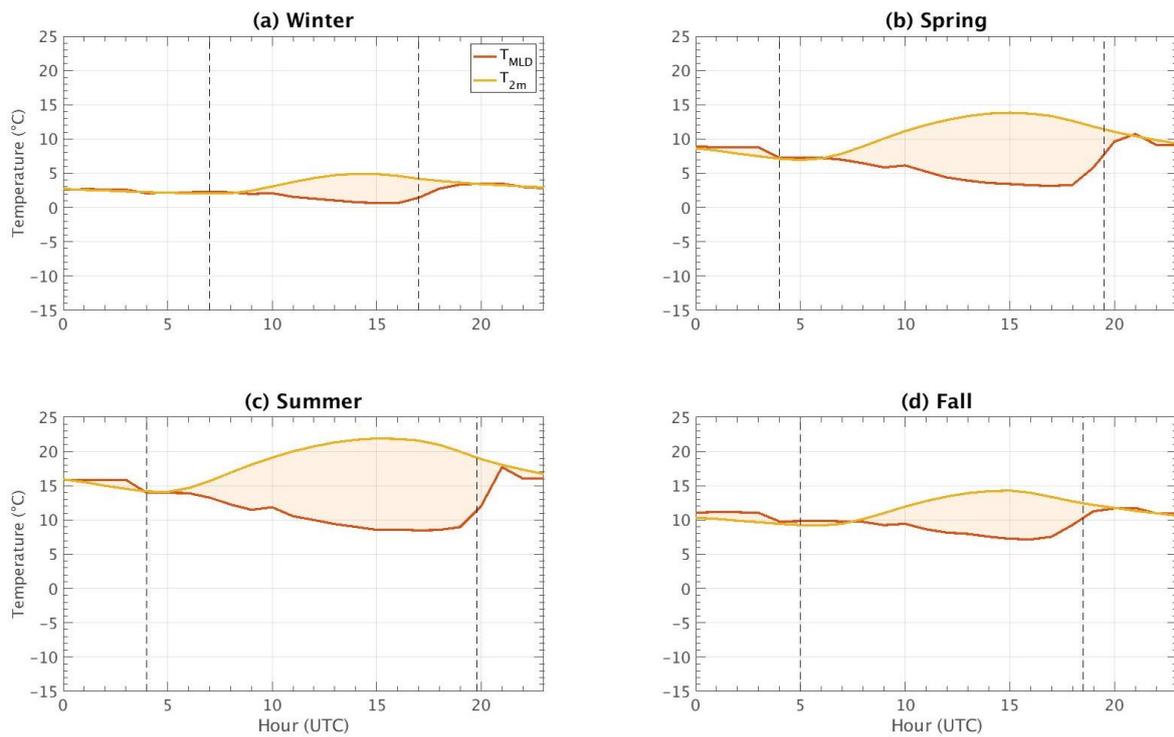
night, there is most of the time an underestimation of the observed temperature variations done by OTEM as seen during the nights of the first days of March 2012 in Figure 3-4 (bottom panel), which could be related to an overestimation of  $R_{CS}$  (red dashed line) which depends only on LW radiation since SW radiation is not available at this time (see Eq. 2.24). These days have high temperatures thanks to a clear-sky condition (not shown) with  $R_{CL}$  close to zero, and the radiative cooling at night is very important enhancing high values of  $F_{LW}^{\uparrow}$  (not shown), but probably an underestimation of  $F_{LW,CS}^{\downarrow}$  is done at this time and high values of  $R_{CS}$  are found during night.

- Some  $\frac{\partial T_{2m}}{\partial t}_{obs}$  peaks are not well reproduced by OTEM, as seen e.g. in the late morning of the 30<sup>th</sup> of March 2012 (Figure 3-4, bottom panel). Indeed, the  $HA$  term has high negative values at this early time (which is not seen on the other previous days), which leads to an underestimation of the observations by OTEM. This very low negative value of  $HA$  could be explained by the parametrization done to estimate this term in Section 2.2.1.2. Indeed, the temperature at MLD ( $T_{MLD}$ ) could be underestimated and thus enhancing a high temperature gradient which yields to a high  $HA$  value (cf. Eq. 2.27). This  $T_{MLD}$  underestimation probably comes from an error of measurement from the lidar instrument used to estimate MLD at this particular hour (10:00 UTC) since it presents an early and sudden peak of 1200 m which is not seen for the other days (not shown).
- When  $R_{CL}$  presents an important negative contribution (e.g. 18<sup>th</sup> August afternoon in Figure 3-4, top panel),  $HA$  becomes weaker and its diurnal cycle is attenuated due to the reduce of solar radiation and hence fewer surface heat fluxes are developed and the temperature gradient between  $T_{MLD}$  and  $T_{2m}$  is weaker than when there are fair-weather conditions. Ground heat exchange and advection terms play a minor role on this time scale and contribute very little to OTEM without detecting a very significant diurnal cycle.
- As for the last days of March 2012 (bottom panel in Figure 3-4),  $R_{CL}$  happens to present a very low (almost negligible) hourly contribution during daytime because of cloudless conditions (not shown), which allows to hourly temperature variations to have a normal gaussian diurnal cycle (similar to  $R_{CS}$ ).



**Figure 3-4:** Hourly evolution from 14<sup>th</sup> to 21<sup>th</sup> August 2011 (top panel) and from 23<sup>rd</sup> to 30<sup>th</sup> March 2012 (bottom panel) of the temperature variation for the different terms, the observations, and OTEM.

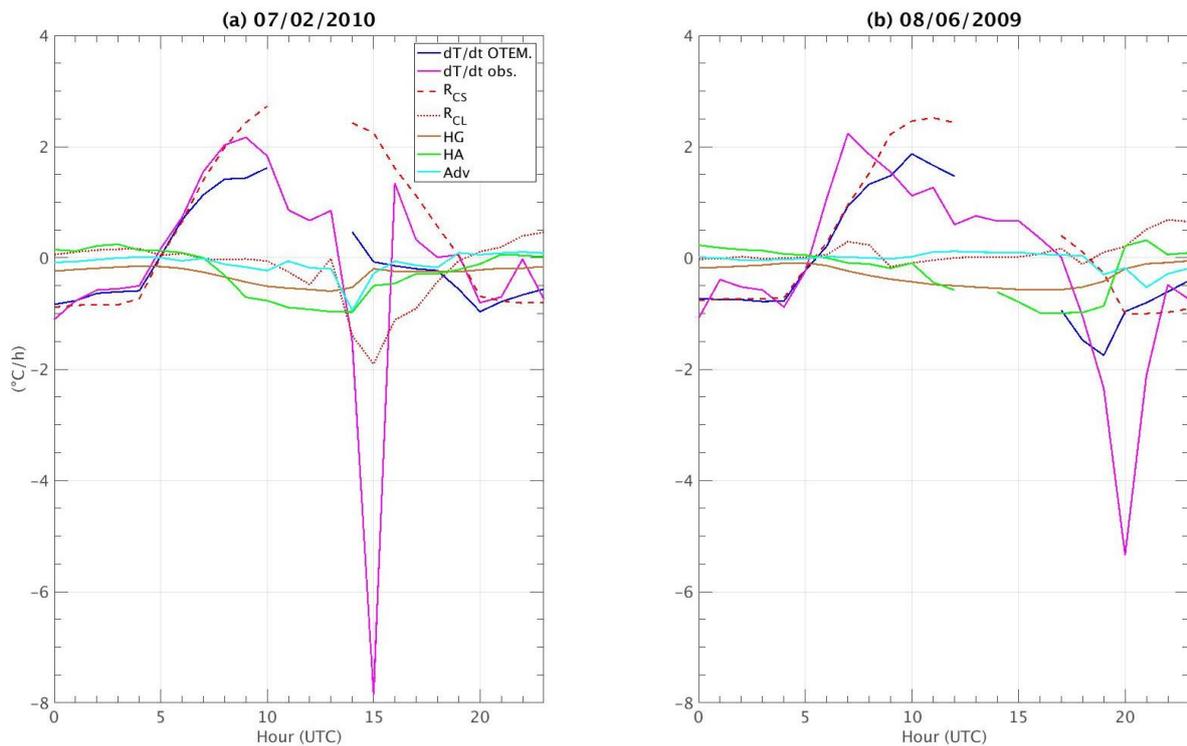
Evaporation and thermal conduction at the surface increase as the day goes on thanks to the increase of solar radiation arriving at the surface. However, the atmospheric heat exchange term is on average a negative contribution to temperature variations in one hour when  $F_{lat}$  and  $F_{sens}$  increase in the afternoons. Figure 3-5 shows the mean diurnal cycle split into seasons for  $T_{MLD}$  and  $T_{2m}$ , and the difference between these two temperatures represented as the colored-semitransparent brown interval, called  $\delta T$ . Thanks to longer hours with sun, the MLD develops more and reach higher values for summer and thus temperatures at these heights are lower (Figure 3-5c) than for the other seasons and the difference between the two temperatures is very high and negative. Since the  $HA$  term depends on this temperature difference, this term has an important negative contribution to temperature variations mostly in summer and spring. On the contrary, during winter MLD is very low and  $T_{MLD}$  has similar values as  $T_{2m}$  with a maximal difference of  $-5$  °C at 15:00 UTC (Figure 3-5a). During nighttime, for all the seasons of the year,  $\delta T$  is on average zero or very close to zero, partly due to the temperature inversion observed on average and the fact that MLD is close to the surface in response to the cooling of the air by the radiatively LW cooled surface and the absence of turbulence during this time (cf. 1.2.1.), therefore, temperature variations at night might be hard to quantify accurately.



**Figure 3-5:** Mean diurnal cycle averaged from 2009 to 2014 of  $T_{MLD}$  (red line),  $T_{2m}$  (yellow line) for (a) winter, (b) spring, (c) summer and (d) fall seasons. The colored-semitransparent brown space between  $T_{MLD}$  and  $T_{2m}$  in each subfigure represents its difference, called  $\delta T$ .

### 3.1.3 Focus on specific cases that OTEM does not capture

As seen in Figure 3-1c, there are times during summer when the observed surface temperature variation (pink line) decreases drastically in one hour ( $\frac{\partial T_{2m}}{\partial t}_{obs} < -4 \text{ } ^\circ\text{C/h}$ ), and OTEM is not able to capture this behavior. These cases represent barely 0.1% of the database: there are only few cases but it is interesting to understand the reason why OTEM is not able to capture them. In order to better understand this, two different cases are studied in details (Figure 3-6). Figure 3-6a corresponds to the 2<sup>nd</sup> of July 2010 when a  $-8.0 \text{ } ^\circ\text{C/h}$  observed temperature variation occurred at 15:00 UTC, and Figure 3-6b shows the values of 6<sup>th</sup> of August 2009 with a negative pic of  $\frac{\partial T_{2m}}{\partial t}_{obs} = -5.3 \text{ } ^\circ\text{C/h}$  at 20:00 UTC (22:00 LT), with a  $\text{SZA} > 89^\circ$  which usually corresponds to the sunset or early night at that time in August.



**Figure 3-6:** Hourly evolution for the two days in which the lowest hourly observed temperature variations are found: (a) 2<sup>nd</sup> of July 2010 and (b) 6<sup>th</sup> of August 2009. Gaps in  $\frac{\partial T_{2m}}{\partial t}_{O TEM}$  (blue line) are due to the absence of  $F_{LW}^{\uparrow}$  for these hours that do not allow to estimate  $\Delta F_{NET,CS}$  and thus  $R_{CS}$  (cf. Eq. 2.22).

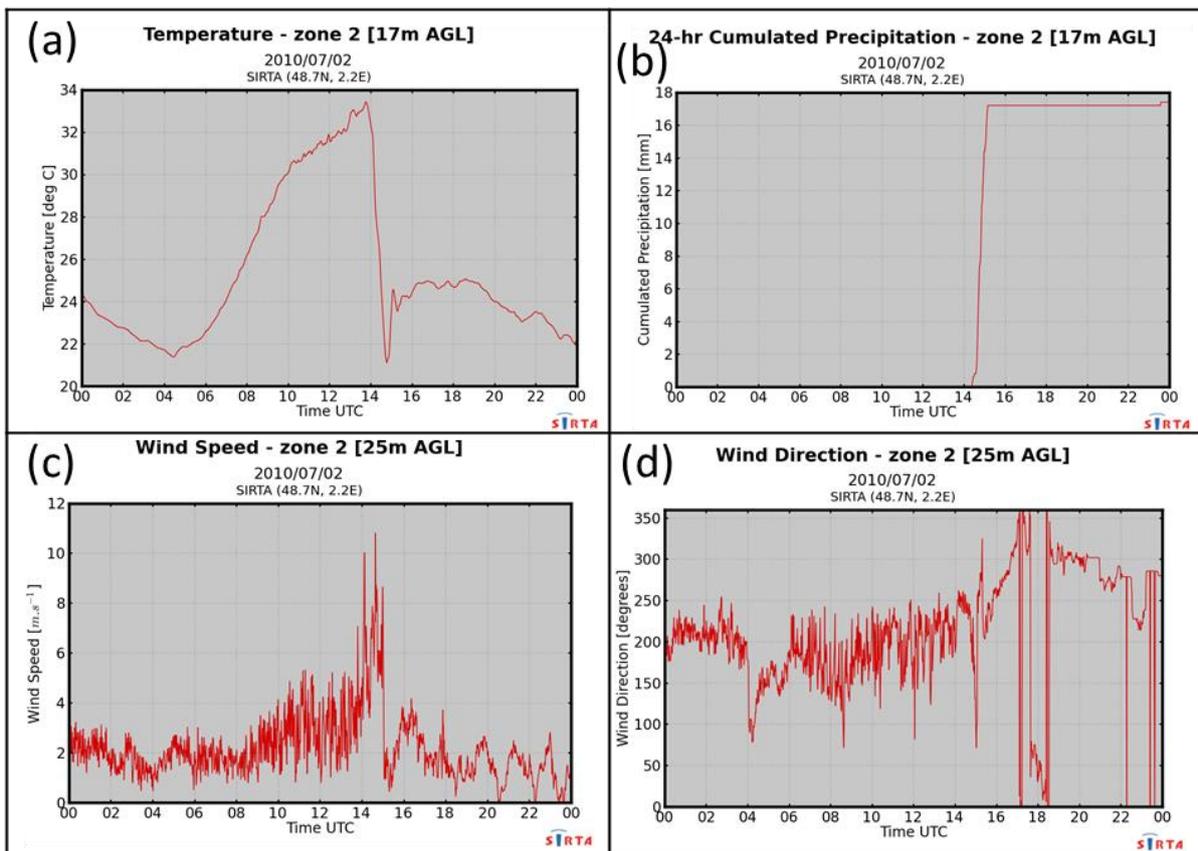
### 3.1.3.1. 07/02/2010

In the 07/02/2010 at 15:00 UTC,  $R_{CL}$  has an important negative contribution with a peak of -2 °C/h and the  $Adv$  term has a negative pic one hour before. The  $HA$  term remains negative but close to zero, as  $HG$ , and the  $Adv$  term presents a peak one hour before. Anyway, none of these terms are able to capture the behavior that we have in reality, as OTEM value at this hour is only -0.2 °C/h.

To analyze this anomaly of temperature variation, Figure 3-7 presents the Quicklooks for this day of different meteorological variables measured at SIRTA with a temporal resolution of 1-min (available at <https://sirta.ipsl.fr/sirta/data/quicklooks/>). The important and sudden decrease of temperature at 15:00 UTC could be associated with a cold pool event, which is more occurrent in summer thanks to higher surface temperatures than those in winter. This 7<sup>th</sup> of July was particularly warm as seen in Figure 3-7a with  $T_{2m}$  lowest value of 22 °C and reaching 33°C. The mentioned cold pool is characterized by storms and heavy precipitations. Indeed, as shown in Figure 3-7b, during all this day it did not rain until around 14:20 UTC when it rained a total of 17 mm in just 40 minutes and then it immediately stopped just after 15:00 UTC with no

more rain detected for the rest of the day. Figure 3-8 presents pictures from the SIRTA sky imager: during the morning just few clouds were present, and between 14:30 and 15:00 UTC a strong cloud cover is seen. Furthermore, at the time this event occurred, wind speed increases as a new air mass arrived at the site coming from two completely opposite directions within just few minutes (Figure 3-7c and Figure 3-d, respectively).

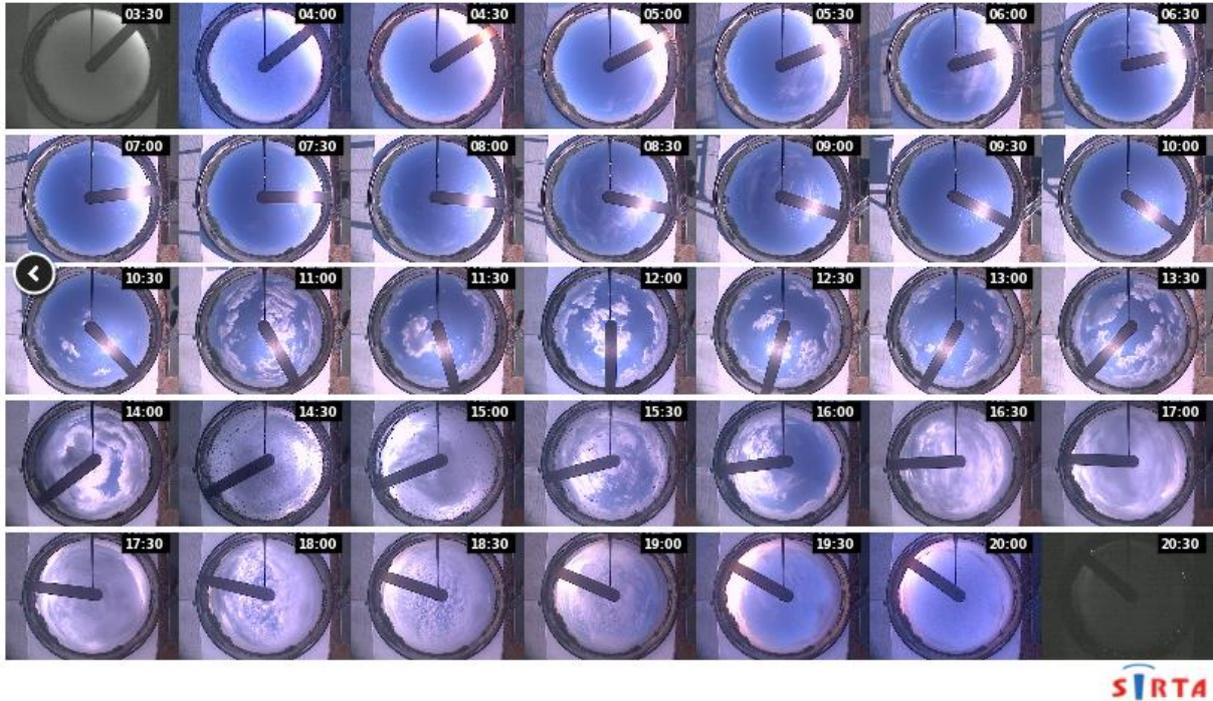
All these figures show that this event occurred in less than an hour, which is the resolution of the data in SIRTA-ReOBS and ERA5/ERA5-Land datasets, and then in OTEM. Therefore, the hourly values in the observations (as well as in both reanalysis) do not allow in the first place to capture this extreme event properly in OTEM, since it is related to rapid cloud formation ( $<1$  h) that is not well captured by the  $R_{CL}$  term, and rapid wind change direction not captured by the  $Adv$  term. This could lead to an underestimation of the contribution of both terms in OTEM. However, since the cold pool event is rare and ephemeral on a local scale (Llasat and Puigcerver, 1990; Conangla et al., 2018), its presence does not bias significantly the general performance of the model.



**Figure 3-7:** 07/02/2010 quicklooks 1-min values of (a)  $T_{2m}$ , (b) 24-h cumulated precipitation, (c) wind speed, and (d) wind direction. These quicklooks are available at <https://sirta.ipsl.fr/sirta/data/quicklooks/>.

## TSI-440 Total Sky Imager

2010/07/02  
SIRTA (48.7N, 2.2E)



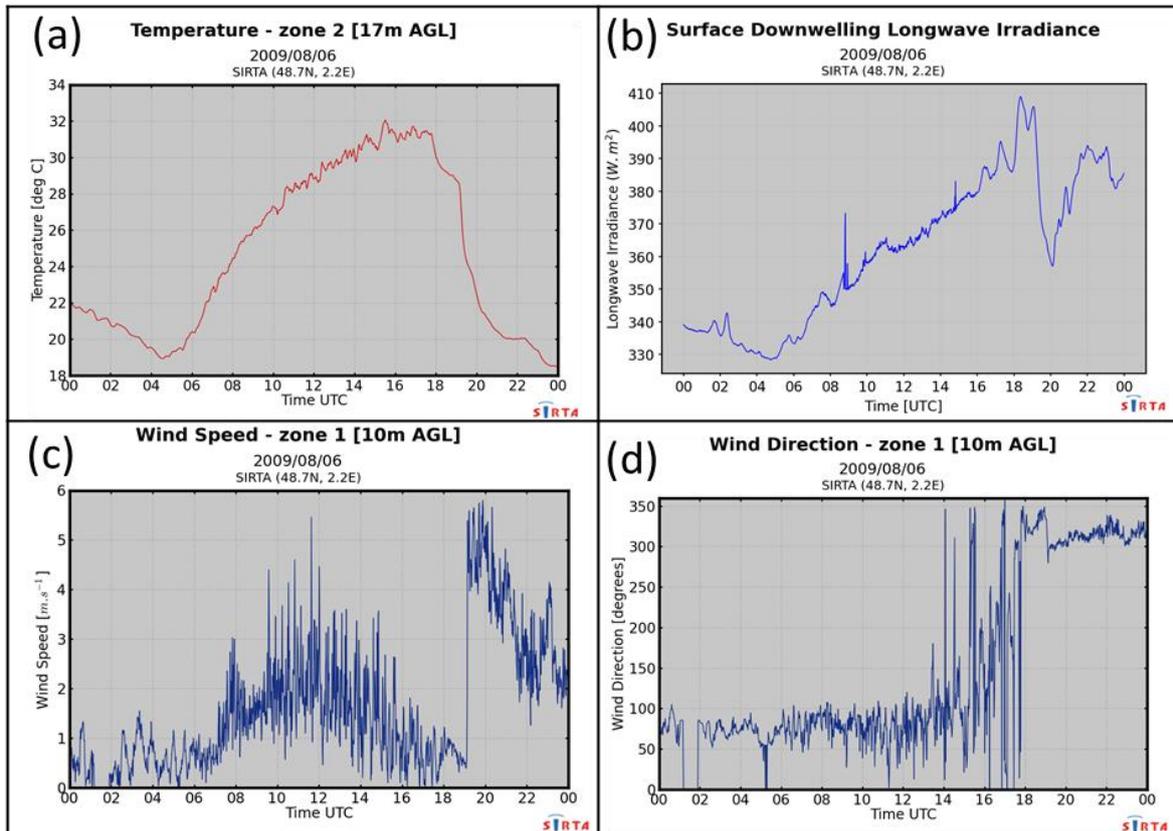
**Figure 3-8:** 07/02/2010 sky imager photographs took every 30 min from 03:30 to 20:30 UTC.

### 3.1.3.2. 08/06/2009

For this second date,  $\frac{\partial T_{2m}}{\partial t}_{obs} = -5.3 \text{ }^\circ\text{C/h}$  at 20:00 UTC, with the temperature decreasing from 28 °C at 19:00 UTC to 23 °C at 20:00 UTC (Figure 3-9a) just at sunset, and OTEM underestimates this abrupt temperature decrease with a value of only -1.8 °C/h.  $R_{CL}$  has a positive contribution to temperature variations (Figure 3-6b), which is expected at this hour since  $SW \downarrow \approx 0$  and only  $LW \downarrow$  is contributing (cf. Eq. 2.22), as shown in Figure 3-9b. The  $HA$  term is positive at this hour because  $T_{MLD} > T_{2m}$ , when an hour before it was negative. This change of sign is due to the rapid diminution of MLD from 2100 m at 19:00 to 300 m at 20:00 UTC (not shown), meaning an important – and unusual – decrease of 1800 m in one hour, and since during nighttime the hypothesis of working on a stable atmosphere is set, there is an underestimation of  $HA$  term, when in reality the atmosphere could still be in an unstable state.

The most likely reason why OTEM does not capture well this important temperature decrease is the underestimation of the  $Adv$  term. Indeed,  $Adv$  has a negative – but still very weak – peak one hour after the event (Figure 3-6b), and a sudden increase in wind speed is detected at 19:00 UTC along with a change wind direction changing from a

N-E to a N-W regime (Figure 3-9c and d, respectively). Since the  $Adv$  term is estimated with a zonal and meridional temperature gradient (cf. Eq. 2.31) considering the closest point to the one where SIRTA is located (Appendix A.5, Figure A-5), this abrupt change of direction could not be captured by ERA5 time and spatial resolution. Indeed, this sudden and rapid wind change direction occurred within few minutes (<1 h) and thus the  $Adv$  term does not capture this event.



**Figure 3-9:** 08/06/2009 quicklooks 1-min values of (a)  $T_{2m}$ , (b) surface downwelling longwave irradiance, (c) wind speed, and (d) wind direction.

### 3.2 EVOLUTION OF THE DIFFERENT OTEM TERMS

The monthly-hourly mean contribution of each term is here assessed in order to better understand the extent to which SEB processes contribute to temperature variations, and the times when one of them modulates or mitigates the others for different seasons and times of the day.

#### 3.2.1 Annual contribution of each term

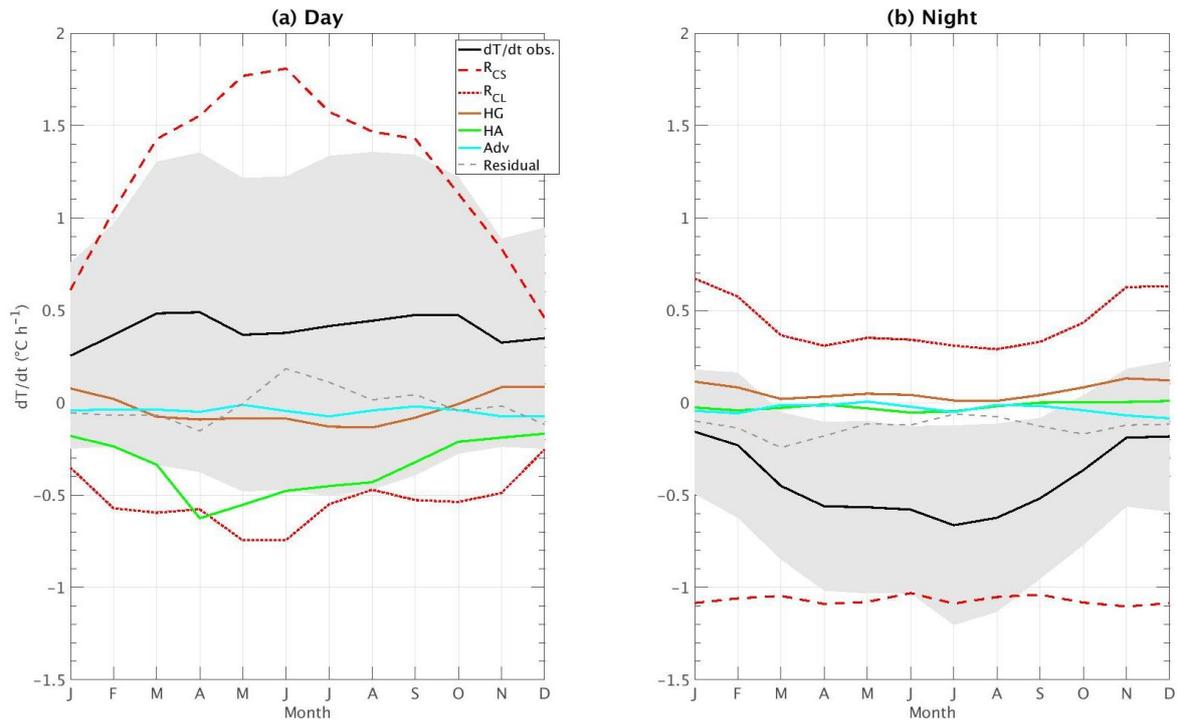
The mean contribution of each term is averaged monthly for day and night separately (Figure 3-10).

Figure 3-10a shows that all the terms have a seasonal cycle during daytime. The annual

cycle of the  $R_{CS}$  ( $R_{CL}$ ) term is important, by having a positive (negative) contribution during the day, mostly dominating the temperature variations at this scale (and also when mean diurnal cycle are estimated in Appendix B.1, Figure B-6). On average,  $R_{CS}$  has the biggest contribution amplitude values ranging from 0.48 °C/h for January to a maximal of 1.75 °C/h in June. On the other hand,  $R_{CL}$  presents a low negative contribution also in January around -0.25 °C/h when the solar radiation is weak and thus preventing clouds to strongly reduce it, and a maximal and more important negative contribution in both Mai and June of -0.75 °C/h. This latter term presents less month-to-month variability compare to  $R_{CS}$  especially during the J.A.S.O.N. months when clouds cool the surface with a constant value of -0.5 °C/h whereas the clear-sky radiative contribution varies from 1.6 °C/h in July to more than half of this value (0.75 °C/h) in November. This constant value in  $R_{CL}$  is more controlled by the difference  $F_{SW}^{\downarrow} - F_{SW,CS}^{\downarrow}$  rather than  $F_{LW}^{\downarrow} - F_{LW,CS}^{\downarrow}$  because the last one remains on average constant not only during these months but through the whole year whereas the difference in SW radiation varies (not shown) and thus drives the shape of the  $R_{CL}$  term. In addition, the  $HA$  term plays an important role in temperature variations in April, having a similar cooling effect as  $R_{CL}$  of -0.7 °C/h, and for the rest of the year this term continues cooling the surface with its contribution decreasing while approaching to November, December and January when hourly solar availability decreases. Furthermore,  $\frac{\partial T_{2m}}{\partial t_{obs}}$  finds two maximal average values during daytime in March and April and then in September and October, values that do not coincide with the maximal contribution of the OTEM terms. This could be linked to the fact that for spring and fall, lesser cold anomalies occurred compared to winter and summer, related – and explained – partly to the presence of weather regimes for these two latter seasons. Indeed, several studies have shown that weather regimes are largely related to surface temperature anomalies mostly in winter and summer seasons (Rojas et al., 2013a; Sousa et al., 2018b; Yiou et al., 2008; Folland et al., 2009; Cassou et al., 2005) and to other meteorological anomalies such as wind speed (Cortesi et al., 2019) and precipitations (Xoplaki et al., 2004b). Note that the standard deviation of  $\frac{\partial T_{2m}}{\partial t_{obs}}$  (represented as the shaded gray area in Figure 3-10a) has also great values for these transition months, explaining also the increase in observed temperature variations.

According to Figure 3-10b for nighttime, the clear-sky and cloud radiative forcing terms contribute the most in values to hourly temperature variations at the surface and the other terms contribution remain close to zero.  $R_{CL}$  is always positive, has an annual cycle more pronounced than the other terms, and has an opposite behavior with respect to the one at daytime: its highest contribution occurs in January, November and December with a warming effect of 0.70 °C/h mostly due to the increase in cloud cover and the lesser cumulative energy received by the surface during daytime, while it is getting weaker and constant from March to October with a mean value around 0.30 °C/h for all these months. As for  $\frac{\partial T_{2m}}{\partial t_{obs}}$ , it has almost the same shape as  $R_{CL}$  since

this term is the one contributing the most at this time.



**Figure 3-10:** Mean annual cycle averaged monthly from 2009 to 2014 of the five terms of OTEM, the observations, and the residual (dashed gray lines) for (a) day and (b) night. The shaded gray area in each of the panels represents the standard deviation of  $\frac{\partial T_{2m}}{\partial t}_{obs}$ .

### 3.2.2 Monthly-hourly contribution of each term

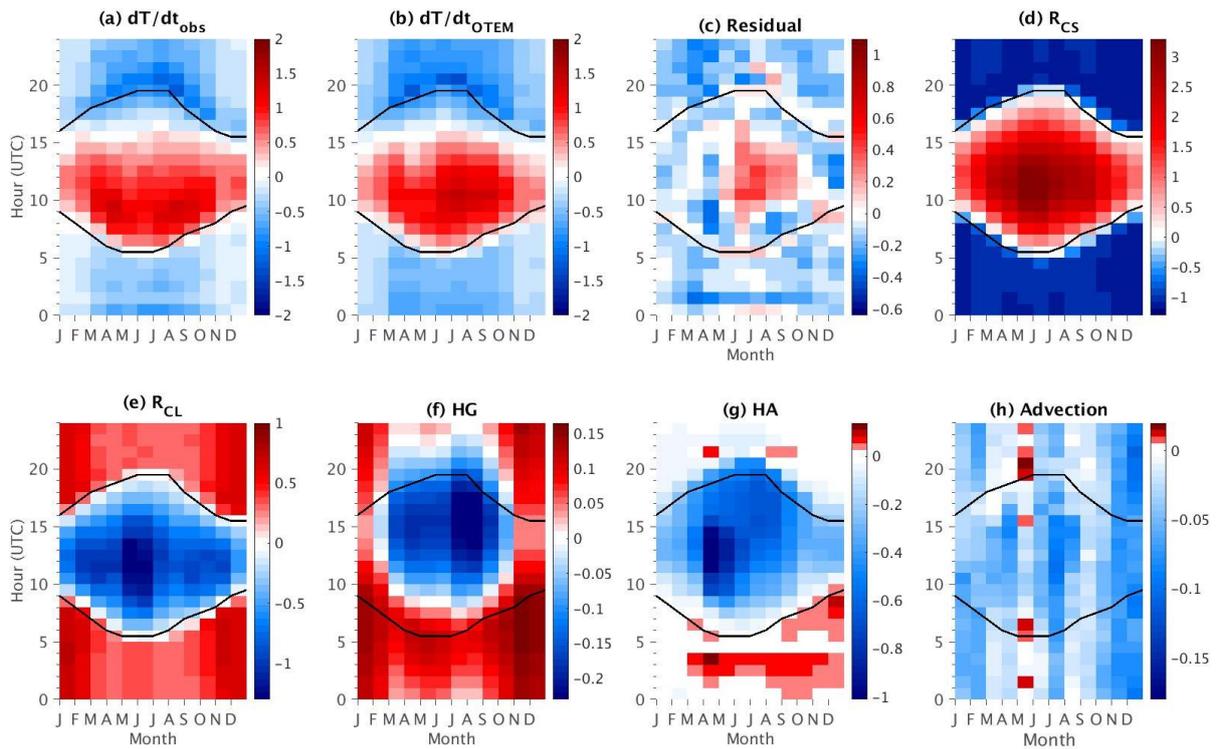
Figure 3-11 presents the values of the monthly-hourly mean values from January 2009 to February 2014 for the observations, for OTEM, the residual term which is calculated as the difference between OTEM and the observations (i.e.  $\frac{\partial T_{2m}}{\partial t}_{OTEM} - \frac{\partial T_{2m}}{\partial t}_{obs}$ ), and the five terms of OTEM.

A diurnal cycle clearly appears for all the terms except for the *Adv* term (Figure 3-11h), which is mainly driven by the solar radiation intensity during the day. Further, a maximal negative (positive) contribution to temperature variations of  $R_{CL}$  ( $R_{CS}$ ) is found during the late mornings for all the months, cooling (warming) the surface with a maximal mean value of  $-1.3$   $^{\circ}\text{C/h}$  ( $3.2$   $^{\circ}\text{C/h}$ ) in May and June at 10:00 UTC and 11:00 UTC. In spring and fall seasons there is an important contribution of *HA* term (Figure 3-11g) in the late morning and the afternoon, especially during spring, with an hourly maximum averaged contribution of  $-1.1$   $^{\circ}\text{C/h}$ . The difference between day and night *HA* contribution is the largest during spring and summer compared to the other seasons when the MLD can reach high values in the afternoon due to increased turbulence,

especially in spring when the absence of latent heat flux data overestimates its contribution in the morning.  $HA$  contribution is minimal in winter due to the low development of the MLD. This result is in contrast with the negative (and very low) contribution of the  $HG$  term, which is maximal in summer in the afternoon with an average value of  $-0.25$  °C/h when the surface temperature is the highest and thus strengthens its cooling action during daytime. An opposite effect is found in winter when the ground is usually warmer than the surface, yielding to a positive contribution of  $HG$  (Figure 3-10a and Figure 3-11f). As expected, the  $Adv$  term does not present a strong monthly-hourly cycle compared to the other terms, although one can distinguish a mean negative contribution (still very low) at all seasons with a mean minimum in July in the afternoon of  $-0.12$  °C/h, as shown Figure 3-11h.

During nighttime, Figure 3-11a shows that the observed temperature variations have a minimal averaged value just after the sunset for all the months of the year, that OTEM also reproduces with good accuracy, since the difference between the two datasets is quite low at these times (Figure 3-11b and Figure 3-11c). Further, for winter,  $\frac{\partial T_{2m}}{\partial t}_{obs}$  is on average low, especially after midnight when surface temperatures remain almost constant, whereas in spring and summer, its hourly variation increases. Indeed, the surface reaches high temperatures during daytime, and in the night, the energy emitted is much greater than the energy received, especially in summer when the cumulative energy the surface receives during daytime is much larger thanks to larger sunny hours, than in winter. This explains partly why the contribution of clouds is lower at night while approaching to warmer seasons (Figure 3-10b, red dotted line).

Lastly, the residual, defined as the difference between OTEM and the observations (i.e.  $\frac{\partial T_{2m}}{\partial t}_{OTEM} - \frac{\partial T_{2m}}{\partial t}_{obs}$ ), also has a seasonal variability and is mainly negative, but a minimum difference of  $-0.45$  °C/h occurs in April in the morning (Figure 3-11c), which is related to the negative increase of the  $HA$  term at that time. This is partly due to the important absence of latent heat flux data especially for this month, which implies an increase in the residual value due to  $\tau_a$  estimation (see Table 2.1 and Section 2.2.1.2.b). Nevertheless, during the months when solar radiation is strong, the residual reaches positive values (with a maximum value around  $0.4$  °C/h) but OTEM overestimation remains low. These values for the residual are low compared to the values of  $R_{CS}$ , but despite it remains almost at the same order of magnitude as the  $R_{CL}$  and  $HA$  terms, a good correlation is found during daytime (cf. Section 3.1.1.).



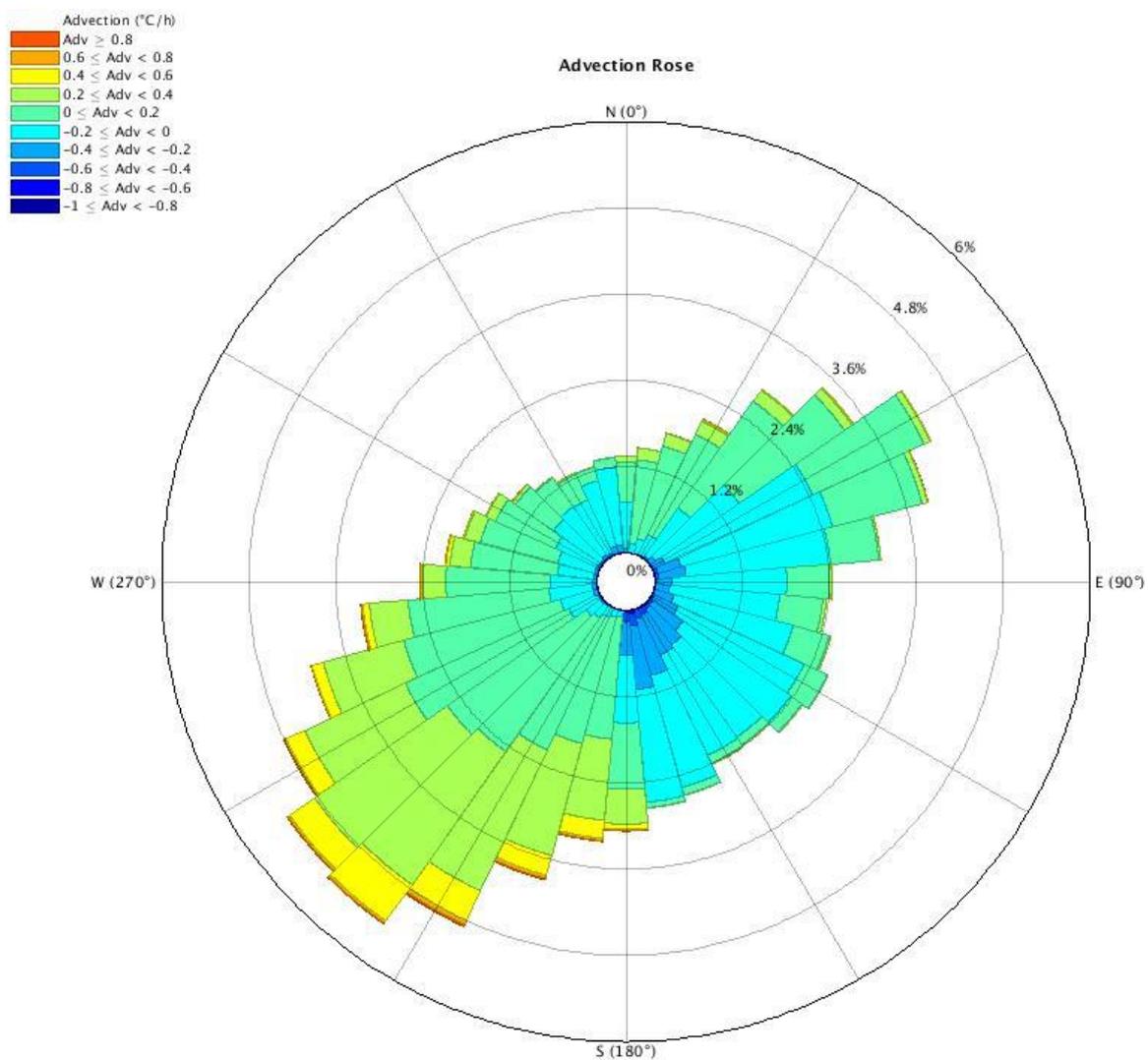
**Figure 3-11:** Monthly-hourly mean values for (a)  $\frac{\partial T_{2m}}{\partial t}_{obs}$ , (b)  $\frac{\partial T_{2m}}{\partial t}_{OTEM}$ , (c) the residual (i.e. difference between OTEM and the observations), for (d)  $R_{CS}$ , (e)  $R_{CL}$ , (f) HG, (g) HA and (h) Adv. Units in the color bars are all in  $^{\circ}\text{C}/\text{h}$ , and their scale is different for each subfigure. The black contour lines in each figure correspond to sunrise (bottom line) and sunset (top line) approximative hours.

Focusing on the transition periods (sunrise and sunset, black lines in Figure 3-11), the residual presents low values at these times. Indeed, there is a slight underestimation of the model of about  $-0.13\text{ }^{\circ}\text{C}/\text{h}$  for some months (e.g. February) at sunrise hours, whereas a low overestimation with close-to-zero residual mean values are found for May and June. For the sunset, a similar behavior is found as for the sunrise (with very similar values for the residual term). Therefore, a good agreement is found between the model and the observations for these specific hours.

### 3.2.3 Advection dependence on wind direction

In order to study the dependence of advection on wind direction and its contribution to hourly surface temperature variations, an advection distribution (or advection rose) is estimated and plotted in Figure 3-12. The wind direction is directly estimated from the two horizontal components of wind velocities ( $u$  and  $v$ ), retrieved from ERA5-Land (cf. Section 2.2.1.2.d). This figure shows us that:

- The negative contribution of the *Adv* term is mostly coming from N-E and S-E wind regimes, whereas the positive contributions are from S-W regime. Few percentages of winds are coming from N-W. Indeed, this term contributes strongly to cooling the surface for a S-E regime.
- Figure 3-12 also illustrates the two majority winds regimes coming to the Paris region: the well-known Siberian High (D'Arrigo et al., 2005) which brings on average cold winds, and the Westerly winds coming from the Atlantic Ocean with warmer and humid near-surface air.



**Figure 3-12:** Advection distribution at SIRT from 2009 to February 2014.

### 3.3 IMPORTANCE ESTIMATION OF EACH TERM: THE RANDOM FOREST APPLICATION

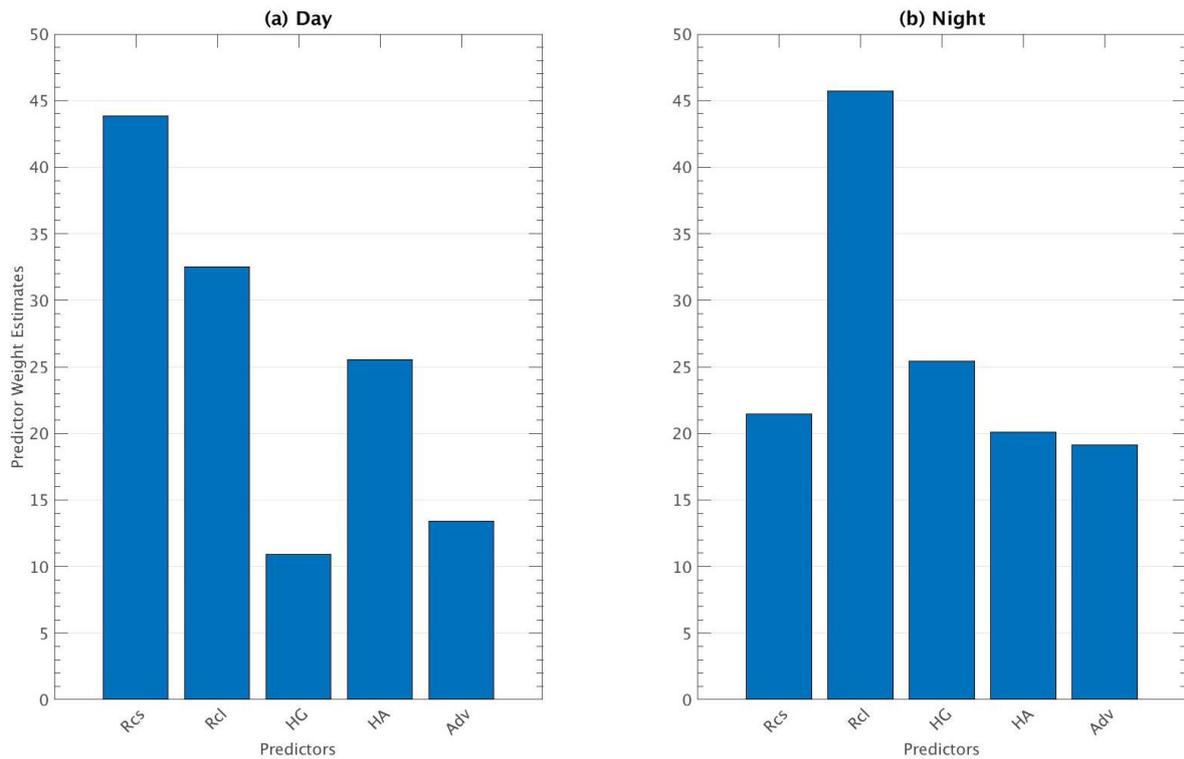
Here, the random forest method is used to establish which term dominates (in terms of weight and not in contribution) the temperature changes in the model ( $\frac{\partial T_{2m}}{\partial t_{OTEM}}$ ) during day and nighttime. As explained in Chapter 2, Section 2.2.2., this machine learning method consists of bootstrapped-aggregated decision trees, a method that combines and gathers together the results of these trees to construct more powerful prediction models. Since in the present study the model has already been constructed, one of random forest method most impressive feature is used, which consists of the ability to provide a fully nonparametric estimation of the importance of each term (or predictor) on the model (see Section 2.1.2.). Additionally, this method is used because of the nonlinear relationship between each single term and OTEM (not shown). This non-linearity does not yield to estimate how each process at the surface affects the temperature variations (contrary to what is done in Miller et al., 2017).

#### 3.3.1 General estimation

Figure 3-13 presents the predictor (i.e. each term involved in  $\frac{\partial T_{2m}}{\partial t_{OTEM}}$ ) weight estimate value for daytime and nighttime. Figure 3-13a corroborates that  $R_{CS}$  is the most important term during daytime, followed by  $R_{CL}$  and then  $HA$ , whereas  $HG$  is the least important term OTEM for the timescale considered. Next, Figure 3-13b illustrates that  $R_{CL}$  dominates hourly temperature variations during nighttime, followed by  $HG$  and  $R_{CS}$ . These results agree with Kukla and Karl (1993), who suggested that the key regulator of nighttime warming temperatures is downward thermal radiation when cloudy cases are present, an effect that is damped most of the time by the radiative cooling effect of the ground.  $HA$  term play a very minor role since surface turbulence decreases due to the absence of solar radiation.

#### 3.3.2 Diurnal cycle of the weight estimates

The separation of daytime/nighttime does not allow the consideration of the importance of processes occurring during transition times of the diurnal cycle when the increase/decrease in temperature is the strongest. So, to better understand the influence of each term on hourly temperature variations, the weight estimation value is processed for each hour using the random forest method. Figure 3-14 presents the results of this method for each season. This diurnal cycle is estimated by applying the random forest to each hour separately. For instance, to estimate the importance of each term on the observed temperature variations at 10:00 UTC in summer, decision trees are built and trained from all the data sample corresponding to this condition of season and hour, and then the importance estimation of each term is retrieved. This process is repeated for each hour for every season of the year and then a diurnal cycle is obtained.



**Figure 3-13:** Predictor weight estimates obtained by the random forest method for (a) day and (b) night. The abscissa in both cases represents each predictor (or term) of OTEM, and the y-axis their weight (unitless) defined as the sum of their mean square error when permutation in the decision trees is done.

Figure 3-14 supports a clear and reliable estimation of the importance (i.e. weight) of each term split into seasons for every hour of the day, which is not seen by estimating the diurnal and annual cycle of each term (cf. Section 3.2.). Indeed, depending on the hour of the day (and thus the state of the atmosphere), one term will become more important over the rest and temperature variations will be more sensitive to its change even if its contribution in terms of magnitude remains smaller compared to other terms. For instance,  $R_{CS}$  absolute contribution during nighttime is higher than that of  $R_{CL}$  (Figure 3-10b), and yet the latter is more important during nighttime due to its seasonal and hourly contribution variability to  $\frac{\partial T_{2m}}{\partial t}_{obs}$ .

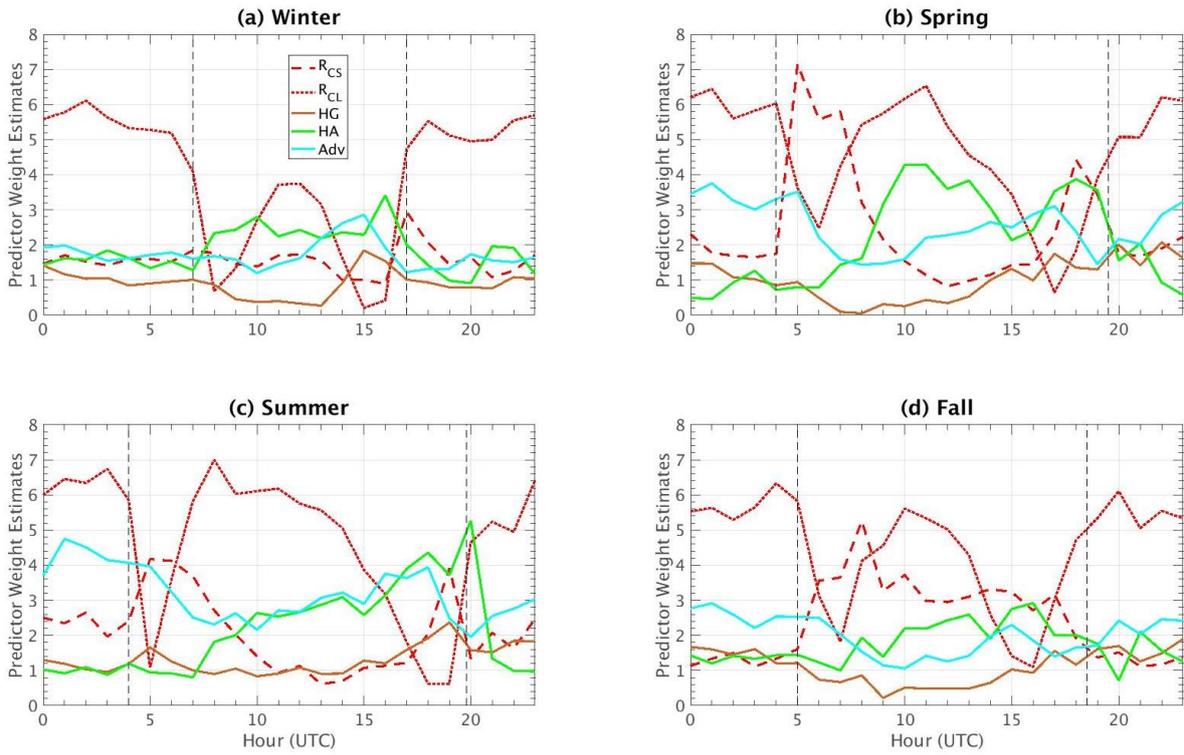
As expected (and previously exposed), for all the seasons  $R_{CL}$  is the term dominating during nighttime, just after sunset, and before sunrise (indicated by vertical dashed lines in black). After sunrise, the surface heating produced by the sun in the early morning enhances a high growth rate in the temperature variations, whose effect makes  $R_{CS}$  the dominant term driving  $\frac{\partial T_{2m}}{\partial t}_{OTEM}$  at those hours of the day for all seasons, except for winter. In winter (Figure 3-14a), the weight of  $R_{CS}$  is almost the same as  $R_{CL}$  one and is weak. This effect is due to the weak SZA for this season, and the surface heating by the sun in clear-sky conditions is not strong enough to modulate

temperature variations. On the contrary,  $R_{CL}$  weight is minimal just after sunrise and before sunset, where, depending on the season,  $R_{CS}$ ,  $HA$  or  $Adv$  are the main terms controlling temperature variations.

At the beginning of the day, it is  $R_{CS}$  that dominates for all the seasons, then  $R_{CL}$  becomes the term dominating temperature variations until the afternoon. This effect is partly explained by the values of the terms themselves:  $R_{CS}$  standard deviation for a given hour/season is weak, and thus its influence to explain the difference between one day to another at a specific hour remains minimal and it is not a strong predictor at diurnal cycle scale, especially in the summer and spring (Figure 3-14b and Figure 3-14c, respectively) when other variables will modulate temperature variations. On the other hand,  $R_{CL}$  turns into the main modulator of  $\frac{\partial T_{2m}}{\partial t}_{OTEM}$  for all the seasons due to its strong standard deviation for a given hour/season, reaching its maximal weight in summer (Figure 3-14c). Therefore, the hourly temperature variations are more sensitive to cloud changes rather than solar radiation which does not vary significantly for a specific hour from day to day within a season.

Concerning the other terms,  $HA$  weight is growing along the day (particularly in summer). Its variation is greater than that of the other terms, making  $HA$  the second most important modulator of temperature variation, except for winter (Figure 3-14a). For this latter season, its contribution is weak due to the lack of solar radiation and vegetation and few turbulent heat fluxes develop, along with a weak MLD. It becomes sometimes the most important term in the late afternoon just before sunset. This is explained by an increase in instability of the atmospheric boundary layer enhanced by its large extension and surface temperature gradient: actually  $HA$  depends on the difference between  $T_{MLD}$  and  $T_{2m}$  (Eq. 2.27) which is maximal in the late afternoon for all the seasons (cf. Figure 3-5) and is generated by the increase in turbulence fluxes near the surface.

Regarding the  $Adv$  term, it shows an important weight in some hours in the late afternoon in winter, which makes it the term controlling on average hourly temperature variations at that time (then it is  $HA$  who becomes more important). In summer (Figure 3-14c), it presents an important increase as the day goes on similar to  $HA$  after 10:00 UTC, but  $HA$  is even more important thanks to a development of turbulent heat fluxes at the surface in the late afternoon.  $Adv$  becomes the second most important modulator of the temperature variation during nighttime before sunrise for all seasons. A thermally-induced circulation linked to the urban heat island which is set up in Paris could be at the origin of this weight: this circulation is likely to create a greater variation of temperature some nights when cooler air from rural areas is advected to warmer zones towards the city center. The SIRTA observatory, located in a suburban area at 20 km from the center of Paris could be indeed regularly affected by this circulation.

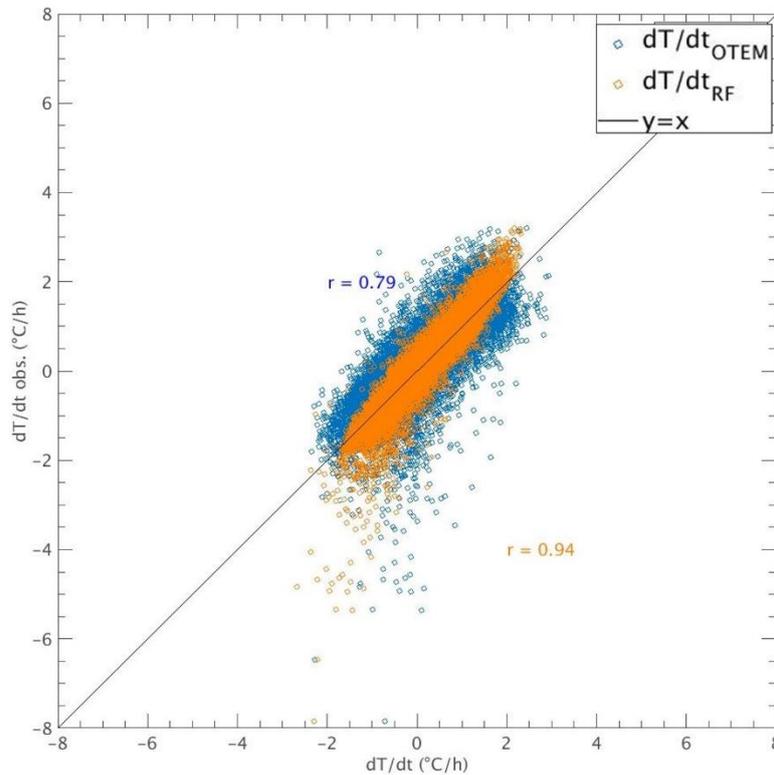


**Figure 3-14:** Diurnal cycle of the predictor weight estimate for each term of OTEM, for (a) winter, (b) spring, (c) summer and (d) fall. Vertical dashed black lines are for sunrise and sunset mean hours.

### 3.3.3 Validation of the random forest method

However, this machine learning method is generally used in other studies to train and to have better estimations of a particular model. In order to validate the random forest method skill on predicting new observed temperature variations, the method is used to predict  $\frac{\partial T_{2m}}{\partial t}_{obs}$  (rather than  $\frac{\partial T_{2m}}{\partial t}_{OTEM}$  as it is done to estimate the weight of each term in Section 3.3.2.). The output for this case is called  $\frac{\partial T_{2m}}{\partial t}_{obs,RF}$ . A comparison between  $\frac{\partial T_{2m}}{\partial t}_{OTEM}$  (i.e. the linear sum of the five terms) and the new  $\frac{\partial T_{2m}}{\partial t}_{obs,RF}$  is done, the results of this validation are shown in Figure 3-15. Indeed, the scatterplot before performing the random forest method ( $\frac{\partial T_{2m}}{\partial t}_{OTEM}$ ) shows the distribution of values between the observations and the model (i.e.  $\frac{\partial T_{2m}}{\partial t}_{obs}$  vs  $\frac{\partial T_{2m}}{\partial t}_{OTEM}$ , blue points) as found in Figure 3-2a. Then, when the random forest method is performed and the data are trained based on  $\frac{\partial T_{2m}}{\partial t}_{obs}$  (instead of  $\frac{\partial T_{2m}}{\partial t}_{OTEM}$ ), better predictions are obtained between  $\frac{\partial T_{2m}}{\partial t}_{obs}$  and  $\frac{\partial T_{2m}}{\partial t}_{obs,RF}$  (orange points) and the correlation coefficient has a higher value (0.94). In such case, the RF method gives better estimations of temperature variations but the

retrieve of the function used to have these results is not available. Nevertheless, this result validates the fact of considering temperature variations as the sum of the five terms to estimate their importance using the RF method (when it is used to predict the modeled temperature variations).



**Figure 3-15:** Scatter plot of  $\frac{\partial T_{2m}}{\partial t}_{obs}$  as a function of the developed model  $\frac{\partial T_{2m}}{\partial t}_{OTEM}$  before applying the random forest method (blue circles) and  $\frac{\partial T_{2m}}{\partial t}_{obs}$  as a function of the model trained after the RF method is applied (orange circles). The correlation coefficient “r” for both scatterplots are also presented.

### 3.3.4 Basic settings on random forest method to study the weight of each term

In Section 3.3.2, the random forest method is used to determine the importance of the predictors (terms) on the near-surface temperature variations. Detailed information on how this method works is given here.

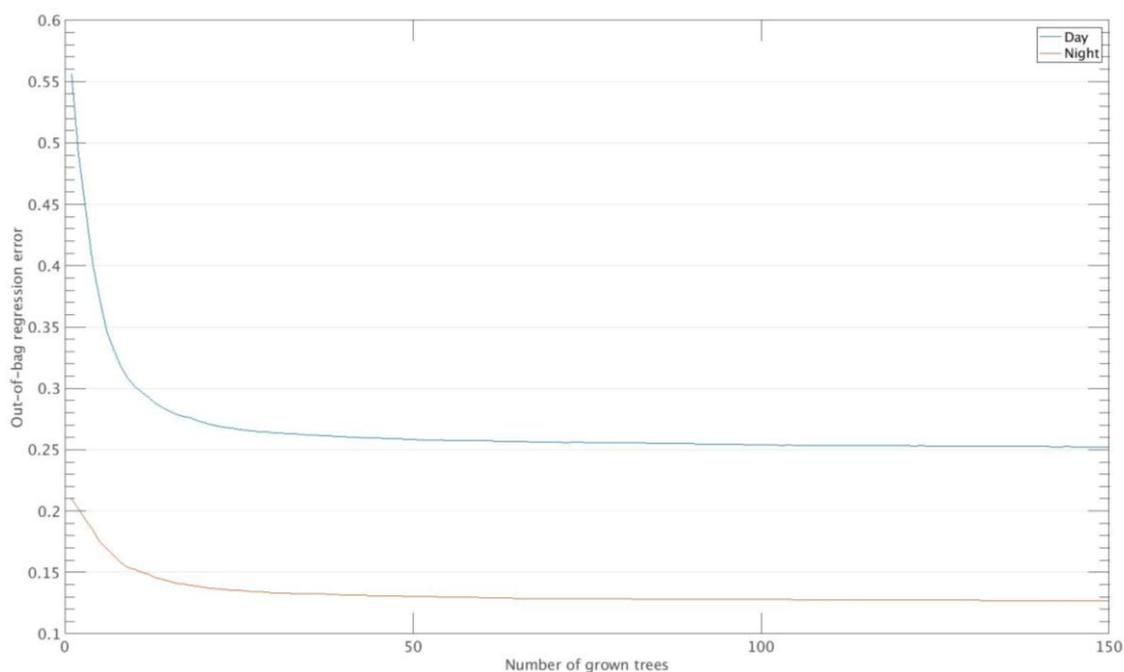
The training algorithm for random forests applies the general technique of bagging, where the key to bagging is that trees are repeatedly fit to bootstrapped subsets of the observations. The bootstrapped term refers here to the fact of choosing randomly data that can be chosen several times to build decision trees (no selection restrictions). A training dataset is chosen randomly with replacement (bootstrapping) from the original

dataset to create a decision tree. In regression techniques, the training dataset correspond to  $\sim 2/3$  of the total of the sample. The  $\sim 1/3$  remaining data not used to train that decision tree is used later as testing data but also to determine the importance of a specific term (James et al., 2013). This procedure is repeated for all the decision trees used in the random forest. Finally, it is not necessary to have a validation dataset in this study because the main interest of using this machine learning is to determine the importance of the terms on the model developed, as it is known that random forest protect against overfitting by constructing training samples through bootstrapping.

Some hyperparameters are tuned in order to optimize the analysis:

- The random forest method is set to have 150 decision trees, because at that number the error converges to a small value, as seen in the Figure 3-16 (converging value of 0.25 during daytime, 0.12 during nighttime).
- For the split criteria, since our model is a simple one-degree regression, the method is set to use the mean square error (MSE) to do the split at each leaf.

The number of random features to consider at each split and the number of bootstrapped dataset used to train each decision tree in the random forest method is approximatively to be  $2/3$  of the total of predictors and  $2/3$  of the total of sample, respectively (James et al., 2013).



**Figure 3-16:** Out-of-bag error over the number of grown regression trees for day (blue line) and night (orange line)

## CONCLUSION

A statistical analysis was performed in this chapter in order to estimate how well the OTEM representation follows the real observed temperature variations. After founding a good agreement between OTEM and the reality (correlation coefficient of 0.79), OTEM still encounters difficulties to reproduce extreme hourly temperature variations ( $\frac{\partial T_{2m}}{\partial t_{obs}} < -5 \text{ }^\circ\text{C/h}$ ), mainly because these (rare) event occurs within a smaller scale than the data resolution (e.g. cold pool event).

The mean contribution of each term is estimated in an annual and monthly-hourly cycle, since it is important to know to what extent they contribute to temperature variations during different hours of the day. Table 3-3 summarizes the contribution of each term, showing their minimum and maximum as well as their mean value for day and nighttime. From this chapter we conclude that:

- The radiative terms are on average the ones contributing the most to near-surface temperature variations for both day and nighttime.
  - o  $R_{CS}$  can both warm and cool the surface during daytime, but it is always negative at night (radiative cooling due to absence of SW solar radiation).
  - o  $R_{CL}$  has a similar behavior as  $R_{CS}$  warming and cooling during daytime but with a stronger cooling than warming effect, whereas at night it warms up to  $1.21 \text{ }^\circ\text{C/h}$  with a weak cooling effect.
- $HA$  mean contribution is almost negligible at night but it presents a wide range of contribution going from  $0.89$  to  $-1.03 \text{ }^\circ\text{C/h}$ . On the other hand, at daytime, it has a stronger cooling than warming effect, mostly due to the increase in the MLD in late afternoons.
- On average,  $HG$  and  $Adv$  mean contributions are close to zero at this scale, but  $Adv$  still can warm (cool) the surface up to  $1.57 \text{ }^\circ\text{C/h}$  ( $-1.53 \text{ }^\circ\text{C/h}$ ) during daytime, whereas at night it has a similar maximal contribution, but it can cool the surface up to  $-2.16 \text{ }^\circ\text{C/h}$ .

The use of a random forest method makes it possible to identify which term dominates over the others, depending on the period:

- On a diurnal scale, cloud radiation becomes the main contributor to temperature variation during daytime just after early mornings. Indeed, at this time, it is  $R_{CS}$  that controls temperature variations when surface heating occurs and no other (or very few) turbulence processes happen.
- For all the seasons on a diurnal scale, the late afternoon temperature variations are mostly driven by heat fluxes in the atmosphere when radiative contribution become weak and the MLD is at its maximal value.

- At night, cloud radiation is controlling temperature variations and on a diurnal scale contributing to warm the surface, as well as heat fluxes in ground that becomes the second most important term.
- In winter, spring and fall, advection of the air is the second most important contribution at night whereas in summer is clear-sky radiation.

**Table 3-3:** Mean, maximum and minimum value contribution of each term in OTEM split in day and nighttime.

Term	Daytime			Nighttime		
	Mean	Max	Min	Mean	Max	Min
RCS (°C/h)	1,33	3,97	-0,76	-0,95	-0,41	-1,47
RCL (°C/h)	-0,55	0,75	-3,72	0,42	1,21	-0,15
HA (°C/h)	-0,39	1,16	-2,56	-0,02	0,89	-1,03
HG (°C/h)	-0,05	0,54	-0,76	0,07	0,54	-0,60
Adv (°C/h)	-0,04	1,57	-1,53	-0,04	1,50	-2,16

Since clouds are found to be the main modulator of solar radiation on hourly temperature variations for all seasons, the next chapter studies the specific role of clouds and how they affect other atmospheric variables under different conditions.



## 4 CHAPTER 4: SPECIFIC INFLUENCE OF CLOUDS ON 2 M TEMPERATURE VARIATIONS UNDER CLOUDY CONDITIONS

---

### INTRODUCTION

Chapter 3 shows that clouds are the main modulator of solar radiation effect on the hourly temperature variations at the surface during the day (and the main contributor during the night). Knowing how each term of OTEM affects the hourly temperature variations, a deeper analysis is carried out in this fourth Chapter, in order to better understand the specific role of clouds under different conditions. This analysis is performed by considering other variables available in the SIRTA-ReOBS dataset to characterize both the atmosphere and the clouds, including lidar measurements to characterize the vertical profile of clouds.

### 4.1 CONTRIBUTION OF OTEM TERMS IN CLOUDY CONDITIONS

As previously stated in Chapter 3, the cloud radiative term controls on average hourly temperature variations for both day and nighttime, thus a special focus only in cloudy conditions is here addressed. Cloudy cases are then identified based on a criterion on the absolute value of CRE (Eq. 2.23) that must be higher than  $5 \text{ W/m}^2$  in both SW and LW domains (Chiriaco et al., 2018) during daytime, whereas for nighttime, only LW radiation is considered for this criterion, following:

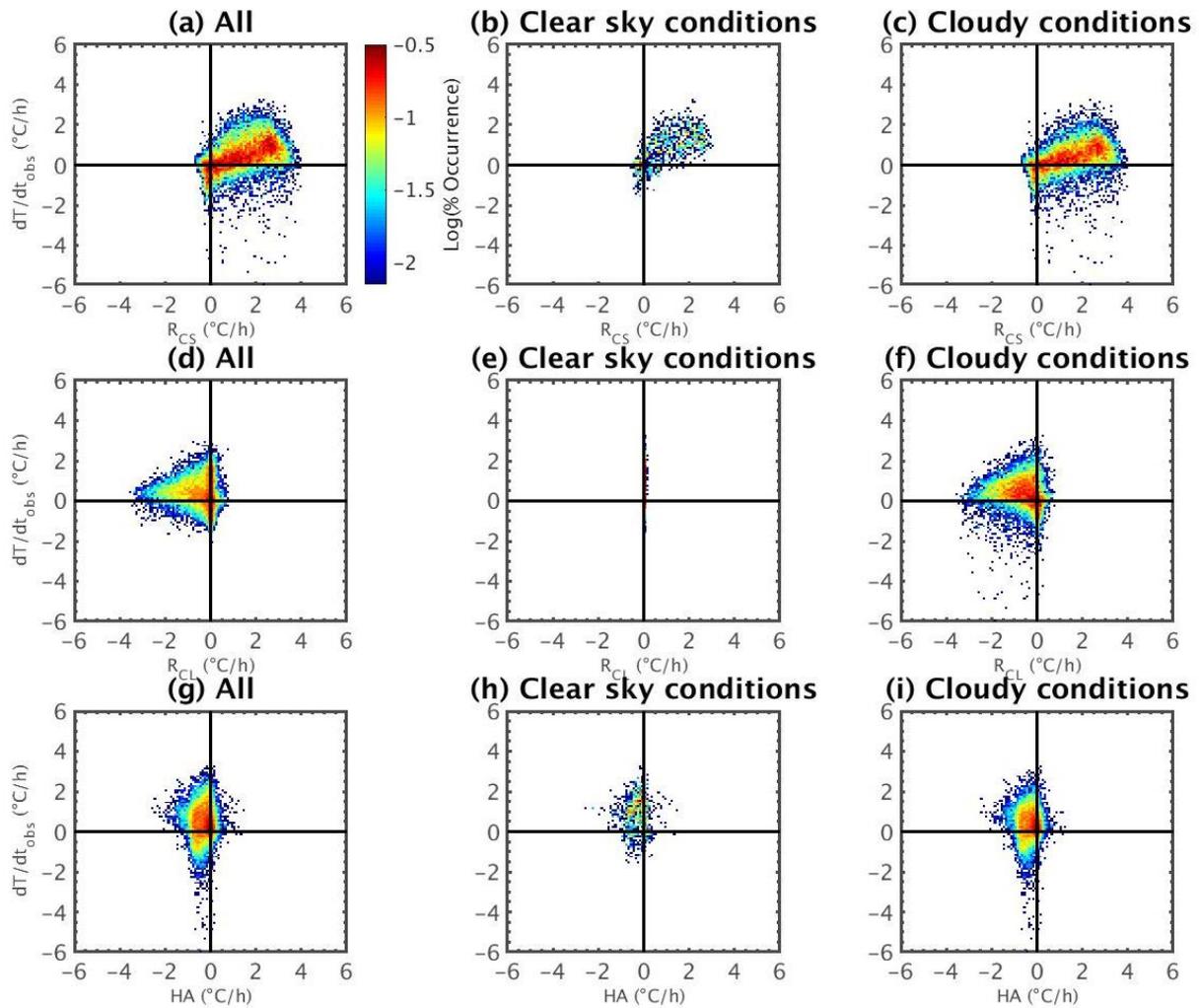
$$SWCRE = F_{SW}^{\downarrow} - F_{SW,CS}^{\downarrow} \quad (4.1)$$

$$LWCRE = F_{LW}^{\downarrow} - F_{LW,CS}^{\downarrow} \quad (4.2)$$

According to this criterion of classification, cloudy conditions correspond to 84% of the total cases from January 2009 to February 2014 at SIRTA, and clear-sky conditions represent the remaining 16%. To analyze the difference between clear-sky and cloudy conditions, Figure 4-1 shows scatter plots of occurrence of  $\frac{\partial T_{2m}}{\partial t}_{obs}$  as a function of the three main terms of OTEM ( $R_{CS}$ ,  $R_{CL}$  and  $HA$ ), during daytime separated into cloudy and clear-sky conditions. In Figure 4-1a for all daytime cases (first column), when both  $\frac{\partial T_{2m}}{\partial t}_{obs}$  and  $R_{CS}$  are positive, it is the radiative clear-sky term that guides the sign of hourly temperature variations. The main results from this figure are then:

- First column is very similar to third column, meaning that general behavior is dominated by cloudy cases.

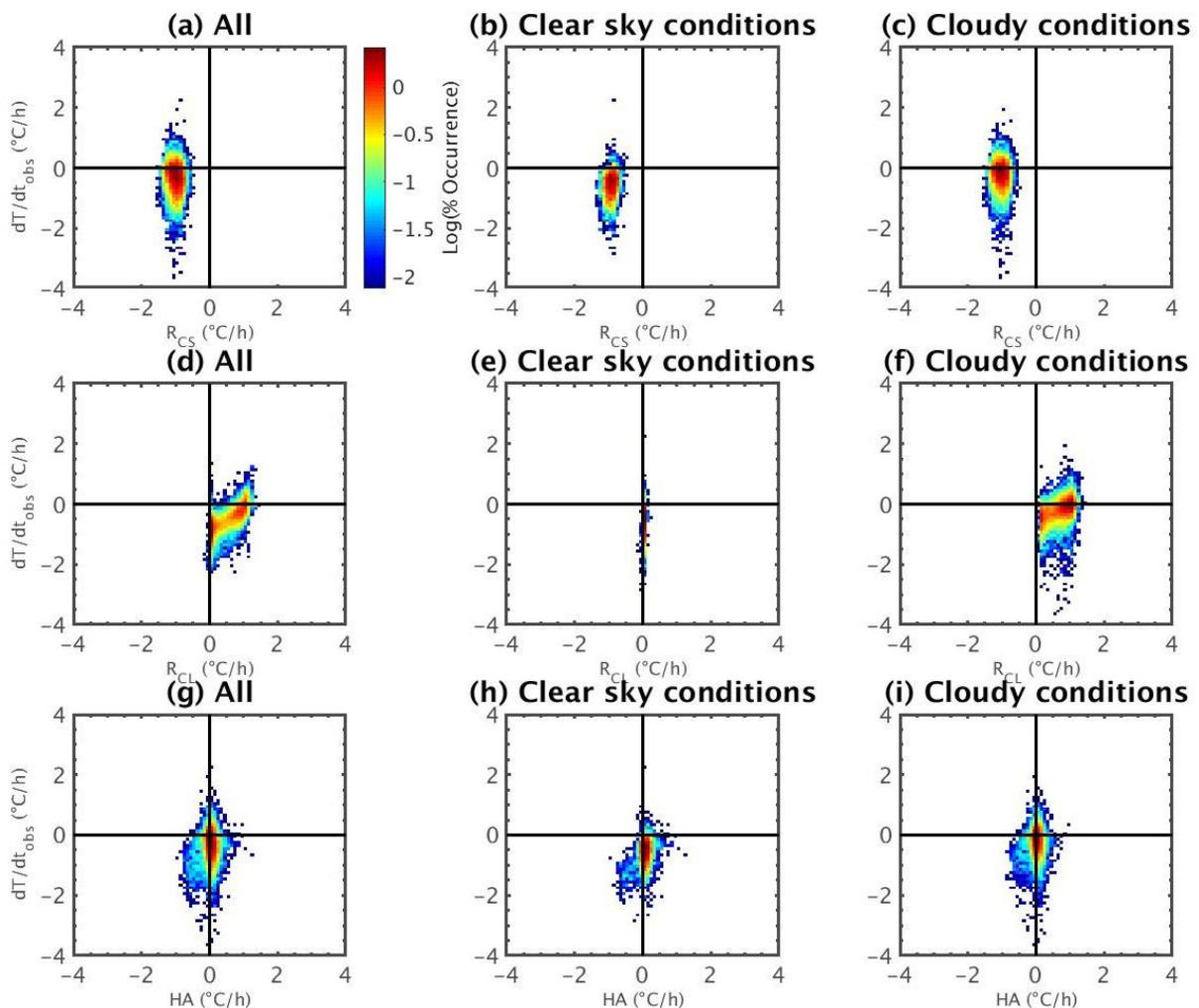
- Even when focusing on cloudy conditions (third column), positive hourly temperature variations are dominated by  $R_{CS}$  (Figure 4-1c).
- All the negative extremes of  $\frac{\partial T_{2m}}{\partial t}_{obs}$  are detected only in cloudy conditions.
- Logically,  $R_{CL}$  has negligible values in clear-sky conditions whereas all its contribution corresponds to a cloudy atmosphere (Figure 4-1e and Figure 4-1g, respectively) with a maximum cooling effect around -4 °C/h.
- HA term contributes mostly to cool the surface when  $\frac{\partial T_{2m}}{\partial t}_{obs}$  is positive (Figure 4-1h), damping the important positive contribution of  $R_{CS}$  (Figure 4-1i).



**Figure 4-1:** Scatter plot of the occurrence (in  $\text{Log}_{10}$ ) of daytime  $\frac{\partial T_{2m}}{\partial t}_{obs}$  as a function of  $R_{CS}$  (first row),  $R_{CL}$  (second row) and HA (third row) terms for all-sky (first column), clear-sky (second column) and cloudy (third column) conditions.

During nighttime, a similar procedure is addressed than for daytime (Figure 4-2). This time, cloudy conditions are less occurrent than during daytime and correspond to 70% of the total cases for the period of study, and clear-sky conditions represent the other 30%. The main characteristics of each term for this time are listed below:

- $R_{CS}$  contribution to temperature variations presents almost no difference between a clear-sky (second column) and cloudy atmosphere (third column) since it remains almost constant for these two conditions (Figure 4-2b and Figure 4-2c).
- As expected  $R_{CL}$  contribution is presented almost exclusively in cloudy conditions, warming the surface up to 1.5 °C/h (Figure 4-2f).
- $HA$  tends more to warm the surface during clear-sky conditions than during cloudy cases when it contributes mostly to cooling it (Figure 4-2h and Figure 4-2i).
- Overall, clouds modulate the negative contribution of  $R_{CS}$  and  $HA$  during nighttime (Figure 4-2d, Figure 4-2a and Figure 4-2g, respectively).



**Figure 4-2:** Same as Figure 4-1 but for nighttime.

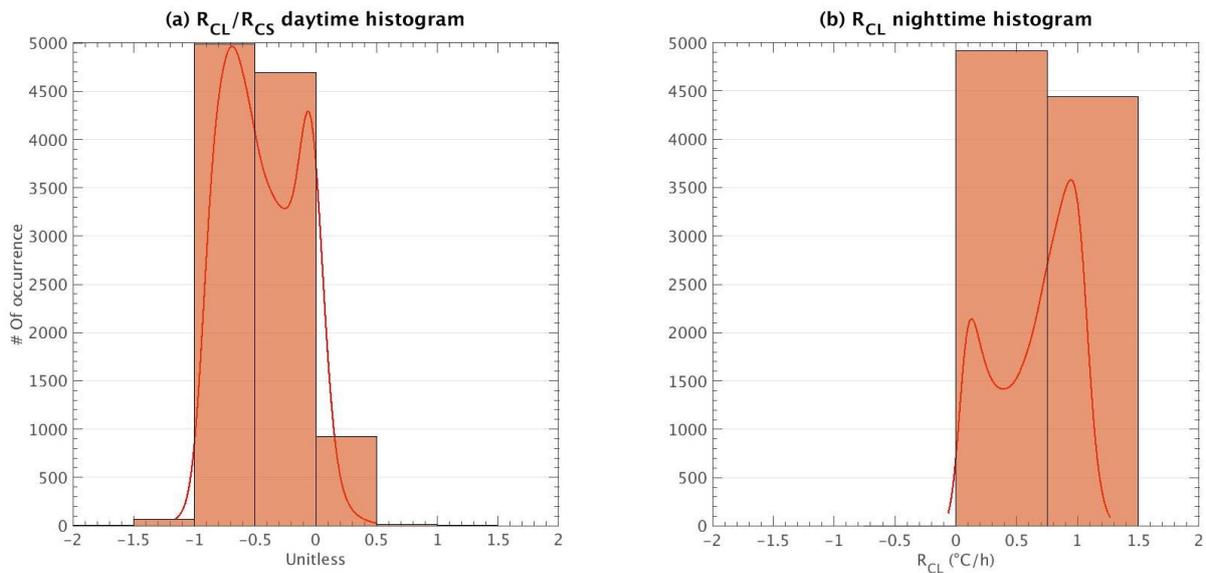
## 4.2 CREATION OF CLOUDS CATEGORIES

The study of the main terms of OTEM separated in clear-sky and cloudy conditions showed a more important (and notable) contribution of these terms to hourly temperature variations when the atmosphere is cloudy. Due to the wide range of contribution of  $R_{CL}$  for day and nighttime cases (Figure 4-1f and Figure 4-2f, respectively), some categories are created based on this contribution.

### 4.2.1 Cloud categories based on their radiative effect

As shown in Section 3.3.1 and Figure 3-13a, the  $R_{CS}$  term dominates hourly temperature variations during daytime. During the day the ratio of  $R_{CL}$  by  $R_{CS}$  is estimated to assess to what extent clouds' contributions are driven by the clear-sky term. Figure 4-3a shows the distribution of this ratio for daytime hours. Two peaks can be easily identified: one between -0.5 and -1 and one slightly negative with a tail in positive values. From this distribution, three different bins can be created in order to separate the effect of clouds on observed surface temperature variations, going from the highest cooling effect to the warming effect (rectangular brown boxes in Figure 4-3a): (i)  $-1.0 \leq \frac{R_{CL}}{R_{CS}} < -0.5$  (bin 1), (ii)  $-0.5 \leq \frac{R_{CL}}{R_{CS}} < 0$  (bin 2) and (iii)  $0 < \frac{R_{CL}}{R_{CS}} \leq 0.5$  (bin 3). The cases with very close-to-zero values of  $R_{CS}$  (10% of the cases), which occurred in early mornings and late afternoons, are excluded since they give divergent values. Same thing for very rare negative value of  $R_{CS}$ , as they affect the sign. Furthermore, most values are negative. Very few values (less than 100 values) are found for  $R_{CL} > R_{CS}$  (i.e.  $R_{CL}/R_{CS} < -1$ ) so they are not enough to allow us to perform a further and complete seasonal analysis.

For nighttime cases,  $R_{CL}$  is the term dominating and controlling temperature variations (Section 3.3.1. and Figure 3-13b) and since  $R_{CS}$  is on average constant at this time (Figure 3-10b), it is not necessary to divide it by  $R_{CS}$  as done previously for daytime hours. Therefore, two bins are created from the PDF of  $R_{CL}$  shown in Figure 4-3b where two peaks are spotted when  $R_{CL} > 0$ :  $0 < R_{CL} < 0.75$  °C/h, a bin when cloud warming effect is said to be weak, and  $0.75 \leq R_{CL} \leq 1.5$  °C/h, a bin when cloud warming effect is classified as strong. During nighttime there is no SW radiation so clouds always have a warming effect on surface temperature and no negative values are found in Figure 4-3b.



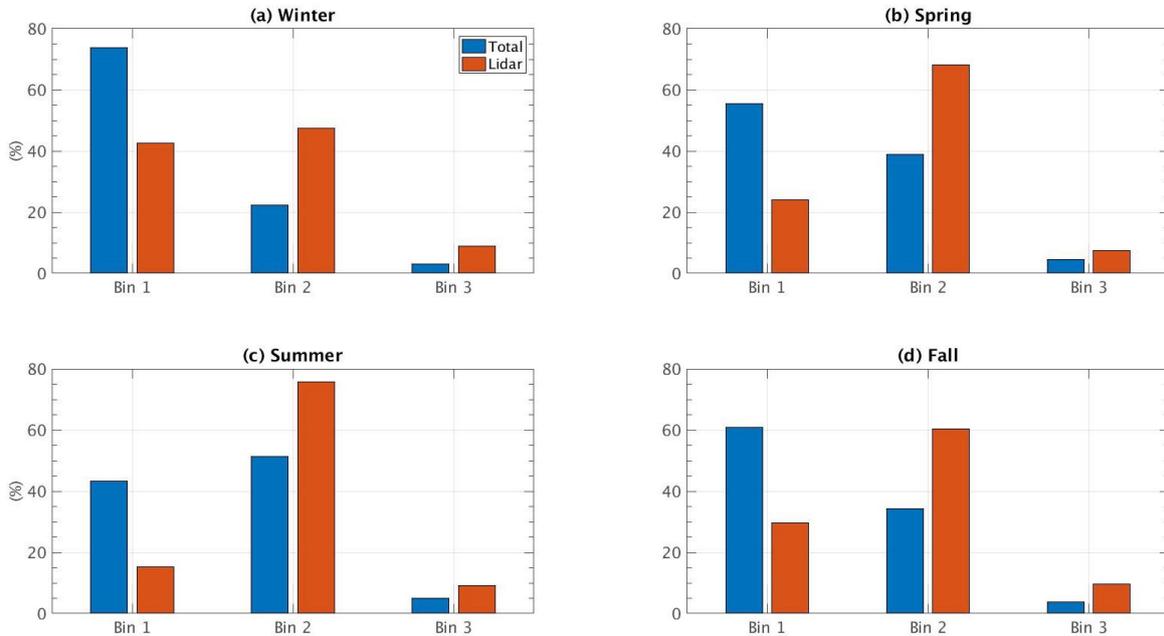
**Figure 4-3:** (a) Daytime histogram of  $\frac{R_{CL}}{R_{CS}}$  and (b) nighttime histogram of  $R_{CL}$ . The red line in both figures represents the PDF and the rectangular semi-transparent brown boxes the different bins created to analyze clouds influence. These histograms are built by considering only cloudy hours. Negative and close-to-zero values are removed for daytime hours (see text for further information).

#### 4.2.2 Specific lidar sampling for daytime cases

In the following sections, the database that regroups cloudy daytime cases (i.e. the hours in which SWCRE and LWCRE are greater than  $5 \text{ W/m}^2$ ) is called the “total cloudy sampling”. From this sampling, the “lidar sampling is created”, corresponding to cases with lidar measurements (the lidar instrument does not operate all the time, cf. Section 2.1.1.2.). This last sampling is used to study the vertical structure of the troposphere (from to the ground up to 15 km). Figure 4-4 presents the repartition from one bin to another for these two samplings, distributed by seasons. These histograms show that the patterns are different for the lidar sampling comparing to the total one: bin 2 is preponderant whatever the season, whereas it is preponderant only in summer for the total sampling. In winter, spring and fall, the majority of cases are in bin 1 for the total sampling. Finally, bin 3 represents only a few cases for the two samplings.

The difference found between the total cloudy and lidar samplings can be explained by the fact that the lidar does not operate when it is raining, hence the lidar sampling does not represent the same repartition of the different type of clouds. To further understand that, the hourly precipitation rates are considered in Figure 4-5. The percentage of occurrence is calculated as the ratio between non-zero precipitation rates cloudy cases and the total daytime cloudy cases. From this figure, both the occurrence of precipitation and the mean precipitation rates are the highest for bin 1

(blue histograms) and then bin 2 (green histograms) at all seasons. Indeed, these mean precipitations rates for bin 1 (Figure 4-5b) are higher than the mean precipitation rates for the period of interest (which is 0.08 mm/h from 2009 to 2014), meaning that this bin 1 is the most affected by this sampling differentiation on cloudy and lidar cases.

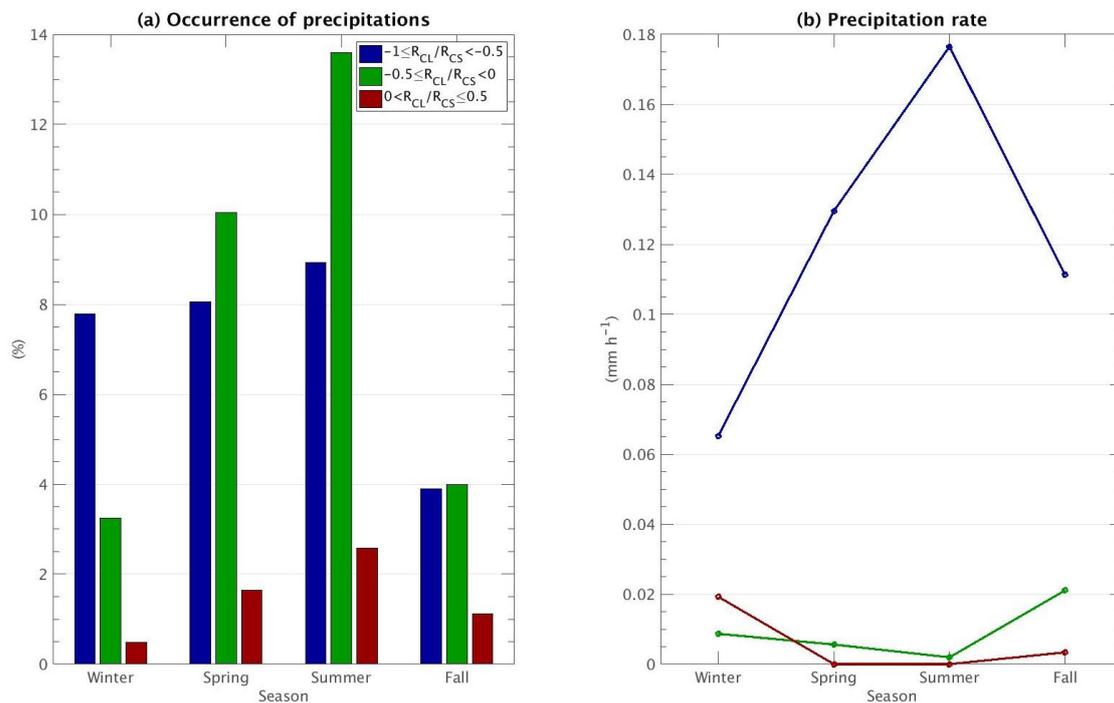


**Figure 4-4:** Histograms of total cloudy and lidar percentage sampling availability for each bin created, with respect to the total daytime cloudy cases of each season.

When focusing only in the lidar sampling availability from the total cloudy sampling for each season (all bins combined), lidar sampling availability is 8% for winter, 12% for spring, 20% for summer and 14% for fall. Furthermore, the lidar sampling is available 22% of the time with respect to the total daytime cloudy cases, all seasons and bins combined.

### 4.3 DAYTIME VERTICAL LOW-ATMOSPHERE STRUCTURE AND CHARACTERIZATION UNDER DIFFERENT CLOUD EFFECTS

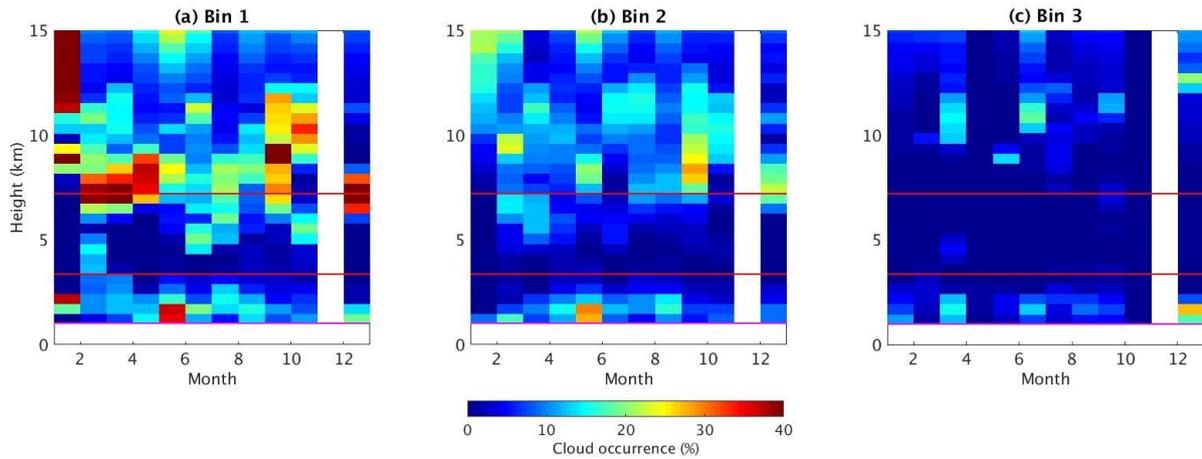
In this section, the main objective is to characterize the atmospheric conditions by evaluating different meteorological variables in order to assess how clouds affect them, as a function of their radiative effects (i.e. separated in the three bins previously created). Lidar profiles are used to characterize the type of clouds detected for the different bins.



**Figure 4-5:** Cloudy daytime (a) percentage occurrence of precipitations and (b) mean precipitation rates for each bin split into seasons.

### 4.3.1 Seasonal cycle of cloud occurrence

Figure 4-6 shows the seasonal cycle of cloud occurrence estimated from the SR values measured by the lidar instrument. This percentage of occurrence of clouds is calculated at each level as the number of cloudy profiles ( $SR(z) \geq 5$ ) over the number of profiles with  $SR(z) \geq 0.01$ . The red horizontal lines on each figure correspond to low- ( $P > 680$  hPa), mid- ( $440 < P < 680$  hPa), and high-level ( $P \geq 440$  hPa) clouds limits (Chepfer et al. 2010; Chiriaco et al. 2018). These lidar profiles correspond to non-precipitant cases. Cloud occurrence is higher for cases with important cloud cooling radiative effect (bin 1, Figure 4-6a). Indeed, high-level clouds are more frequent for bin 1 in winter than for the other months, whereas low-level clouds are in general more present in spring and summer. The majority of mid-level clouds are spotted for bin 1, with only a few for bin 2 and almost no presence of these clouds for bin 3. Actually, for bin 3 the occurrence of clouds is very low despite the fact that for reminder here we only consider cloudy cases ( $SWCRE$  and  $LWCRE > 5 \text{ W/m}^2$ ). This quasi-absence of clouds enhances a positive radiative effect, coupled with a presence of high-level clouds for some months (e.g. March and August) and the absence of very thick opaque mid-level clouds. Note that in November no lidar data were available for any bin, winter being the season with less amount of lidar data available from the total cloudy sampling (not shown).

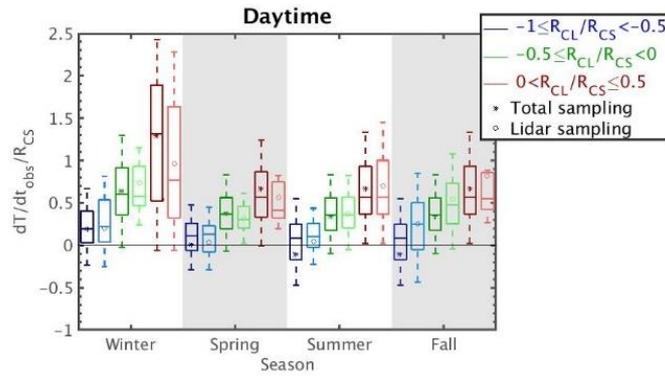


**Figure 4-6:** Seasonal cycle of occurrence of clouds estimated from SR values for (a) bin 1:  $-1.0 \leq \frac{R_{CL}}{R_{CS}} < -0.5$  (b) bin 2:  $-0.5 \leq \frac{R_{CL}}{R_{CS}} < 0$  , and (c) bin 3:  $0 < \frac{R_{CL}}{R_{CS}} \leq 0.5$ . The occurrence is defined as the number of cloudy profiles ( $SR(z) \geq 5$ ) over the number of profiles for which  $SR(z) \geq 0.01$ . November appears as a white column because there were no data at this period.

#### 4.3.2 Clouds with strong cooling effect

Since the occurrence of clouds is different for each bin and each month, it is important to analyze clouds' influence on hourly temperature variations separating it by seasons, and then study their influence on different meteorological variables. To do so, first  $\frac{\partial T_{2m}}{\partial t}_{obs}$  is divided by  $R_{CS}$  for each bin in order to estimate how much hourly temperature variations are driven by the solar radiation: when this ratio is close to 1, it means that the variations of temperature are the ones that would be expected in clear-sky conditions with no other modulators, and when this ratio deviates from 1 it means that the temperature variations are either damped or enhanced by the other terms.

Figure 4-7 shows the values of this ratio and Figure 4-8 shows the distribution of different relevant meteorological observations (represented as box-and-whiskers) available in the SIRTA-ReOBS dataset, for each bin created, split into seasons, for both the total cloudy and the lidar samplings. In addition, lidar  $SR(z)$  histograms (cf. Eq. 2.1) presented in Figure 4-9 are estimated by cumulating all lidar  $SR(z)$  observations available for one bin and one season in particular. The following analysis consists then on characterize some standard meteorological variables (e.g. temperature, RH,  $F_{sens}$ , etc.) as a function of the radiative effect of clouds, whether this effect is strong, weak, negative or positive.



**Figure 4-7:**  $\frac{\partial T_{2m}/\partial t_{obs}}{R_{CS}}$  daytime values for the four seasons for three  $R_{CL}/R_{CS}$  bins:  $-1.0 \leq \frac{R_{CL}}{R_{CS}} < -0.5$  in blue,  $-0.5 \leq \frac{R_{CL}}{R_{CS}} < 0$  in green and  $0 < \frac{R_{CL}}{R_{CS}} \leq 0.5$  in red. Dark-colors box plots with the mean represented as '\*' correspond to cases when all meteorological variables are available at the same time (without considering lidar availability), whereas light-colors box plots with 'o' as their mean value represent the sample when both all meteorological variables and Lidar profiles are available simultaneously. Distributions are represented by box-and-whiskers plots, where the boxes indicate the 25<sup>th</sup> and the 75<sup>th</sup> data percentiles, the whiskers indicate 5<sup>th</sup> and 95<sup>th</sup> percentiles, the middle line represents the median, and the \* or ° is the mean.

When clouds have an important cooling radiative effect (which is the case for bin 1), Figure 4-7a shows that the hourly temperature variations (normalized by  $R_{CS}$ ) are the lowest at all the seasons. In winter this ratio stays mostly positive (temperature increase in one hour) but less than 0.5 (less than half the increase that could be reached if the sky was cloudless) and it can become negative during summer and fall, i.e. the warming induced by  $R_{CS}$  can be counterbalanced by the clouds in these seasons. Besides, the cloud cover is almost total and the radiative effects are strong, as seen in Figure 4-8d, Figure 4-8e and Figure 4-8f. In addition, the seasonal variability of SWCRE is directly related to the seasonal variability of  $F_{SW,CS}^{\downarrow}$ , and not (or only slightly) to the seasonal variability of cloud properties, since the ratio  $\frac{SWCRE}{F_{SW,CS}^{\downarrow}}$  remains almost constant during the entire year (not shown) and cloud cover presents the highest mean values among the different bins (Figure 4-8d).

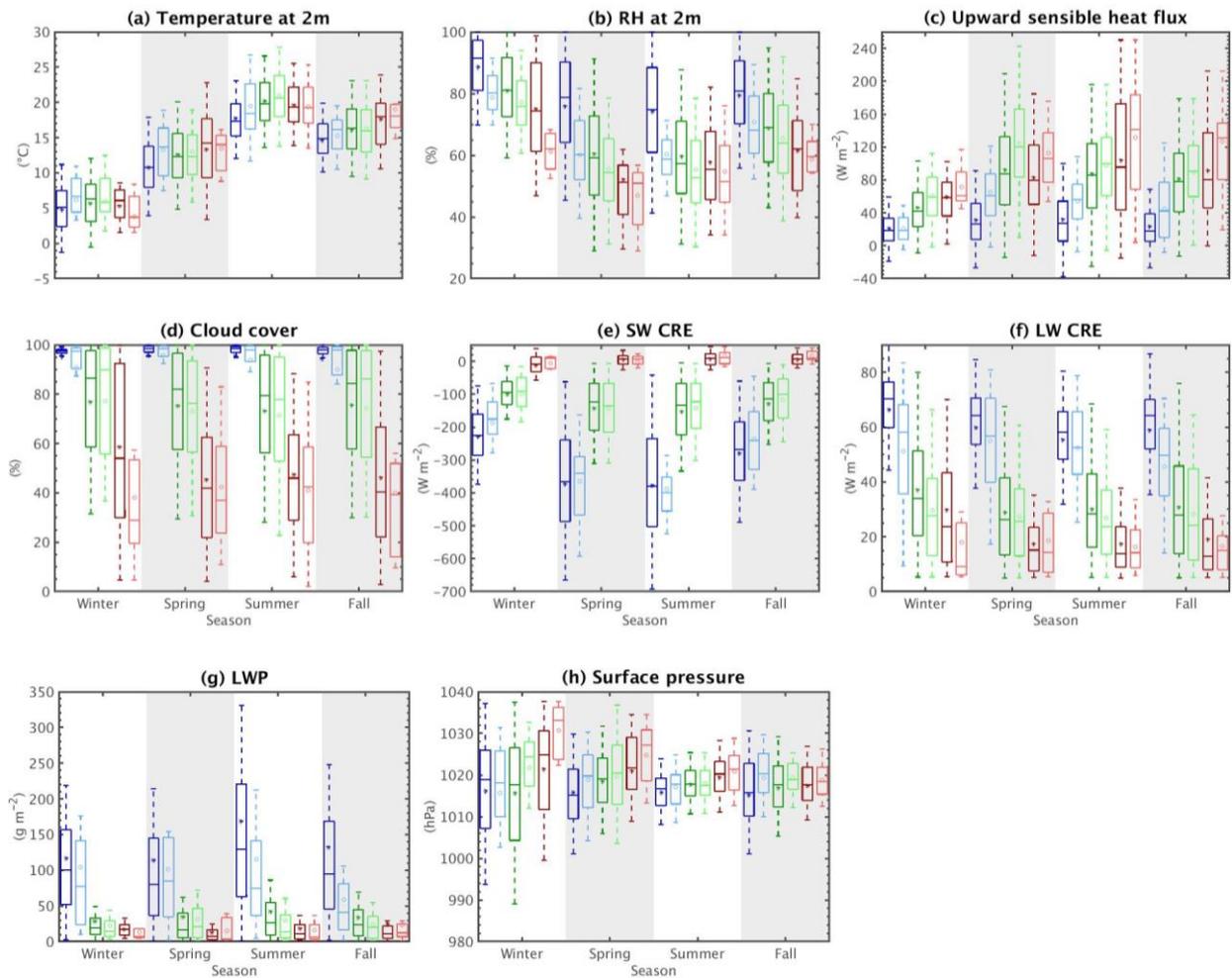
The light blue bin corresponds to lidar sampling, that are only non-rainy cases. Indeed, lower RH values (Figure 4-8b), higher amount of  $F_{sens}$  (Figure 4-8c), higher (i.e. less negative) values of SWCRE (Figure 4-8e) and both lower liquid water path (LWP) and surface pressure (Figure 4-8g and Figure 4-8f, respectively) values are found for this sampling than for the total cloudy one. Furthermore, minimum surface pressure values correspond always to the total cloudy sampling for all the tree bins at all the seasons, whereas the lidar sampling exhibits fearer weather conditions since precipitations hours are not considered here. Logically, lower LWCRE is also found for the lidar sampling. Concerning the precipitations rates shown in Figure 4-5b for this bin 1,

summer is the season with the highest mean rate which is seen in the high values of LWP.

As expected, temperatures are higher for the lidar sampling (i.e. when it is not raining) at all the seasons except summer, as shown in Figure 4-8a, along with more cases with negative temperature variation (Figure 4-7a). For this sampling that excludes rainy cases, both higher and lower values of SWCRE are found for spring rather than for summer (Figure 4-8e): the minimum value is about  $-600 \text{ W/m}^2$  for spring whereas for summer it is about  $-500 \text{ W/m}^2$  (a situation not captured by the total sampling where summer has a lower value of SWCRE than spring). Figure 4-9b (where the  $SR(z)$  histograms are presented, cf. Section 2.1.1.2.) shows that low- and high-level clouds are more frequent for spring than for summer (Figure 4-9c) and thus stronger negative values of SWCRE are found for spring. Then, the important presence of mid-level clouds with a high value of lidar  $SR(z)$  ( $> 80$ ) spotted in summer in Figure 4-9c is potentially at the origin of the strong negative values of SWCRE despite the very low presence of other clouds. In addition, despite the strong negative values of SWCRE for this bin for all seasons, surface temperatures are not so low (Figure 4-8a), partly due to the high LWCRE values (Figure 4-8f) that slightly dampen the strong SWCRE.

In addition, some differences in cloud presence can be detected between Figure 4-9a, Figure 4-9b, Figure 4-9c and Figure 4-9d for the first bin. For the fall season, the  $SR(z)$  histogram in Figure 4-9d exhibits an important presence of high-level optically thin clouds (especially above 8 km,  $5 < SR(z) < 15$ ), along with mid-level optically thick clouds ( $SR(z) > 80$ ). Figure 4-9b also shows an important presence of high-level clouds in spring but within a smaller vertical range ( $7 < z < 10 \text{ km}$ ) compared to fall. Indeed, based on both the  $SR(z)$  value and height, these high-level clouds could correspond to cirrostratus or cirrocumulus which form when a mass of warm air meets a mass of cold air, where the lighter warm air rises and could form these cirrus clouds. In addition, spring and summer exhibit high amounts of low-level thicker clouds ( $SR(z) > 40$ ) that are not detected in the other two seasons of the year, which could correspond to the moments just before a rainstorm is set. Indeed, the highest amount of occurrence of precipitations and mean precipitation rates are found for these two seasons as seen in Figure 4-5, and the mean cloud cover in Figure 4-8d for the lidar sample is around 95% for both seasons and less clear-sky conditions are found ( $0.01 < SR(z) < 1.2$ ) (Figure 4-9b and Figure 4-9c).

To conclude this first case, clouds with a strong radiative effect are associated mostly to rainy hours with temperature variations that can both decrease and increase, with a moist atmosphere and surface turbulent fluxes that are very low. As for the clouds detected by the lidar, this radiative effect could be associated mostly to optically thick mid-level clouds for all the seasons, especially in summer.



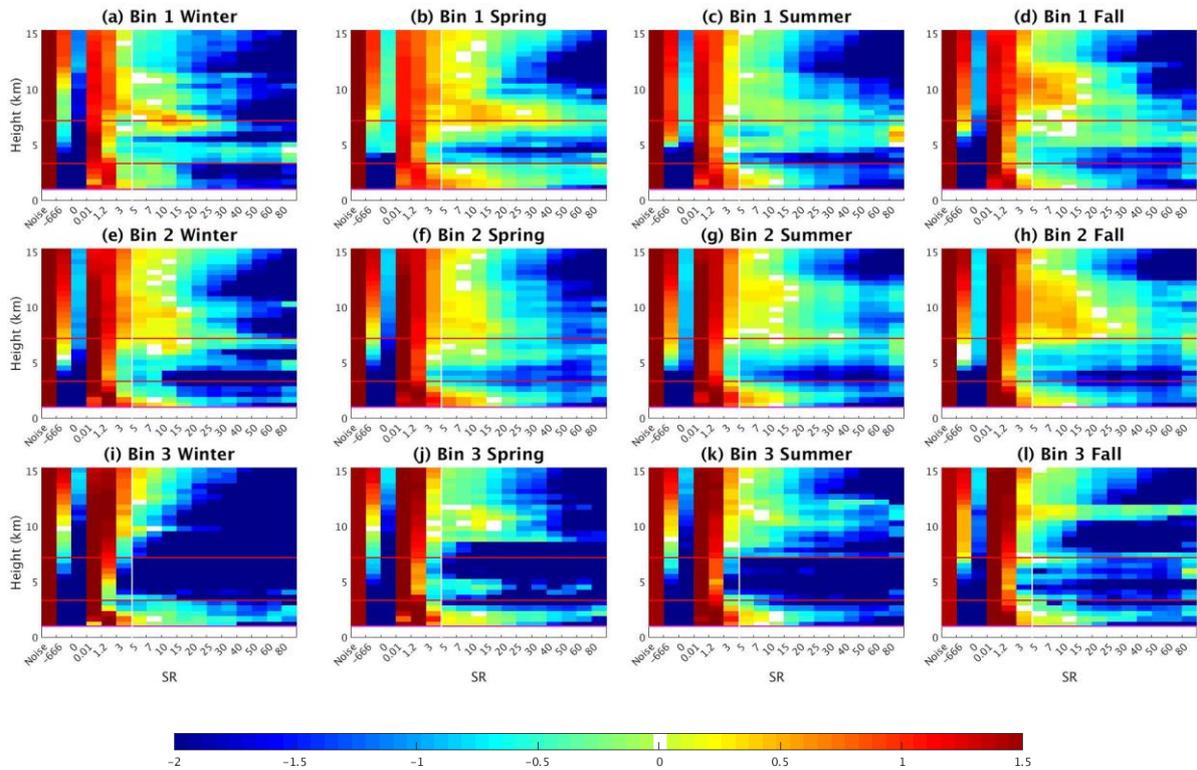
**Figure 4-8:** Daytime values for (a) temperature, (b) relative humidity, (c) upward sensible heat flux, (d) cloud cover retrieved from a sky imager, (e) shortwave and (f) longwave cloud radiative effect, (g) liquid water path and (h) surface pressure. Colors and boxplots follow the same definition as in Figure 4-7.

### 4.3.3 Clouds with weak radiative effect: cooling or warming

When  $R_{CS}$  becomes dominant with respect to  $R_{CL}$  and clouds present a weak radiative effect (bins 2 and 3), temperature most of the time increases in the hour, especially in winter (Figure 4-7a). In comparison with the first bin, the air becomes drier (lower RH, higher sensitive heat flux, Figure 4-8b and Figure 4-8c), which agrees with less mean cloud cover (Figure 4-8d), lower LWCRE and less negative SWCRE (Figure 4-8e and Figure 4-8f), and lower LWP (Figure 4-8g).

Furthermore, the occurrence of precipitations is higher when clouds have a weak cooling effect on surface temperature variations than when their effect is strong, as shown in Figure 4-5a, but their precipitation rate remains smaller for the first case. Light precipitation is then sometimes present for bin 2, corresponding to clouds optically

thinner than those in bin 1, which allow more solar radiation to reach the surface and thus SWCRE maximum values do not exceed  $-350 \text{ W/m}^2$  for cases with a weak cloud cooling effect.



**Figure 4-9:** Lidar scattering ratio ( $SR(z)$ ) histogram obtained by cumulating all lidar observations during daytime for bin 1 (a)–(d), bin 2 (e)–(h) and bin 3 (i)–(l) for winter (first column), spring (second column), summer (third column) and fall (fourth column). The color bar is the logarithm of the percentage of occurrence (the sum of one level is equal to  $\log_{10} 100 \%$ ); lidar data showed in each subplot start above the instrument’s recovery altitude ( $z = 1 \text{ km}$ ); the red horizontal lines represent the limits of low-, mid- and high-level clouds; and the white vertical line shows the threshold of clouds detection ( $SR(z) = 5$ ). A single “Noise bin” is presented here, which is simply the sum of the -999 and -777 bins that correspond to noisy and non-normalized profiles (cf. Section 2.1.1.2.).

Lidar histograms in Figure 4-9 show fewer differences between bin 1 and bin 2 ( $R_{CL} < 0$ ) than between bin 2 and bin 3 ( $R_{CL} > 0$ ) because this figure only considers non-rainy situations and in Section 4.2.2. it is noted the important differences between the two samplings (total cloudy and lidar, cf. Figure 4-4) for bin 1 for most of the variables. Furthermore, for the lidar sampling, LWCRE in winter and fall is slightly higher than for the two other seasons for the second bin (Figure 4-8f). This is partly due to the presence of high-level thin clouds ( $5 > SR(z) > 20$ ), such as cirrus, detected in the histograms of  $SR(z)$  for this bin in Figure 4-9e and h for these two seasons, which is slightly more than for spring and summer.

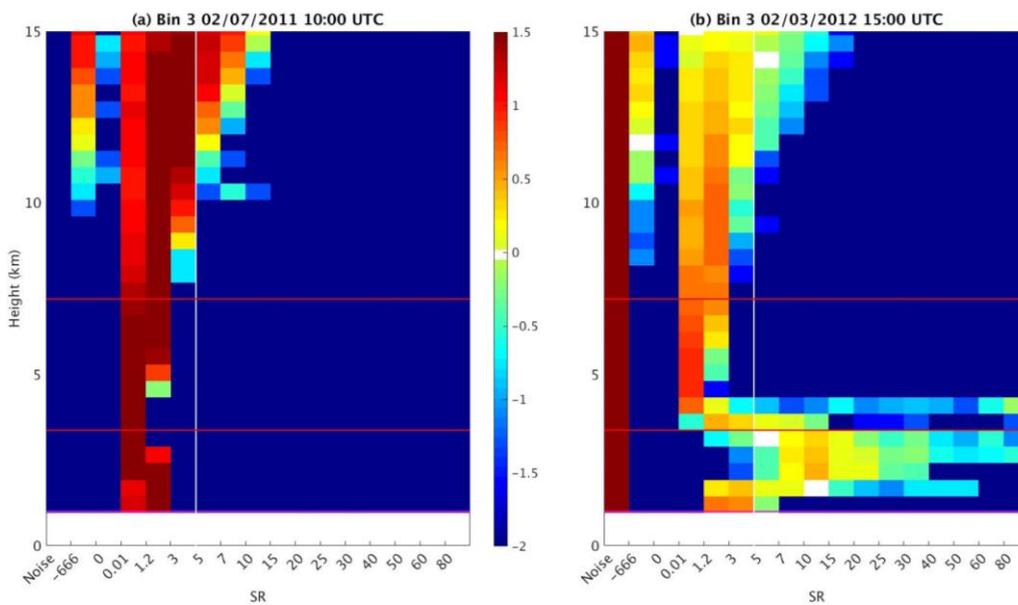
When clouds have a warming (but still weak) effect on surface temperature variations, precipitations are very low and almost undetectable for spring, summer and fall (red lines in Figure 4-5b). This leads to have very low mean values of RH not exceeding 60% and have fair weather surface pressure with minimum values not decreasing below 1010 hPa for the cloudy sampling as exposed in Figure 4-8b and Figure 4-8h, respectively. On the other hand, the mean precipitation rate in winter is considerably higher with respect to the other seasons (Figure 4-5b), and even higher than for bin 2, leading to very low-pressure values and a high percentage of relative humidity for this season.

In bin 3, the absence of mid-level clouds and the lower occurrence of high clouds compared to bin 2 is obvious, while the difference for low-level clouds is not clear (Figure 4-9e to Figure 4-9l). This is consistent with lower LWCRE (Figure 4-8f) but not necessarily with the very low values (i.e. close to zero) of SWCRE observed for bin 3 (Figure 4-8e). Note that the cloud cover retrieved from the sky imager can drop by up to a minimum of 5% (Figure 4-8d) and thus a lower amount of clouds is detected by the SR(z) histograms (Figure 4-9i to Figure 4-9l) for the lidar sampling than for the total one, and yet SWCRE is sometimes positive. These SWCRE values close to zero or greater than zero are explained by the fact that most of the hours corresponding to this bin 3 coincide just after sunrise (not shown), when solar radiation is still weak, thus explaining that  $F_{SW}^{\downarrow} \approx F_{SW,CS}^{\downarrow}$ . The SWCRE positives values reveal that  $F_{SW}^{\downarrow}$  is larger than  $F_{SW,CS}^{\downarrow}$ , an effect observed when the direct part of solar radiation is not fully attenuated and when some diffuse radiation reaches the surface in addition to the direct flux, radiation which is strengthened by scattering processes in clouds and atmospheric particles. This effect thus increases radiation beyond a respective clear-sky scenario, and it mostly happens in winter (but could occur the rest of the year) in early mornings and/or late afternoons hours when the solar elevation angle is weak (Stapf et al., 2020; Wendisch et al., 2019). This distribution towards early morning hours might explain that despite the high values of the ratio between temperature variations and  $R_{CS}$  (Figure 4-7a) and lower cloud cover and cloud effects, surface temperatures are not always maximal for this bin (these maximal temperatures happen only for spring and fall), as seen in Figure 4-8a.

To further investigate the cases and the type of clouds detected when CRE is positive, two different cases are here considered when these conditions appear and are shown in Figure 4-10. Figure 4-10a presents the SR(z) histogram of the 7<sup>th</sup> of February 2011 at 10:00 UTC for this bin 3, which presents very low – and positive – values of SWCRE = 12 W/m<sup>2</sup> and LWCRE = 6.24 W/m<sup>2</sup>, values explained by the complete absence of low- and mid-level clouds, but the presence of some thin cirrus ( $5 < SR(z) < 10$ ) combined with the weak SZA at that time of the day lead to having  $F_{SW}^{\downarrow} > F_{SW,CS}^{\downarrow}$  and hence a positive SWCRE. Secondly, Figure 4-10b shows another case corresponding to the 3<sup>rd</sup> of February 2012 at 15:00 UTC, for which SWCRE = -57 W/m<sup>2</sup> but LWCRE = 59 W/m<sup>2</sup> leading to have  $R_{CL} > 0$ , despite the presence of low-level clouds. Indeed, at this hour (13:00 LT) for February, solar radiation is approaching to very low values and the LW

radiation emitted by these low-level clouds becomes greater and they tend to warm the surface, a reason that might explain why  $R_{CL}$  is on average positive just before sunset not only for this month but for the rest of the year (see Figure 3-11e).

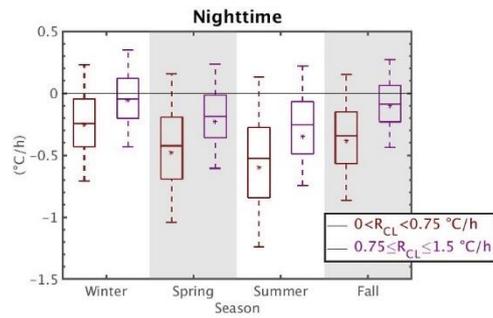
The atmosphere when clouds have a weak cooling effect presents a low cloud cover and temperature variations that are mostly positive all the time, and because SWCRE decreases, the clear-sky radiation tends to dominate these variations, with mostly low- and high-level clouds which are optically thin. As for the clouds warming the surface, the atmosphere presents a very low occurrence and temperature variations are the biggest among the other cases created, but the lidar does capture mostly optically thin cloud at low- and high-levels, especially in summer and fall.



**Figure 4-10:** Same as Figure 4-9 but for bin 3 ( $0 < \frac{R_{CL}}{R_{CS}} \leq 0.5$ ) for (a) 02/07/2011 at 10:00 UTC and (b) 02/03/2012 at 15:00 UTC.

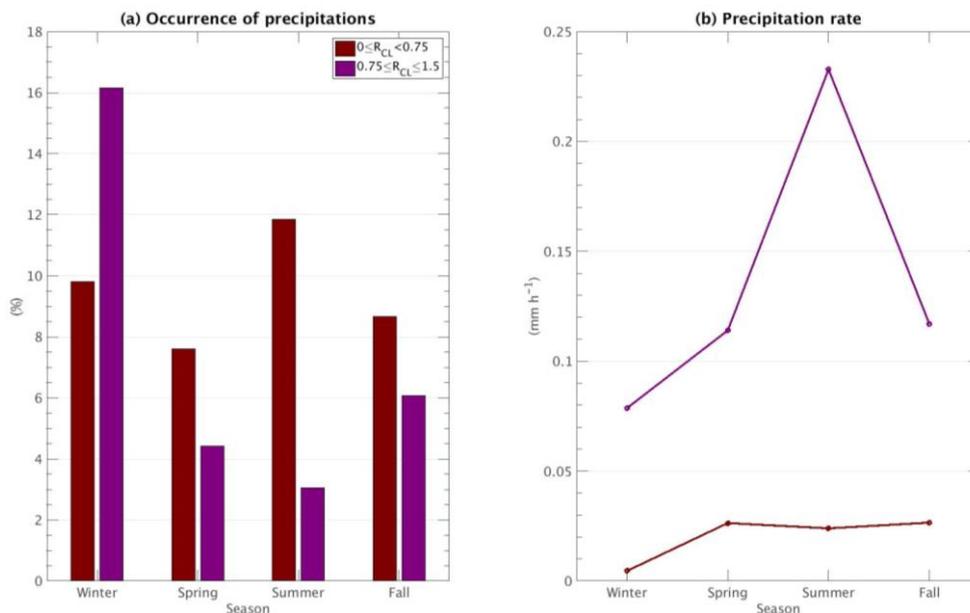
#### 4.4 NIGHTTIME LOW-ATMOSPHERE CHARACTERIZATION UNDER POSITIVE CLOUD EFFECTS

During nighttime, a similar procedure is addressed as for daytime to analyze the clouds' influence on different meteorological variables. Since during night hours the lidar instrument does not operate, the analysis focuses only in total cloudy sampling. The distribution of  $\frac{\partial T_{2m}}{\partial t}_{obs}$  for the two bins created for all the seasons is presented in Figure 4-11, Figure 4-12 is the same as Figure 4-5 but for the two bins created for the nighttime study, and Figure 4-13 shows the same meteorological variables as in Figure 4-8. Neither cloud cover nor cloud fraction profiles are shown here because both the sky imager (based on visible images) and the lidar do not operate during nighttime.

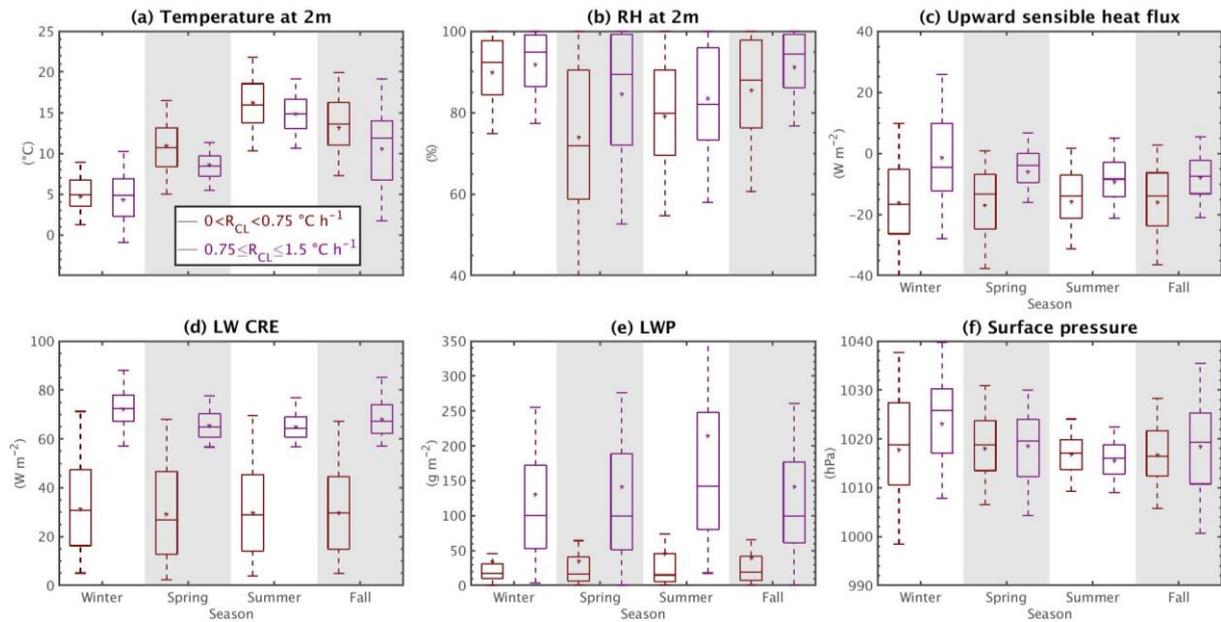


**Figure 4-11:**  $\partial T_{2m}/\partial t_{obs}$  nighttime values for two  $R_{CL}$  bins:  $0 < R_{CL} < 0.75^\circ\text{C}/\text{h}$  in maroon and  $0.75 \leq R_{CL} \leq 1.5^\circ\text{C}/\text{h}$  in purple. Distributions are represented by box-and-whiskers plots, where the boxes indicate the 25<sup>th</sup> and the 75<sup>th</sup> data percentiles, the whiskers indicate 5<sup>th</sup> and 95<sup>th</sup> percentiles, the middle line represents the median, and the \* is the mean.

Figure 4-11 shows that the temperature variations are generally stronger for the second bin than for the first bin, i.e. the negative temperature variations induced by longwave cooling during nighttime are damped when the effect of clouds is stronger. This agrees with what is expected since, at hourly time scales, clouds are the most important factor in the temperature variations during nighttime (see section 3.3.2.). As expected, the second bin has stronger LWCRE values (Figure 4-13e). The effect of clouds seems enhanced when cool and moist air is present over SIRTa (Figure 4-13a and Figure 4-13b). More details for each bin are given in the following subsections.



**Figure 4-12:** Same as Figure 4-5 but for nighttime cloudy cases.



**Figure 4-13:** Nighttime values for (a) temperature, (b) relative humidity, (c) upward sensible heat flux at 2 m, (d) longwave cloud radiative effect, (e) liquid water path and (f) surface pressure. Colors and boxplots follow the same definition as in Figure 4-7.

#### 4.4.1 Clouds with weak warming effect

For the first bin ( $0 < R_{CL} < 0.75 \text{ °C/h}$ ), LWCRE presents a wide range of values going from  $5 \text{ W/m}^2$  up to  $75 \text{ W/m}^2$  with no seasonal variability detected (Figure 4-13d), and yet temperatures are the highest for this bin (Figure 4-13a) except in winter. As expected, upward sensible heat flux is most of the time negative due to surface cooling, as values of  $\frac{\partial T_{2m}}{\partial t}_{obs}$  are predominantly negative (Figure 4-13c and Figure 4-11, respectively), with drier air conditions since RH has the lowest both mean and extremes values for all the seasons (Figure 4-13b).

Figure 4-12 shows the occurrence of precipitations and mean rates split into seasons for nighttime cases. For this bin 1, mean precipitation rates are very low but yet their occurrence is higher than for bin 2 at all the months, except winter. These low precipitation rates are related to the very low LWP found for this bin (Figure 4-13e) (Tian et al., 2019; Geoffroy, 2008) but since the occurrence of precipitation is very high, they could be at the origin of the extreme high RH values that can reach up to 100 %.

#### 4.4.2 Clouds with strong warming effect

The temperature distribution has the lowest values for all seasons (except spring) for the category of clouds that have the strongest warming effects (second bin, in purple on Figure 4-13a). This category of clouds is tightly linked to strong LWCRE ranging from  $55 \text{ W/m}^2$  up to  $90 \text{ W/m}^2$  (Figure 4-13d). These low temperatures could be

associated with situations when the incoming air masses are colder, which is linked to large-scale situations (Pinaridi and Masetti, 2000; Wang et al., 2005) since surface pressures are always lower for this bin 2 (except in winter) than for bin 1 (Figure 4-13f).

Finally, the distribution of sensible heat flux is positive in winter with values reaching up to  $40 \text{ W/m}^2$  (purple box plot, Figure 4-13c). These unusual and positive values during nighttime are explained by the presence of a strong stable atmospheric layer, where surface temperatures are the lowest (except for spring) (Figure 4-13a) and the atmosphere is warmer at higher altitudes since clouds contribute to warm the most at this time, and therefore a positive upward sensible heat flux develops. Similar behavior was previously found by Miao et al. (2012) in Beijing also during nighttime in cloudy conditions, where sensible and latent heat fluxes were close to zero due to the presence of a stable atmospheric layer, especially in winter when the sensible heat flux is found to be slightly greater than the latent heat flux. Indeed, almost half of temperature variation values are found to be positive for this season, as illustrated in Figure 4-11, and along with the positive upward sensible heat flux found, higher temperatures are found for this bin which is not seen in the other seasons where the sensible heat flux is close to zero or mostly negative and temperatures are higher for the first bin (cloud with a low warming effect).

## CONCLUSION

This chapter highlights the role of clouds on one-hour surface temperature variations. Different cases have been created in order to separate the contribution of clouds, for both day and nighttime. A deeper analysis is performed based on lidar profile observations available at SIRTA during the non-rainy day in order to better characterize the type of clouds for each condition/case. This analysis shows that:

- During daytime, clouds with a strong cooling effect that considerably counterbalanced the warming contribution of the sun are characterized by a large cloud cover, a very humid air with high precipitation rates, and little development of sensible heat flux. Among those kinds of clouds, the non-precipitant ones are often mid- and high-level clouds for all the seasons. This is partly related to the geographical position of SIRTA since it is located in a transition zone where warm and wet air coming from the Atlantic Ocean can nudge against the cold and dry masses of air coming from the Siberian region which favors the formation of these clouds.
- Clouds that warm the surface correspond to weak cloud cover, with almost zero precipitation which could be linked to the presence of fair weather cumulus. Nevertheless, these cases mostly occur in early morning, when the sun radiative effect is weak, and so they do not correspond to high values of surface temperatures. Moreover, higher surface temperatures are even found when clouds have

a cooling effect, because they also correspond to cases with high solar radiation warming effect (except for summer).

- During nighttime, temperature variations could partly be due to the sensitivity of LW radiation to the presence of atmospheric gases and aerosols, and not only to clouds. Furthermore, nightly temperature variations are also influenced by the stability of the atmosphere that creates positive sensible heat flux when clouds have an important warming effect.

The role of clouds and the contribution and importance of other surface terms on temperature variations have been deeply analyzed at SIRTA in Chapters 3 and 4, but depending on the zone of study and the temporal scale considered, these results will vary. The next Chapter consists in studying the variability of these results. For that, the contribution of the OTEM terms are separated into weather regimes, and the OTEM model is applied to other data at the Météopole site in Toulouse, France.



## 5 CHAPTER 5: METEOROLOGICAL AND GEOGRAPHICAL VARIABILITY OF THE OTEM TERMS

---

### INTRODUCTION

The previous results were at one place (SIRTA), and regardless of the large-scale circulation conditions. However, one can easily anticipate that these results can be influenced by the place and by these large-scale conditions. This variability is studied in this chapter.

First, previous results are analyzed for different weather regimes existing at SIRTA (NAO<sup>-</sup>, NAO<sup>+</sup>, Atlantic Ridge and Blocking) in summer and winter. These weather regimes indeed have different characteristics and defined patterns that will conditionate the behavior of the terms controlling surface temperatures, and thus a direct relationship between these large- and small-scale processes is set. Finally, OTEM is applied to another dataset from Toulouse (France) Météopole in order to evaluate how the location could influence the five OTEM terms, their contribution and importance.

### 5.1 DEPENDENCE OF THE CONTRIBUTION OF OTEM TERMS TO WEATHER REGIMES

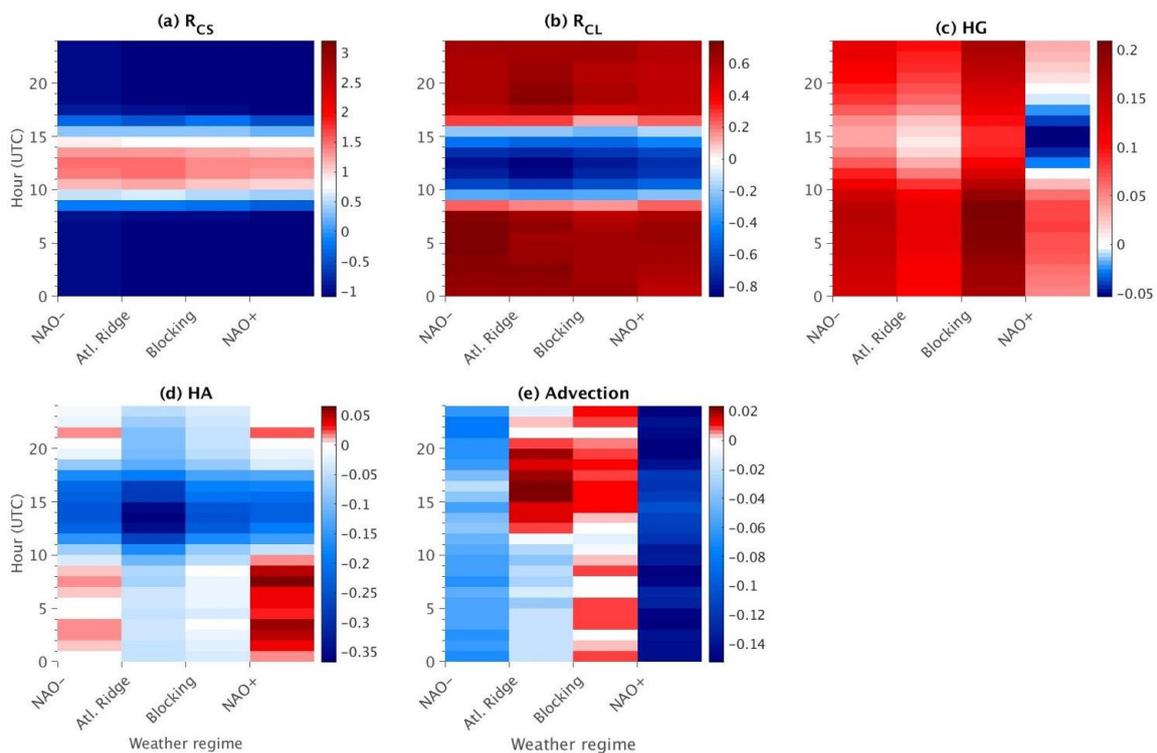
#### 5.1.1 Mean contribution of OTEM terms to weather regimes

In Chapter 1 it was mentioned that at first order, the large-scale atmospheric circulation dominates climate and weather variability. In Western Europe, the four main weather regimes controlling these processes in Winter and Summer are the North Atlantic Oscillation in both its negative and positive regimes (NAO<sup>-</sup> and NAO<sup>+</sup>), the Atlantic Ridge and the Blocking regime. In this section, the contribution of each OTEM term is analyzed independently for each weather regime in order to free itself as much as possible from the influence of the large scale: OTEM results separated onto weather regimes can be really interpreted as local processes influence. Also, comparison of OTEM results from one weather regime to another can be interpreted as the influence of large-scale.

Figure 5-1 and Figure 5-2 show the hourly mean values of the five terms of OTEM for winter and summer, respectively, separated onto weather regimes. For winter, the separation in time regimes has no effect on the radiative terms (Figure 5-1a and Figure 5-1b) because the mean contribution of these terms does not vary significantly among the weather regimes. However, HG and *Adv* have a strong negative contribution during the day for the NAO<sup>+</sup> regime, which agrees with the fact that this regime presents an

intensification of winds and precipitations linked to the increase of the Azores high- and the Icelandic low-pressure systems. Furthermore, HG was found to have a mean positive contribution during the winter season (cf. Figure 3-11f), but when separating in weather regimes this term continues to be positive except for NAO<sup>+</sup> with a negative contribution in the afternoon, and a positive and weak value for the rest of the hour.

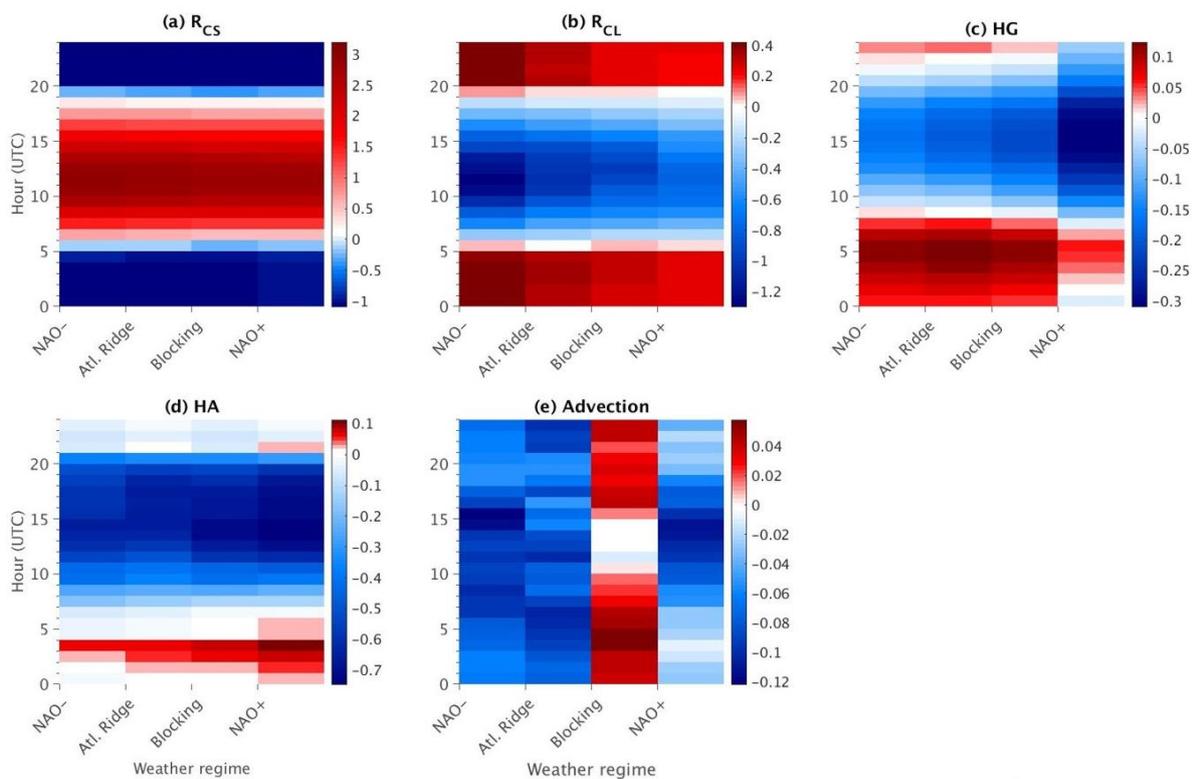
On the other hand, *Adv* contribution decreases in the NAO<sup>-</sup> comparing to other regimes, following the attenuation of wind speed at this weather regime. This term follows a similar behavior for both the Atl. Ridge and Blocking regimes having mostly a positive but still weak mean value. The advection term is in fact the one that varies the most from one regime to another compared to the other terms, variation that is not seen in Figure 3-11h when mean values for the winter months are all negative. This weather-to-weather variation shows that *Adv* plays a significant role even on the hourly time scale in these weather regimes.



**Figure 5-1:** Weather-hourly mean values for winter of (a)  $R_{CS}$ , (b)  $R_{CL}$ , (b) HG, (d) HA and (e) Adv and (f) the residual (i.e. difference between OTEM and the observations). Units in the color bars are all in  $^{\circ}\text{C}/\text{h}$ , and their scale is different for each subfigure.

For the contribution of each OTEM term in summer season (Figure 5-2), it is expected that  $R_{CS}$  and  $R_{CL}$  have a higher contribution than for winter, but in summer HG presents an opposite pattern than in winter by contributing to cool the surface during the day and to warm it during night (except for NAO<sup>+</sup>), as shown in Figure 5-2c. Furthermore, the NAO<sup>+</sup> is also characterized by the presence of fair-weather and warm days, hence

a weak contribution of  $R_{CL}$  (which is the main modulator of  $R_{CS}$  during the day, as shown in Section 3.2). In addition, the variation between each regime is minimal for all the terms except for  $Adv$ . The  $Adv$  term has a negative contribution for all the weather regimes (Figure 5-2e), except during the blocking one when the heatwaves occur more frequently and thus the other terms present a weak mean negative contribution. Indeed, this difference is not detected when the mean monthly-hourly values of  $Adv$  are calculated in Figure 3-11g for the summer months, when a mean negative contribution is found.



**Figure 5-2:** Same as Figure 5-1 but for summer.

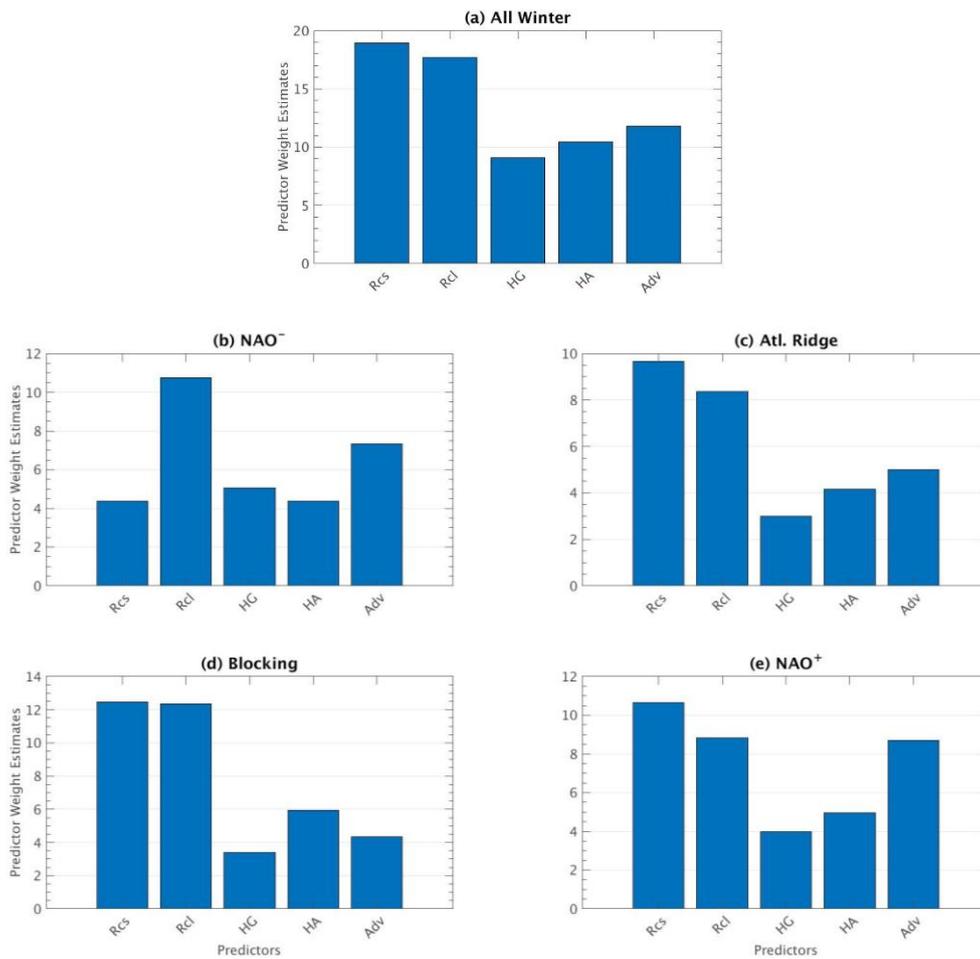
### 5.1.2 Weight of OTEM terms for the weather regimes

The random forest method used previously in Section 3.3. is here applied to determinate the weight of each OTEM term in each weather regime. Figure 5-3 shows the weight estimates during winter and for the four weather regimes. Here we notice that:

- The  $NAO^+$  regime (Figure 5-3d) has a relatively similar behavior than the whole winter season (Figure 5-3a):  $R_{CS}$  is the most important term followed by  $R_{CL}$  and  $Adv$ , the latter being in fact an important process as the westerly winds become stronger.

- For NAO<sup>-</sup> (Figure 5-3b), the contribution of clouds controls temperature variations, as bad weather conditions are typical of this regime. Moreover, because of the weakening of the westerly winds during this weather regime, Siberian and polar air can easily invade France lowering surface temperatures by advecting cold air and thus *Adv* is the second most important term. The weight estimates for NAO<sup>-</sup> are in fact the ones presenting most difference with respect to the estimates for the whole winter.
- The temperature variations during the Atlantic ridge regime are largely controlled by the radiative terms (Figure 5-3c) followed by *Adv*. The latter is certainly a key term in dampening the ridge circulation effect on surface temperatures during winter season, as shown by (Sousa et al., 2018b). For this weather regime, no significant differentiation with respect to the whole winter weight estimates is spotted.
- The blocking regime (Figure 5-3d) presents a dominance of the radiative term (similar to the Atlantic ridge), followed by the *HA* term. Since during this regime fair weather conditions are established thanks to the Siberian High installed at that moment, the boundary layer can further develop and this *HA* term becomes an important modulator along with *R<sub>CL</sub>* and *R<sub>CS</sub>*, which is not seen when considering all the winter season (Figure 5-3a).

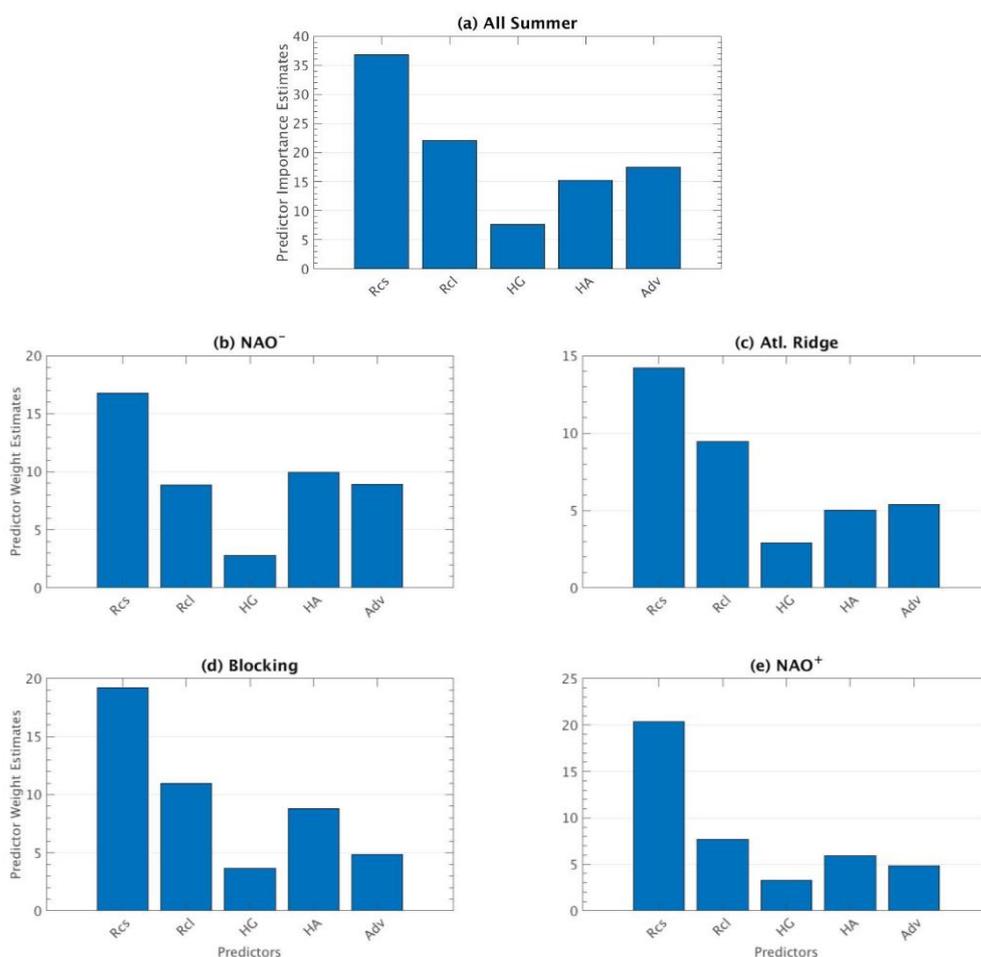
Furthermore, the fact that in winter there are more night than day hours makes that *R<sub>CL</sub>* play a predominant role for all the regimes.



**Figure 5-3:** Predictor weight estimates obtained by the random forest method for (a) NAO<sup>-</sup>, (b) Atl. Ridge, (c) Blocking and (d) NAO<sup>+</sup> during winter.

During summer (Figure 5-4), It is found that:

- Most of the weights for each term follow the same pattern whether the analysis is done by regime or the whole summer season.
- Differences between whole summer and the specific regimes are mostly found for the NAO<sup>-</sup> regime when HA is the second most important term after R<sub>CS</sub>.
- Hence R<sub>CL</sub> is the second most important term for all the regimes except NAO<sup>-</sup>, when generally a dry and mild weather is installed during this regime.



**Figure 5-4:** Same as Figure 5-3 but during summer season.

In summer, the large-scale circulation condition does not play a major role on the repartition of the weight of the different terms that control the one-hour temperature variations (similar behavior is found whether we separate or not the analysis onto weather regimes). In winter on the contrary, large-scale conditions are decisive: depending on whether we are in a particular weather regime, it could be sun radiation or clouds or air advection that controls at first order one-hour temperature variation.

## 5.2 SPATIAL VARIABILITY OF OTEM TERMS: COMPARISON BETWEEN TOULOUSE AND SIRTÀ

Chapter 3 and 4 have been consecrated to study the influence of OTEM terms on temperature variations exclusively in the SIRTÀ Observatory. In this section, OTEM is applied to a dataset from the Météopole site in Toulouse (cf. Section 2.1.2.) from January 2016 to December 2017, for the purpose to evaluate how local conditions affect the contribution of each surface term and compared them with the values found

at SIRTa. Each OTEM term was previously estimated in Section 2.2.1.3.

The two observatories are indeed located in two zones where weather conditions are different. As mentioned in Chapter 1, SIRTa is located in a zone where Siberian and Atlantic circulations meet, controlling temperatures and precipitations. However, in the case of Toulouse, weather is controlled by the westerly winds coming from the Atlantic Ocean and the S-E wind regimes coming from the Mediterranean. Furthermore, the Météopole is located in an urban area close to the city center (cf. Figure 2-1b) and have a different soil type than SIRTa (cf. Section 2.2.1.3.) that will affect the surface energy balance and thus the distribution and contribution of the terms controlling surface temperature variations. Actually, SIRTa presents limestone soil type, whereas Toulouse has a predominance of clay loam soil. These aspects need to be considered when comparison between the two sites is done in the following sections.

### 5.2.1 Statistics of OTEM in the Météopole

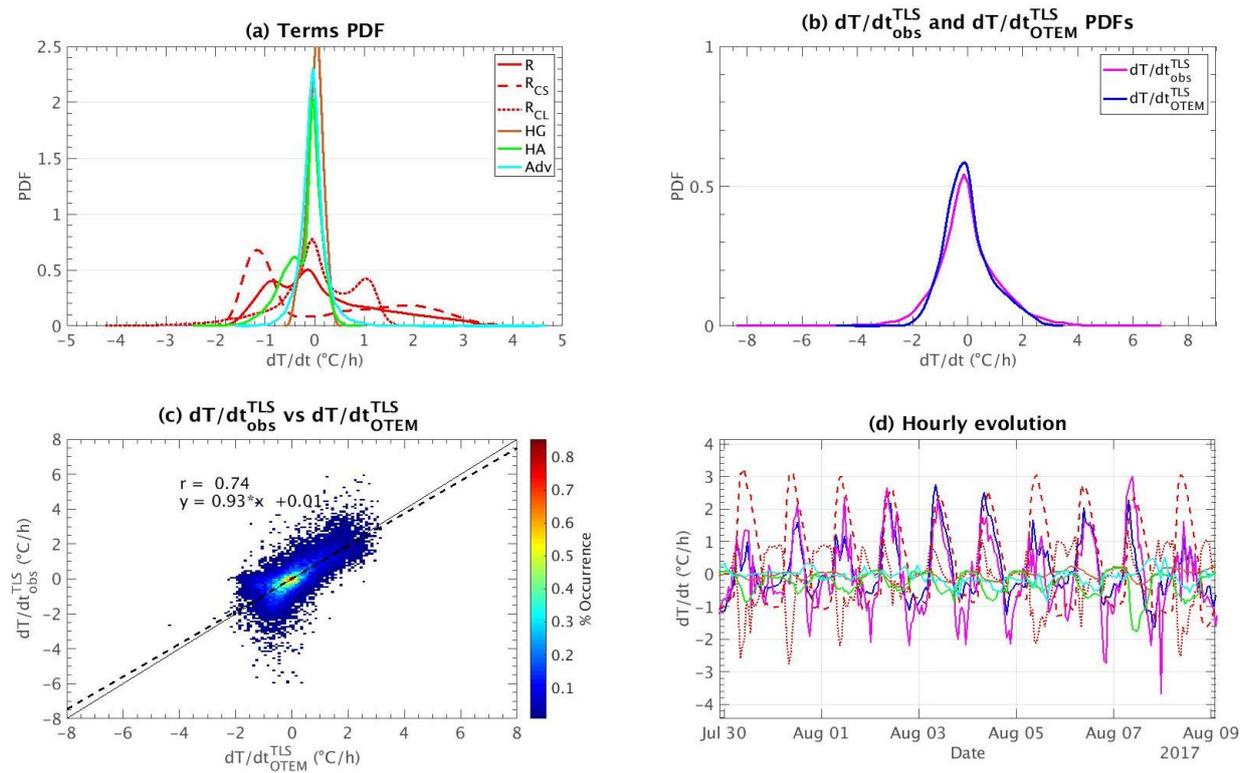
As done for SIRTa, a statistical evaluation of OTEM is done but for the Météopole dataset in order to evaluate the reliability of the results. Figure 5-5a and Figure 5-5b show the PDF of the five terms,  $\frac{\partial T_{2m}^{TLS}}{\partial t}_{obs}$  and  $\frac{\partial T_{2m}^{TLS}}{\partial t}_{OTEM}$ :

- The radiative term  $R$  (red line) has values between  $-2.0$  °C/h and  $3.6$  °C/h with a negative peak at  $-1.0$  °C/h and another around  $-0.2$  °C/h.
  - o  $R_{CS}$  contribution values go from  $-2.5$  to  $3.6$  °C/h with a peak at  $-1.2$  °C/h.
  - o  $R_{CL}$  distribution ranges from  $-3.1$  °C/h to  $1.7$  °C/h, with a considerable amount of cases around  $1.0$  °C/h.
- Both  $HG$  and  $Adv$  follow a Gaussian distribution, with contribution values ranging from  $-0.6$  °C/h to  $0.6$  °C/h for the former, and from  $-1.8$  °C/h to  $4.6$  °C/h for the latter.
- The  $HA$  term has a majority of negative values with a negative contribution reaching down to  $-4.4$  °C/h, and a weak positive one of  $1.0$  °C/h.

Concerning the PDFs of the observed and OTEM temperature one-hour variations (Figure 5-5b), the two curves follow a similar distribution. Furthermore, observed temperature variations at the Météopole site for the two years of study can go down to  $-8.2$  °C/h and up to  $6.6$  °C/h, whereas OTEM varies a bit less ranging from  $-4.4$  °C/h to  $3.12$  °C/h.

Figure 5-5c shows the scatterplot of  $\frac{\partial T_{2m}^{TLS}}{\partial t}_{obs}$  vs  $\frac{\partial T_{2m}^{TLS}}{\partial t}_{OTEM}$ , as well as the correlation coefficient "r" and the linear equation fitting the best for the two datasets. Indeed, the

correlation of 0.72 found at the Météopole is lower than that at SIRTA (0.79, cf. Section 3.1.1.) but still remains a good estimation value between the two datasets. Finally, Figure 5-5d shows the hourly evolution of each of the five terms,  $\frac{\partial T_{2m}}{\partial t}_{obs}^{TLS}$  and  $\frac{\partial T_{2m}}{\partial t}_{OTEM}^{TLS}$  from 07/30/2017 to 08/09/2017, showing that OTEM follows quite well in average the observations. Moreover, for this particular case, observed temperatures variations presents diurnal fluctuations that OTEM gets to capture quite well, as seen e.g. for the 08/01/2017 or the 08/02/2017. As expected,  $R_{CS}$  has a strong diurnal cycle and the other four terms tend to modulate its important contribution.



**Figure 5-5:** PDFs of (a) radiation (red line), radiative clear sky (red dashed line), radiative cloud (red dotted line), heat ground (brown line) and heat atmospheric (green line) exchange, and Adv (cyan line) terms, and (b)  $\frac{\partial T_{2m}}{\partial t}_{obs}^{TLS}$  (pink line) and the  $\frac{\partial T_{2m}}{\partial t}_{OTEM}^{TLS}$  (blue line). (c) is the scatter plot of  $\frac{\partial T_{2m}}{\partial t}_{obs}^{TLS}$  vs  $\frac{\partial T_{2m}}{\partial t}_{OTEM}^{TLS}$ . The sloping solid line represents the 1:1 line and the dashed one correspond to the least-square best fit linear regression line between the two datasets. The correlation coefficient "r" and the linear equation fitting the best for the two datasets are also indicated. (d) Hourly evolution for some days of August 2017 of the temperature variation for the different terms, the observations, and OTEM (same colors as in (a) and (b)).

## 5.2.2 Comparison between SIRTA and the Météopole

Now that we have made sure that OTEM method is reliable for the Météopole dataset

as well as it was for SIRTA dataset, we can compare the two sites in order to analyze the spatial variability of these surface terms (even if the period is not the same) and the influence of specific local conditions.

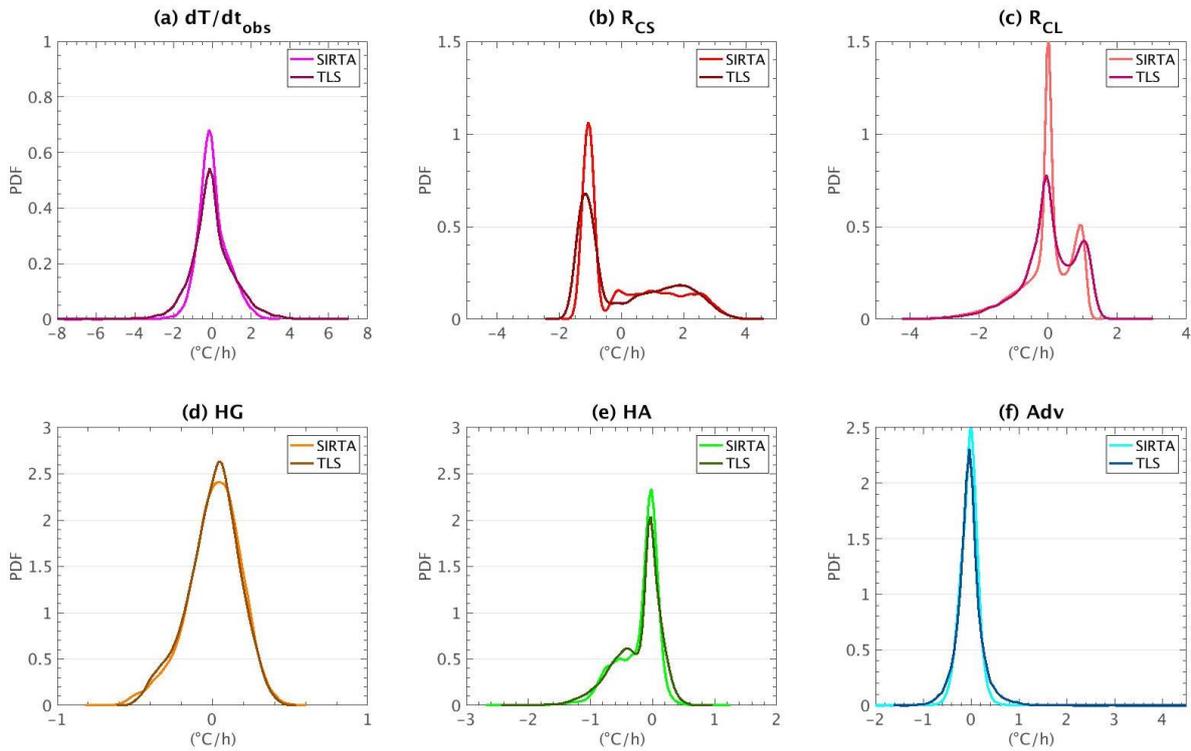
### 5.2.2.1 General distribution of OTEM terms

Figure 5-6 shows the PDFs of the observed temperature variations, and the five terms of OTEM for both SIRTA and Toulouse sites. Concerning  $\frac{\partial T_{2m}}{\partial t}_{obs}$  (Figure 5-6a), Toulouse presents higher positive temperature variation values than SIRTA (mostly at day, see Appendix C.1, Figure C-7), with a maximum of +6.63 °C/h for the former whereas it is +3.2 °C/h for the latter. It is quite expected indeed that  $\frac{\partial T_{2m}}{\partial t}_{obs}$  in Toulouse presents higher values than at SIRTA since the former is located further south in France where on average weather conditions are warmer and get larger direct sunlight than at the north. A similar behavior is found for the negative values, with a higher distribution in Toulouse (at night, see Appendix C.1, Figure C-8) with values around -2.0 °C/h than at SIRTA. Surface temperatures thus present a higher and more rapid hourly growth rate along the day at Toulouse. Concerning the OTEM terms:

- $R_{CS}$  is higher at Toulouse than at SIRTA (Figure 5-6b), with a maximal value around 4.5 °C/h for the Météopole, explained by the fact of having larger direct sunlight than at SIRTA, where the maximum value is around 4.0 °C/h. Nevertheless, the two distributions follow a similar behavior, with a negative peak for both sites located almost at the same value around -1.6 °C/h, and similar patterns for the positive values.
- $R_{CL}$  is also higher at Toulouse (Figure 5-6c) with values reaching up to 2.7 °C/h whereas for SIRTA it goes up to 1.37 °C/h. With respect to the negative maximal values, both sites present similar values of -3.8 °C/h.
- Figure 5-6d shows that HG PDF for both sites are nearly identical. Nevertheless, at SIRTA this term contributes to cool the surface down to -0.80 °C/h whereas at Toulouse its minimal value is -0.53 °C/h. This is explained by the difference of heat capacity between the two soils, which is higher for Toulouse and thus it requires more energy (and hence more time) to change its temperature yielding to negative values higher than at SIRTA.
- For  $HA$ , SIRTA and Toulouse PDF shows a similar distribution with a peak around 0 °C/h, with a higher negative than positive contribution to temperature variations. At Toulouse, this term cools the surface down to a minimal value of -4.3 °C/h whereas at SIRTA is -2.6 °C/h.
- The advection term in Toulouse presents a higher positive contribution than at SIRTA, value that go up to 4.6 °C/h linked to the geographical position of the Météopole site, where advection plays a major role on weather and 2 m temperature variations (see Section 5.1.1.3.). At SIRTA, a lower value of  $Adv$  is found

since SIRTA is located in a northern area that are largely influenced by the cold winds coming from the Siberian region.

The geographical position of each observatory plays indeed an important role on determining the distribution of each OTEM term, since each of them is affected primarily by the incoming solar radiation which itself depends on the latitude. After having studied their distribution, the following sub-section is consecrated to analyze their average contribution from a monthly-hourly point of view (as done at SIRTA in Chapter 3).



**Figure 5-6:** PDF of (a)  $\frac{\partial T_{2m}}{\partial t}_{obs}$ , (b)  $R_{CS}$ , (c)  $R_{CL}$ , (d)  $HG$ , (e)  $HA$  and (f)  $Adv$  for SIRTA (light colors) and the Météopole site (dark colors). X- and y-axis limits are different for each figure.

### 5.2.2.2 Dependence of the monthly-hourly evolution to the site

The monthly-hourly mean contribution of each term at the Météopole site is shown in Figure 5-7, along with a residual value (defined like for SIRTA study as the difference between OTEM and observed temperature variations).

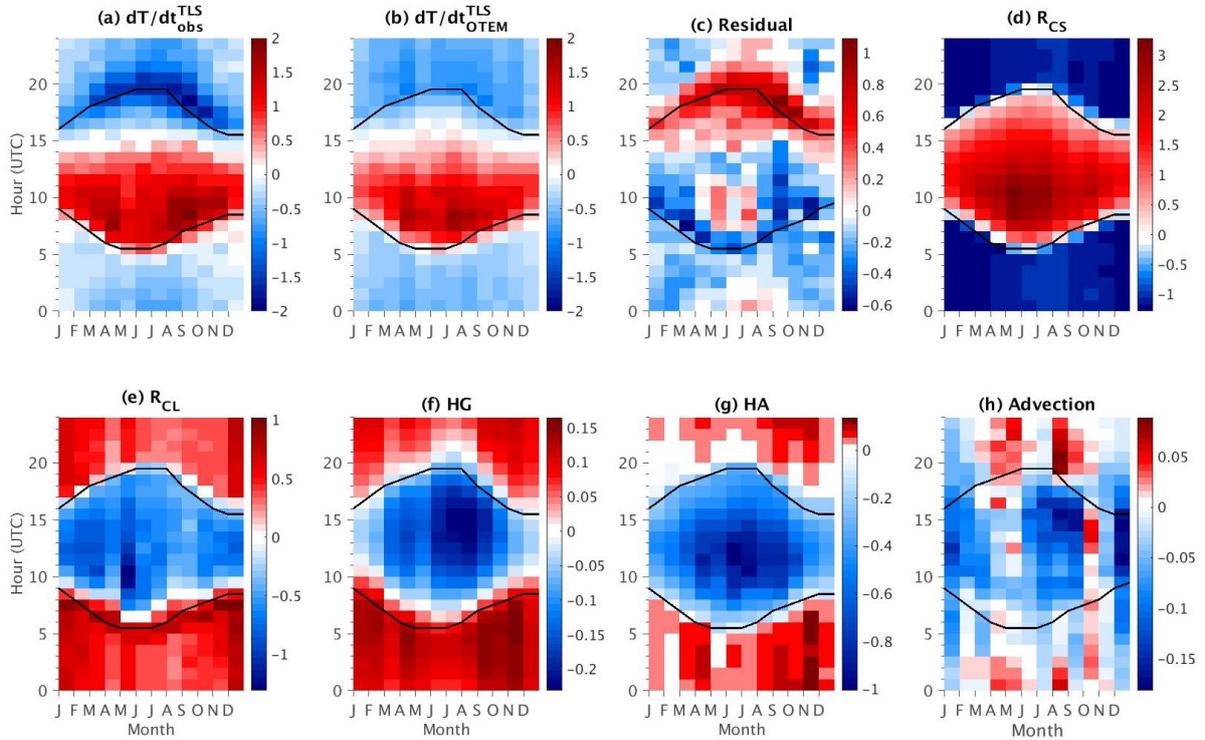
Figure 5-7a shows that the mean values of observed temperature variations are higher in the morning for all the months of the year, with mean values ranging from 1.5 to 2.0 °C/h at 08:00 UTC in September for instance.  $\frac{\partial T_{2m}}{\partial t}_{obs}^{TLS}$  is the lowest in the late

afternoon just at the sunset for all the months, with a minimal value of  $-2.0$  °C/h at 18:00 UTC in September. Indeed, during these well-known transition hours, OTEM encounters difficulties to reproduce the real observed temperature variations mean values, as illustrated in Figure 5-7c by the residual term. Actually, at these hours, the residual reaches values up to  $1.0$  °C/h in September and October at 18:00 UTC, and OTEM underestimates the hourly temperature variations, which is not seen the rest of the day. This underestimation also occurs at sunrise, but residual values are lower and ( $-0.8$  °C/h). These underestimations are higher than what we had at SIRTa (Figure 3-11c), where observed temperature variations were not as high as at the Météopole. Indeed, these transitions periods during the morning and evening are the two most complicated phases of the diurnal cycle of temperature variations over land, affecting OTEM performance. These periods are indeed difficult to model because of the change of the state of the atmosphere from night to day (stable to unstable) and day to night (unstable-convective to stable), which happen to be more marked at the Météopole site than at SIRTa. Furthermore, the hypothesis of assuming that the vertical temperature profile at these hours is the same as for the rest of the day where the atmosphere is considered unstable (cf. 2.2.1.1.) could induce an error that is obviously accentuated at Toulouse where temperature variations are more important at these specific hours. However, during the day, the residual is generally low, with some specific exceptions, for example in May at 12:00 UTC when a mean residual value of  $-0.7$  °C/h is found. On the contrary, during the night the residual is on average close to zero for all months.

All the OTEM terms are again driven by the solar diurnal cycle except  $Adv$  (as it was found for SIRTa). This figure shows us that:

- $R_{CS}$  presents a negative constant value of  $-1.3$  °C/h during all the year. This nighttime negative contribution is lower at Toulouse than at SIRTa, because the LW radiation of the surface is more important for the former thanks to a higher incident solar radiation during daytime.
- $R_{CL}$  contributes to warm the surface during nighttime with mean maximum contribution values of  $1.0$  °C/h whereas at SIRTa the maximum is around  $0.7$  °C/h. This difference is due to the higher amount of LW downwards radiation found at Toulouse than at SIRTa (not shown).
- The  $HG$  term (Figure 5-7f) is positive after the sunset hours (black bottom line) at Toulouse for all the months of the year, whereas at SIRTa this behavior is only seen in the winter months.
- Figure 5-7g shows that  $HA$  contribution to temperature variations is very similar to that found at SIRTa with almost negligible (yet positive) close-to-zero mean values during nighttime, and during daytime it always acts as a modulator of  $R_{CS}$  with a mean negative contribution.
- Mean  $Adv$  term contribution is the lowest among the other terms but it tends to contribute on average to cool the surface (Figure 5-7h). Positive mean values

for this term occur more often at Toulouse than at SIRTA, where a warming effect is only found at some hours in May (Figure 3-11h).



**Figure 5-7:** Monthly-hourly mean values for (a)  $\frac{\partial T_{2m}^{TLS}}{\partial t}_{obs}$ , (b)  $\frac{\partial T_{2m}^{TLS}}{\partial t}_{OTEM}$  (c) the residual (i.e. difference between OTEM and the observations), for (d)  $R_{CS}$ , (e)  $R_{CL}$ , (f) HG, (g) HA and (h) Adv. Units in the color bars are all in  $^{\circ}C/h$ , and their scale is different for each subfigure. The black contour lines in each figure correspond to sunrise (bottom line) and sunset (top line) approximative hours.

### 5.2.2.3 Dependence of the weight of OTEM terms to the site

In Chapter 3, Section 3.1.8., the random forest method was used to estimate the diurnal cycle of the weight of each term involved in OTEM at SIRTA. In this section, we perform the same analysis for the Météopole site and shown in Figure 5-8 (right 4 figures).

During nighttime hours, it is indeed  $R_{CL}$  the term controlling temperature variations at this scale for all the four seasons, followed by the Adv term (except in fall), as it was the case for SIRTA results. Since at night very few turbulent heat fluxes develop, it is not surprising that the advection becomes the second most important term for all the seasons. This is explained by the urban heat flux that develops further at Toulouse than at SIRTA, where the city warms up in summer during the day and cools down slowly than the surrounding countryside, and thus an advection flux from a cold to a warm

point develops.

During daytime, a general behavior is close to what we found for SIRTA (especially for the radiative terms), except for the following differences.

The *Adv* term presents the largest difference between the two sites. At Toulouse in the afternoon, *Adv* term controls hourly temperature variations, followed by *HA* for all seasons. On the contrary, at SIRTA it is in general *HA* that dominates the afternoon hours (Figure 5-8, 4 left figures). Indeed, *Adv* weight estimates are quite important a couple of hours before the sunset for Toulouse, that could be explained by the month-to-month variability of *Adv* mean contribution values at these hours as seen in Figure 5-7h, especially in summer and fall when this term clearly dominates temperature variations (Figure 5-8f and Figure 5-8h). In addition, Toulouse is located in a zone where two wind regimes constantly meet: the Tramontane and the Mistral winds, considered as Mediterranean wind regimes that can affect both the surface energy balance and the hydrological sea and land cycle due to their important wind speeds (Obermann et al., 2018). Indeed, Flamant and Pelon (1996) showed that under such wind regimes, vertical velocities are high and thus vertical turbulent transport increases, which is at the same time related to the growth rate of the ABL that happens to be maximal in the late afternoon and thus *Adv* and *HA* have important weight for all the seasons at these hours.

In the early morning just after the sunrise, *HA* becomes more important than the other terms in winter at SIRTA, whereas at Toulouse it is  $R_{CS}$ . This again is linked to the variability of *HA* at these hours at SIRTA that is higher than at Toulouse.

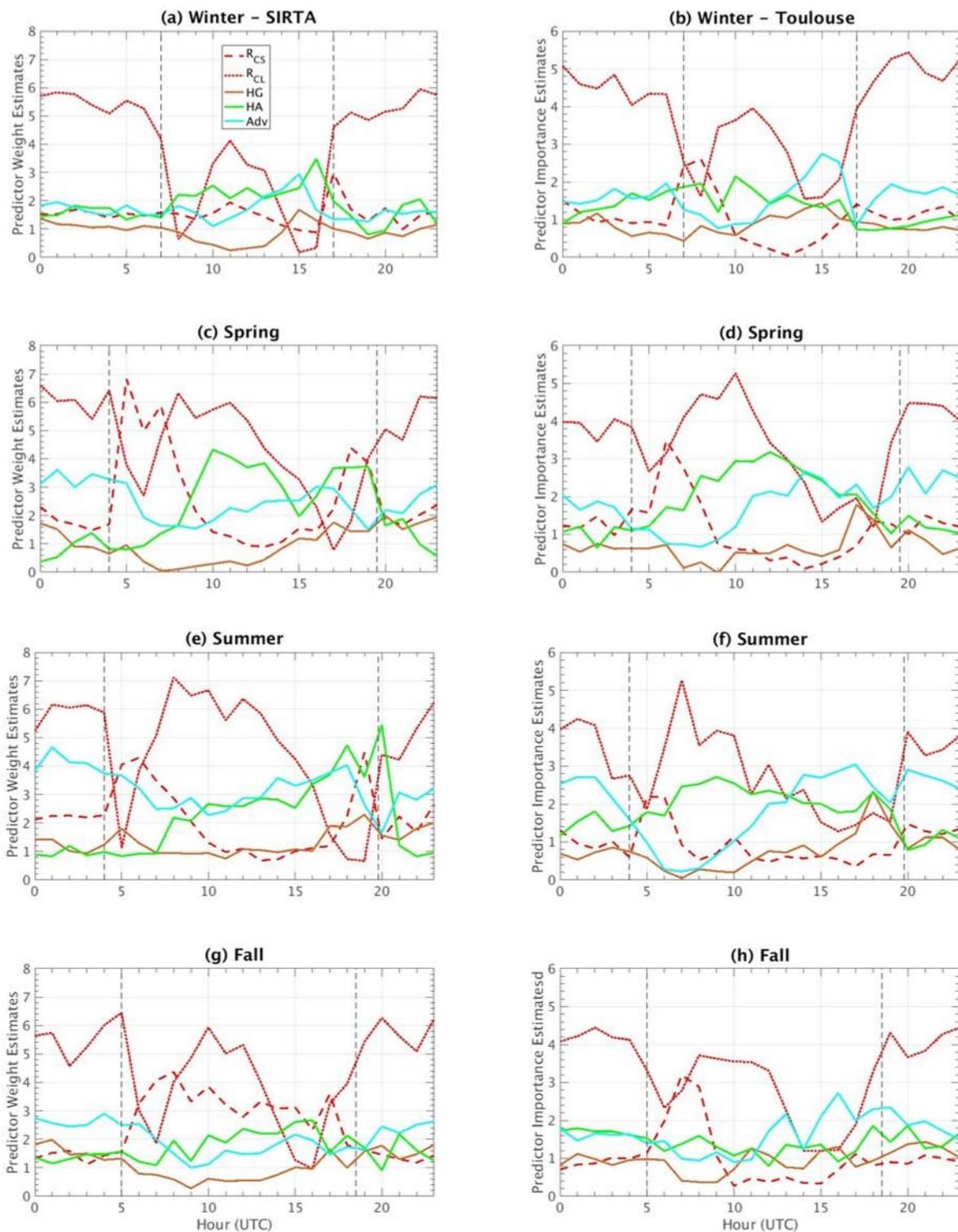
## CONCLUSION

This chapter showed an analysis of the variability of the OTEM terms. Small scale processes (i.e. OTEM terms) are more sensitive to large-scale circulations in winter than in summer when the contribution and importance of the OTEM terms vary more from one regime to another, especially the *Adv* term which is the one that is largely implied in both large and small temporal scales. Large-scale processes then play a major role on both 2 m temperature variations and OTEM terms mostly during winter season.

Local conditions are the factor controlling the differences of OTEM terms between the two sites, especially for the *Adv* term that plays a major role in Toulouse that is not seen at SIRTA during daytime because of the regularly presence of the two regional regime winds at the Météopole. As expected, the contribution of the clear-sky and cloud radiative terms is higher due to the increase in solar radiation at the location of the Météopole. Moreover, OTEM encounters further difficulties to reproduce the observed temperature variations in the transition periods at Toulouse, where observed temperature variations at these hours are extreme, and they drastically decrease at sunset hours. It could be interesting to study what are the mechanism behind this

behavior, as during sunrise and sunset periods on average the contribution of the OTEM terms are close to zero, and a sudden decrease of the ABL is detected at these hours (see Appendix C.2, Figure C-9).

The application of OTEM to different observatories opens the perspective of further analyze the spatial distribution of the terms. Indeed, the SIRTAReOBS approach is currently being applied to other French observatories located in zones with different weather and geographical conditions. It would be e.g. interesting to apply OTEM and to determine the contribution of each term in a mountain zone where local conditions are completely different from those at SIRTAReOBS or at Toulouse.



**Figure 5-8:** Diurnal cycle of the predictor weight estimate for each term of OTEM split into seasons at SIRT (left column, exactly similar than Figure 3-13) and at the Météopole site (right column). Vertical dashed black lines are for sunrise and sunset mean hours.



## CONCLUSIONS AND PERSPECTIVES

---

The European climate and most temperature anomalies are affected not only by the circulation of large-scale air masses (such as the NAO or blocking regimes) but also by smaller-scale processes such as cloud radiation and surface-atmosphere interactions located within the atmospheric boundary layer. The first objective of this thesis was to develop a method (OTEM) to study these small-scale processes and how they affect surface temperature variations in a local scale through these five terms:

- The radiative clear sky and cloud terms, that are controlled by the radiative SW and LW fluxes arriving at the surface.
- The atmospheric heat exchange term that includes the sensible and latent heat fluxes developed at the surface.
- The ground heat exchange term, which represents the contribution of the heat flux into the ground.
- The air advection.

All of these terms can be estimated based almost exclusively on observations, with few hypotheses using the SIRTA-ReOBS dataset in an hourly scale. When each of the terms is estimated, their sum exhibits good accuracy when compared to the real observations of surface temperature one-hour variation, and it is able to well quantify the realistic diurnal and monthly-hourly cycles of each term involved, which themselves are controlled by the diurnal solar cycle. However, there are some cases when OTEM is not able to reproduce the sudden and extreme decrease of observed temperature variations within an hour, partly due to the fact that they correspond to processes that occur in less than an hour (e.g. the cold pool events). In addition, other possible sources of error that lead to OTEM not providing a "perfect" estimate of observed temperature variations are:

- Downwards clear sky fluxes have a 5% error when performing the algorithm that calculates them.
- The pressure at the mixed layer depth (MLD) is estimated with only two radiosoundings per day, and then for the rest of the hours the same observations are used to estimate the pressure at the closest hour available in the radiosounding, thus this pressure does not correspond to the real one.
- The use of the reanalysis to estimate few variables add another source of bias since they are outputs of models/simulations and do not represent real observations.
- The hypothesis of considering always day as an unstable and night as a stable atmosphere (for simplicity) implies the omission of cases when the opposite could happen and OTEM will tend to miscalculate some of its terms (e.g. the

radiative terms). This hypothesis also affects the calculation during the transition sunrise and sunset periods.

- The lack of latent heat fluxes data for the period of interest even if it is compensated by a parametrization. Furthermore, the instruments used to measure the latent and sensible heat fluxes give correct results in 80% of the situations on average because they are sensitive to rain which will disturb the measurement (detection surface impacted if there is water on it).

Despite those possible sources of error, OTEM presents a correlation coefficient of 0.8 (yet quite good) when compared to observations. However, it is during daytime that OTEM presents the best statistics, whereas at night its accuracy decreases. This low (but yet good) accuracy is related to the hypothesis of considering an average MLD value during nighttime (stable atmosphere hypothesis with a constant MLD) that affects both the radiative and atmospheric heat exchange terms. All of these error sources and assumptions do not affect the principal behavior of OTEM, and it remains a good method to estimate and quantify the terms affecting surface temperature variations.

With respect to the individual contribution of OTEM terms at SIRTA, the clear sky radiative term represents a big positive (negative) contribution during daytime (nighttime). In addition, the contribution of the radiative term in clear sky conditions remains on average constant throughout the year during nighttime whereas clouds maximal warming contribution is found during winter at these times. The atmospheric heat exchange becomes predominant in the late afternoons when turbulent heat fluxes further develop due to the increase in atmospheric instability and the ABL reaches its daily maximum, especially in spring and summer. In terms of magnitude, the ground heat exchange and advection terms contribute the least to surface temperature one-hour variations, but still they remain important to obtain the best coefficient estimator between the directly measured observations and OTEM, as they also affect and contribute to surface temperature variations. On average, the cloud radiative term along with the other non-radiative terms act as modulators to the important contribution of the radiative clear sky term.

The use of a random forest analysis makes it possible to identify which term dominates over the others. It confirms that clear sky radiation most influences the 1-hour temperature variations during the day followed by clouds radiation and the atmospheric heat exchange term. At night, it is the cloud radiation that controls temperature variations followed by the ground heat exchange term. Nevertheless, separating the dataset into hours and seasons shows that the cloud radiation effect becomes predominant in several hours for all the seasons because it presents a greater variation of its hourly values than the clear sky radiative effect, especially in spring and summer during daytime when more local cloud cover variability occurs. This cloud dominance is still found despite the fact that the clear-sky radiation magnitudes are

the highest on a monthly-hourly scale. Temperature variations are then more sensitive to cloud changes within an hour rather than the large contribution of clear sky radiation which mostly dominates the early mornings when surface heat fluxes are not yet developed. The other terms remain important as there are times when some of them modulate temperature variations, especially the turbulent heat fluxes in the late afternoons.

To better understand the importance of clouds on temperature variations at this scale (being the second objective of this thesis), their influence on other surface variables as well as their characterization (based on lidar profile observations) is assessed as a function of their radiative effects. The atmosphere presents a high amount of clouds during daytime when clouds have a strong cooling effect. Indeed, an important presence of mid-level thick clouds is spotted for all the seasons (except spring) for these cases, that could correspond to nimbostratus which are very opaque and often produce continuous moderate rain. On the contrary, a weak cooling radiative effect is predominantly associated with low- and high-level thin clouds for all seasons. In addition, situations with daytime positive cloud radiative effect occur when the atmosphere is close to clear-sky conditions but still cloudy with low- and high-level thin clouds occurring mostly in early mornings and late afternoons. Overall, a dominance of mid- and high-level clouds at the SIRTA observatory is detected during daytime. The dominant presence of this specific type of clouds is partly linked to the geographical position of the SIRTA observatory since it is located in a zone where warm and wet air coming from the Atlantic Ocean can nudge against the cold and dry masses of air coming from the Siberian region. This encounter of air masses (whether it is the warm air that encounters cold air or the opposite) tends to form mostly high-level clouds. At night, even when clouds have a significant positive radiative effect, air temperatures are low and are more controlled by large-scale processes, since surface turbulent fluxes are most of the time low and there is no sun radiation. Variability in cloud radiation is still found, partly due to the sensitivity of the radiation reaching the surface to the presence of atmospheric gases (such as water vapor) and aerosols and not only to the radiative properties of the clouds.

After having analyzed the extent to which OTEM terms and in particular clouds contribute to surface temperature variations, a final objective of this thesis is to study the variability of these results, looking at their relation to large-scale circulation processes, and their contribution at another location different from SIRTA.

The OTEM estimation classified into summer and winter weather regimes is studied to better relate the small-scale processes to the large-scale atmospheric dynamics. The contribution and importance of each OTEM term vary more when separating the analysis in weather regime during winter than during summer. Indeed, small-scale processes are more related to large-scale circulations in the former season, affecting

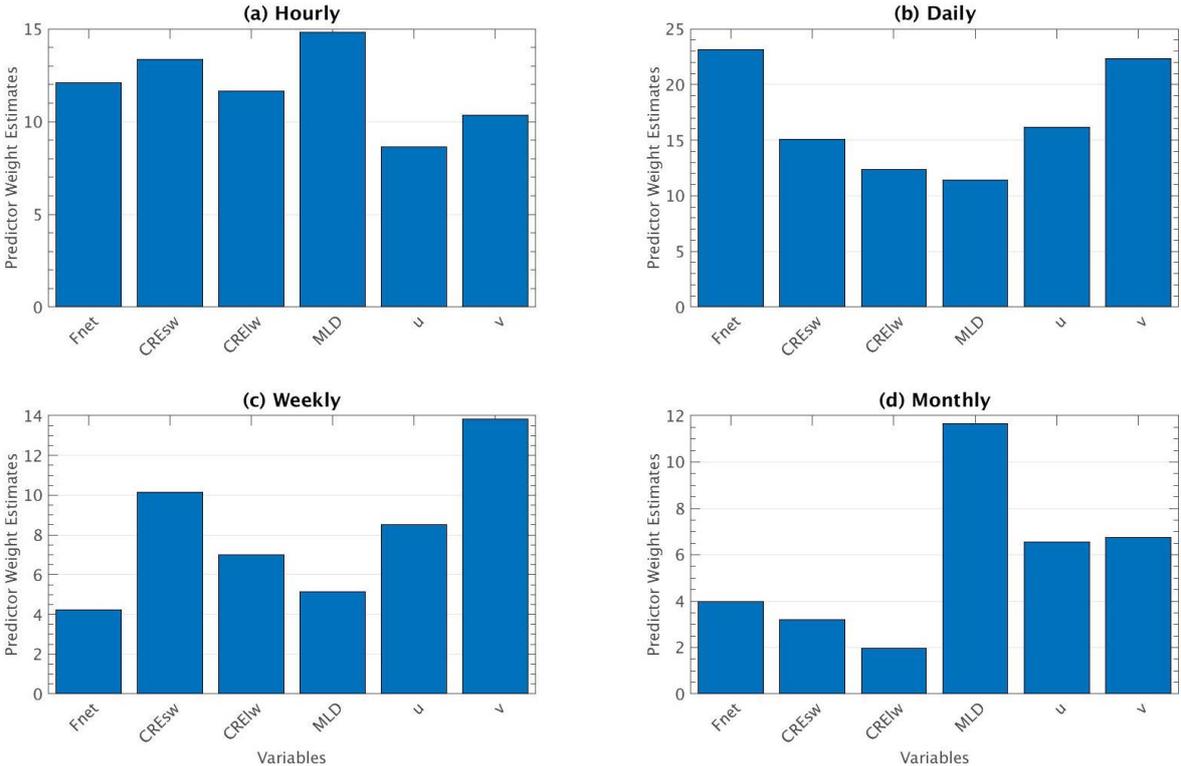
not only SEB variables but also precipitations (Santos et al., 2005), temperatures (Rojas et al., 2013b), and even atmospheric aerosols (Ménégoz et al., 2010).

OTEM is also applied to a dataset retrieved from the Météopole, a METEO-FRANCE observatory located 6 km south of Toulouse city center, in order to study the variability of the terms depending on specific local conditions. As done for SIRTa, OTEM is first evaluated to analyze its capacity to predict observed temperature variations. At the Météopole, a good correlation of 0.72 is found between OTEM and the measurements of temperature variations but it is lower than the one found at SIRTa (0.80), mainly because at Toulouse, the sunrise and sunset gives higher value of temperature variation, so it is more difficult to reproduce these specific hours. Indeed, this aspect opens a perspective to a further work on better understand these transition periods which are known to be the most complicated phases of the diurnal cycle and which affect not only the temperature but also the variability of atmospheric aerosols and gases. All OTEM terms (except the ground heat exchange term) present a higher positive contribution to surface temperature variations at Toulouse than at SIRTa, as meteorological conditions are fairer compared to the Paris region. The clearest difference found between the two sites concerns the advection term, which plays a major role at the Météopole. Toulouse is indeed located in an area subjected to two strong wind regimes coming from the West and the Southeast, that affect not only advection contribution but conditionate also turbulent flux development at the surface.

Extending this study to other observatories opens the perspective to deeply characterize the spatial variability of the results for a more comprehensive variety of local conditions. Another interesting perspective opened by this work is the temporal variability of the results: at what time scale (greater than an hour) such or such term will become preponderant for temperature variations? Increasing the temporal scale requires revisiting completely the equations and assumptions made in Chapter 2. But as an example, Figure 6-i shows the weight estimates of the main variables acting on 2 m temperature (not on temperature variations as previously done) on an hourly, daily, weekly and monthly scale. This simple example does not require new hypothesis because it is only based on variables and not on terms as in OTEM. We see that the hourly temperature depends almost as much on radiations as on clouds and mixing layer depth, whereas the daily temperature is less controlled by clouds and air advection becomes important, as for the weekly temperature. The monthly temperature hardly depends on the radiation anymore: the advection and the mixing layer depth have become the main variables that control temperature, related to the European weather regimes and the large-scale air masses circulations.

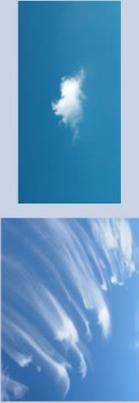
Table 6-i presents (i) the clouds that are most detected by the lidar, and (ii) the mean contribution values of each term of OTEM during daytime at SIRTa, split into seasons. On average, it's the radiative term that contributes the most to warm the surface, and

the other four terms act as modulators of this importance contribution, especially the cloud radiative and atmospheric heat exchange terms. The ground heat exchange and advection remain the terms that contribute the less but when looking at the role they play during the day, the advection term becomes the second most important term that controls temperature variations, just behind the atmospheric heat exchange in the late afternoon, as showed in Table 6-ii, and the ground heat exchange becomes more important than the radiative cloud term, despite the low contribution of the former.

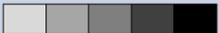
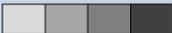
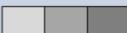
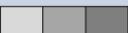
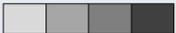
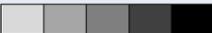
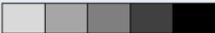
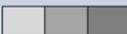
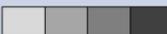
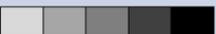


**Figure 6-i:** Predictor weight estimates of the main variables acting on 2 m surface temperatures for (a) hourly, (b) daily, (c) weekly and (d) monthly temporal scales.

The approach developed in this thesis is innovative because it is predominantly based on observations. The estimation of the terms controlling the one-hour surface temperature variations could help to improve the already existing parameterizations of the SEB terms and to better understand their spatial variability as a function of local conditions. Furthermore, a comparison between multi-model regional climate simulations and these estimations can be performed to evaluate if the simulations are able to well reproduce these behaviors, in particular in a warming climate where these processes are expected to change.

	Winter	Spring	Summer	Fall
				
	Cirrus	Cirrus Cumulus	Altostratus	Cirrus Altostratus
Clear-sky radiative term	<b>+1.2 °C/h</b>	<b>+2.2 °C/h</b>	<b>+2.2 °C/h</b>	<b>+1.6 °C/h</b>
Cloud radiative term	-0.5 °C/h	-0.7 °C/h	-0.6 °C/h	-0.5 °C/h
Atmospheric term	-0.3 °C/h	-0.8 °C/h	-0.5 °C/h	-0.4 °C/h
Ground heat exchange term	-0.01 °C/h	-0.1 °C/h	-0.2 °C/h	-0.1 °C/h
Advection	<b>+0.01 °C/h</b>	-0.02 °C/h	-0.03 °C/h	0.02 °C/h

**Table 6-i:** First line: Predominance of clouds detected by the lidar during daytime for each season of the year. Second to sixth line: mean contribution values of each term of OTEM for each season. Highlighted values correspond to both the maximum of positive and negative values for the season.

Terms controlling temperature variations	Early morning	Late morning	Early afternoon	Late afternoon
Clear-sky radiative term				
Cloud radiative term				
Atmospheric term				
Ground heat exchange term				
Advection				

**Table 6-ii:** Importance of each term of OTEM along the day at SIRT. Five cases mean the term is the most important at the time considered, whereas one case means is the less important



## REFERENCES

---

- Al-Hinti, I., Al-Muhtady, A., and Al-Kouz, W.: Measurement and modelling of the ground temperature profile in Zarqa, Jordan for geothermal heat pump applications, *Applied Thermal Engineering*, 123, 131–137, <https://doi.org/10.1016/j.applthermaleng.2017.05.107>, 2017.
- Allan, R. P.: Combining satellite data and models to estimate cloud radiative effect at the surface and in the atmosphere, 18, 324–333, <https://doi.org/10.1002/met.285>, 2011.
- Allan, R. P., Liu, C., Loeb, N. G., Palmer, M. D., Roberts, M., Smith, D., and Vidale, P.-L.: Changes in global net radiative imbalance 1985–2012, *Geophys Res Lett*, 41, 5588–5597, <https://doi.org/10.1002/2014GL060962>, 2014.
- Alvalá, R., Gielow, R., Rocha, H., Freitas, H., Lopes, J., Manzi, A., Von Randow, C., Dias, M., Cabral, O., and Waterloo, M. J.: Intradiurnal and seasonal variability of soil temperature, heat flux, soil moisture content, and thermal properties under forest and pasture in Rondonia, *Journal of Geophysical Research Atmospheres*, 107, <https://doi.org/10.1029/2001JD000599>, 2002.
- Archer, K. J. and Kimes, R. V.: Empirical characterization of random forest variable importance measures, *Computational Statistics & Data Analysis*, 52, 2249–2260, <https://doi.org/10.1016/j.csda.2007.08.015>, 2008.
- Arkin, P. A. and Meisner, B. N.: The Relationship between Large-Scale Convective Rainfall and Cold Cloud over the Western Hemisphere during 1982–84, *Mon. Wea. Rev.*, 115, 51–74, [https://doi.org/10.1175/1520-0493\(1987\)115<0051:TRBLSC>2.0.CO;2](https://doi.org/10.1175/1520-0493(1987)115<0051:TRBLSC>2.0.CO;2), 1987.
- Arnold, N. S., Willis, I. C., Sharp, M. J., Richards, K. S., and Lawson, W. J.: A distributed surface energy-balance model for a small valley glacier. I. Development and testing for Haut Glacier d' Arolla, Valais, Switzerland, 42, 77–89, <https://doi.org/10.3189/S0022143000030549>, 1996.
- Arya, S. P. (Ed.): Chapter 10 Neutral Boundary Layers, in: *International Geophysics*, vol. 42, Academic Press, 141–156, [https://doi.org/10.1016/S0074-6142\(08\)60425-8](https://doi.org/10.1016/S0074-6142(08)60425-8), 1988.
- Barlow, J., Best, M., Bohnenstengel, S. I., Clark, P., Grimmond, S., Lean, H., Christen, A., Emeis, S., Haeffelin, M., Harman, I. N., Lemonsu, A., Martilli, A., Pardyjak, E., Rotach, M. W., Ballard, S., Boutle, I., Brown, A., Cai, X., Carpentieri, M., Coceal, O., Crawford, B., Sabatino, S. D., Dou, J., Drew, D. R., Edwards, J. M., Fallmann, J., Fortuniak, K., Gornall, J., Gronemeier, T., Halios, C. H., Hertwig, D., Hirano, K., Holtslag, A. A. M., Luo, Z., Mills, G.,

Nakayoshi, M., Pain, K., Schlünzen, K. H., Smith, S., Soulhac, L., Steeneveld, G.-J., Sun, T., Theeuwes, N. E., Thomson, D., Voogt, J. A., Ward, H. C., Xie, Z.-T., and Zhong, J.: Developing a Research Strategy to Better Understand, Observe, and Simulate Urban Atmospheric Processes at Kilometer to Subkilometer Scales, 98, ES261–ES264, <https://doi.org/10.1175/BAMS-D-17-0106.1>, 2017.

Barnett, D. N., Brown, S. J., Murphy, J. M., Sexton, D. M. H., and Webb, M. J.: Quantifying uncertainty in changes in extreme event frequency in response to doubled CO<sub>2</sub> using a large ensemble of GCM simulations, *Clim Dyn*, 26, 489–511, <https://doi.org/10.1007/s00382-005-0097-1>, 2006.

Bartolini, E., Claps, P., and D’Odorico, P.: Interannual variability of winter precipitation in the European Alps: relations with the North Atlantic Oscillation., 13, 17–25, <https://doi.org/10.5194/hess-13-17-2009>, 2009.

Bastin, S., Chiriaco, M., and Drobinski, P.: Control of radiation and evaporation on temperature variability in a WRF regional climate simulation: comparison with colocated long term ground based observations near Paris, *Clim Dyn*, 51, 985–1003, <https://doi.org/10.1007/s00382-016-2974-1>, 2018.

Bastin, S., Drobinski, P., Chiriaco, M., Bock, O., Roehrig, R., Gallardo, C., Conte, D., Domínguez Alonso, M., Li, L., Lionello, P., and Parracho, A. C.: Impact of humidity biases on light precipitation occurrence: observations versus simulations, 19, 1471–1490, <https://doi.org/10.5194/acp-19-1471-2019>, 2019.

Belonenko, T. V., Bashmachnikov, I. L., and Kubryakov, A. A.: Horizontal advection of temperature and salinity by Rossby waves in the North Pacific, 39, 2177–2188, <https://doi.org/10.1080/01431161.2017.1420932>, 2018.

Bennartz, R., Shupe, M. D., Turner, D. D., Walden, V. P., Steffen, K., Cox, C. J., Kulie, M. S., Miller, N. B., and Pettersen, C.: July 2012 Greenland melt extent enhanced by low-level liquid clouds, *Nature*, 496, 83–86, <https://doi.org/10.1038/nature12002>, 2013.

Betts, A. K.: Seasonal climate transitions in New England: Seasonal climate transitions in New England, *Weather*, 66, 245–248, <https://doi.org/10.1002/wea.754>, 2011.

Betts, A. K., Desjardins, R. L., and Worth, D. E.: The impact of clouds, land use and snow cover on climate in the Canadian Prairies, *Adv. Sci. Res.*, 13, 37–42, <https://doi.org/10.5194/asr-13-37-2016>, 2016.

Boé, J.: Modulation of soil moisture–precipitation interactions over France by large scale circulation, *Clim Dyn*, 40, 875–892, <https://doi.org/10.1007/s00382-012-1380-6>, 2013.

Bretherton, C. and Hartmann, D.: Large-Scale Controls on Cloudiness, Clouds in the Perturbed Climate System: Their Relationship to Energy Balance, Atmospheric Dynamics, and Precipitation, <https://doi.org/10.7551/mitpress/9780262012874.003.0010>, 2009.

Brient, F., Schneider, T., Tan, Z., Bony, S., Qu, X., and Hall, A.: Shallowness of tropical low clouds as a predictor of climate models' response to warming, *Clim Dyn*, 47, 433–449, <https://doi.org/10.1007/s00382-015-2846-0>, 2016.

Broeke, M. V. D., Reijmer, C., As, D. V., and Boot, W.: Daily cycle of the surface energy balance in Antarctica and the influence of clouds, 26, 1587–1605, <https://doi.org/10.1002/joc.1323>, 2006a.

Broeke, M. V. D., Reijmer, C., As, D. V., and Boot, W.: Daily cycle of the surface energy balance in Antarctica and the influence of clouds, 26, 1587–1605, <https://doi.org/10.1002/joc.1323>, 2006b.

Cassou, C., Terray, L., and Phillips, A. S.: Tropical Atlantic influence on European heat waves, 18, 2805–2811, <https://doi.org/10.1175/JCLI3506.1>, 2005.

Caughey, S. J., Wyngaard, J. C., and Kaimal, J. C.: Turbulence in the Evolving Stable Boundary Layer, 36, 1041–1052, [https://doi.org/10.1175/1520-0469\(1979\)036<1041:TITESB>2.0.CO;2](https://doi.org/10.1175/1520-0469(1979)036<1041:TITESB>2.0.CO;2), 1979.

Chen, T., Rossow, W. B., and Zhang, Y.: Radiative Effects of Cloud-Type Variations, 13, 264–286, [https://doi.org/10.1175/1520-0442\(2000\)013<0264:REOCTV>2.0.CO;2](https://doi.org/10.1175/1520-0442(2000)013<0264:REOCTV>2.0.CO;2), 2000.

Chepfer, H., Bony, S., Winker, D., Cesana, G., Dufresne, J. L., Minnis, P., Stubenrauch, C. J., and Zeng, S.: The GCM-Oriented CALIPSO Cloud Product (CALIPSO-GOCCP), 115, <https://doi.org/10.1029/2009JD012251>, 2010.

Cheruy, F., Campoy, A., Dupont, J.-C., Ducharne, A., Hourdin, F., Haeffelin, M., Chiriaco, M., and Idelkadi, A.: Combined influence of atmospheric physics and soil hydrology on the simulated meteorology at the SIRTA atmospheric observatory, *Clim Dyn*, 40, 2251–2269, <https://doi.org/10.1007/s00382-012-1469-y>, 2013.

Cheruy, F., Dufresne, J. L., Hourdin, F., and Ducharne, A.: Role of clouds and land-atmosphere coupling in midlatitude continental summer warm biases and climate change amplification in CMIP5 simulations, 41, 6493–6500, <https://doi.org/10.1002/2014GL061145>, 2014.

Cherviakov, M.: Variability of Earth's radiation budget components during 2009 - 2015 from radiometer IKOR-M data, 18, EPSC2016-9611, 2016.

Chiriaco, M., Bastin, S., Yiou, P., Haeffelin, M., Dupont, J.-C., and Stéfanon, M.: European heatwave in July 2006: Observations and modeling showing how local processes amplify conducive large-scale conditions, 41, 5644–5652, <https://doi.org/10.1002/2014GL060205>, 2014.

Chiriaco, M., Dupont, J.-C., Bastin, S., Badosa, J., Lopez, J., Haeffelin, M., Chepfer, H., and Guzman, R.: ReOBS: a new approach to synthesize long-term multi-variable dataset and application to the SIRTAs supersite, *Earth Syst. Sci. Data*, 10, 919–940, <https://doi.org/10.5194/essd-10-919-2018>, 2018.

Choi, W., Faloon, I. C., McKay, M., Goldstein, A. H., and Baker, B.: Estimating the atmospheric boundary layer height over sloped, forested terrain from surface spectral analysis during BEARPEX, 11, 6837–6853, <https://doi.org/10.5194/acp-11-6837-2011>, 2011.

Chow, F. K., Weigel, A. P., Street, R. L., Rotach, M. W., and Xue, M.: High-Resolution Large-Eddy Simulations of Flow in a Steep Alpine Valley. Part I: Methodology, Verification, and Sensitivity Experiments, 45, 63–86, <https://doi.org/10.1175/JAM2322.1>, 2006.

Cimini, D., Haeffelin, M., Kotthaus, S., Löhnert, U., Martinet, P., O'Connor, E., Walden, C., Coen, M. C., and Preissler, J.: Towards the profiling of the atmospheric boundary layer at European scale—introducing the COST Action PROBE, *Bull. of Atmos. Sci. & Technol.*, 1, 23–42, <https://doi.org/10.1007/s42865-020-00003-8>, 2020.

Conangla, L., Cuxart, J., Jiménez, M. A., Martínez-Villagrasa, D., Miró, J. R., Tabarelli, D., and Zardi, D.: Cold-air pool evolution in a wide Pyrenean valley, 38, 2852–2865, <https://doi.org/10.1002/joc.5467>, 2018.

Copernicus Climate Change Service: ERA5-Land hourly data from 2001 to present, <https://doi.org/10.24381/CDS.E2161BAC>, 2019.

ERA5: Fifth generation of ECMWF atmospheric reanalyses of the global climate. : <https://cds.climate.copernicus.eu/cdsapp#!/home>, last access: 3 February 2020.

Cortesi, N., Torralba, V., González-Reviriego, N., Soret, A., and Doblas-Reyes, F. J.: Characterization of European wind speed variability using weather regimes, *Clim Dyn*, 53, 4961–4976, <https://doi.org/10.1007/s00382-019-04839-5>, 2019.

Dai, A., Trenberth, K. E., and Karl, T. R.: Effects of Clouds, Soil Moisture, Precipitation, and Water Vapor on Diurnal Temperature Range, 12, 23, 1999.

Dang, R., Yang, Y., Hu, X.-M., Wang, Z., and Zhang, S.: A Review of Techniques for Diagnosing the Atmospheric Boundary Layer Height (ABLH) Using Aerosol Lidar Data, 11, 1590, <https://doi.org/10.3390/rs11131590>, 2019.

D'Arrigo, R., Jacoby, G., Wilson, R., and Panagiotopoulos, F.: A reconstructed Siberian High index since A. D. 1599 from Eurasian and North American tree rings, *Geophys. Res. Lett.*, 32, <https://doi.org/10.1029/2004GL022271>, 2005.

Davis, S. M., Hegglin, M. I., Fujiwara, M., Dragani, R., Harada, Y., Kobayashi, C., Long, C., Manney, G. L., Nash, E. R., Potter, G. L., Tegtmeier, S., Wang, T., Wargan, K., and Wright, J. S.: Assessment of upper tropospheric and stratospheric water vapor and ozone in reanalyses as part of S-RIP, 17, 12743–12778, <https://doi.org/10.5194/acp-17-12743-2017>, 2017.

Deardorff, J. W.: Prediction of Convective Mixed-Layer Entrainment for Realistic Capping Inversion Structure, 36, 424–436, [https://doi.org/10.1175/1520-0469\(1979\)036<0424:POCMLE>2.0.CO;2](https://doi.org/10.1175/1520-0469(1979)036<0424:POCMLE>2.0.CO;2), 1979.

Dee, D. P. and Uppala, S.: Variational bias correction of satellite radiance data in the ERA-Interim reanalysis, 135, 1830–1841, <https://doi.org/10.1002/qj.493>, 2009.

Dewitte, S. and Clerbaux, N.: Measurement of the Earth Radiation Budget at the Top of the Atmosphere—A Review, 9, 1143, <https://doi.org/10.3390/rs9111143>, 2017.

Dione, C., Lohou, F., Chiriaco, M., Lothon, M., Bastin, S., Baray, J.-L., Yiou, P., and Colomb, A.: The Influence of Synoptic Circulations and Local Processes on Temperature Anomalies at Three French Observatories, *J. Appl. Meteor. Climatol.*, 56, 141–158, <https://doi.org/10.1175/JAMC-D-16-0113.1>, 2017.

Domeisen, D. I. V., Grams, C. M., and Papritz, L.: The role of North Atlantic–European weather regimes in the surface impact of sudden stratospheric warming events, 1, 373–388, <https://doi.org/10.5194/wcd-1-373-2020>, 2020.

Dufresne, J.-L. and Bony, S.: An Assessment of the Primary Sources of Spread of Global Warming Estimates from Coupled Atmosphere–Ocean Models, 21, 5135–5144, <https://doi.org/10.1175/2008JCLI2239.1>, 2008.

Dupont, J.-C., Haeffelin, M., Protat, A., Bouniol, D., Boyouk, N., and Morille, Y.: Stratus–Fog Formation and Dissipation: A 6-Day Case Study, *Boundary-Layer Meteorol.*, 143, 207–225, <https://doi.org/10.1007/s10546-012-9699-4>, 2012.

Dutton, E. G., Farhadi, A., Stone, R. S., Long, C. N., and Nelson, D. W.: Long-term variations in the occurrence and effective solar transmission of clouds as determined from surface-based total irradiance observations, 109, <https://doi.org/10.1029/2003JD003568>, 2004.

Efthymiadis, D., Goodess, C. M., and Jones, P. D.: Trends in Mediterranean gridded temperature extremes and large-scale circulation influences, 11, 2199–2214, <https://doi.org/10.5194/nhess-11-2199-2011>, 2011.

Fabiano, F., Christensen, H. M., Strommen, K., Athanasiadis, P., Baker, A., Schiemann, R., and Corti, S.: Euro-Atlantic weather Regimes in the PRIMAVERA coupled climate simulations: impact of resolution and mean state biases on model performance, *Clim Dyn*, 54, 5031–5048, <https://doi.org/10.1007/s00382-020-05271-w>, 2020.

Fawole, O. G., Cai, X., Pinker, R. T., and MacKenzie, A. R.: Analysis of Radiative Properties and Direct Radiative Forcing Estimates of Dominant Aerosol Clusters over an Urban-Desert Region in West Africa, *Aerosol Air Qual. Res.*, 19, 38–48, <https://doi.org/10.4209/aaqr.2017.12.0600>, 2019.

Fischer, E. M. and Schär, C.: Future changes in daily summer temperature variability: driving processes and role for temperature extremes, *Clim Dyn*, 33, 917, <https://doi.org/10.1007/s00382-008-0473-8>, 2008.

Fischer, E. M., Rajczak, J., and Schär, C.: Changes in European summer temperature variability revisited, 39, <https://doi.org/10.1029/2012GL052730>, 2012.

Flamant, C. and Pelon, J.: Atmospheric boundary-layer structure over the Mediterranean during a Tramontane event: ATMOSPHERIC ATMOSPHERIC BOUNDARY-LAYER STRUCTURE, Q.J.R. Meteorol. Soc., 122, 1741–1778, <https://doi.org/10.1002/qj.49712253602>, 1996.

Foken, T., Aubinet, M., and Leuning, R.: The Eddy Covariance Method, in: *Eddy Covariance: A Practical Guide to Measurement and Data Analysis*, edited by: Aubinet, M., Vesala, T., and Papale, D., Springer Netherlands, Dordrecht, 1–19, [https://doi.org/10.1007/978-94-007-2351-1\\_1](https://doi.org/10.1007/978-94-007-2351-1_1), 2012.

Folland, C. K., Knight, J., Linderholm, H. W., Fereday, D., Ineson, S., and Hurrell, J. W.: The summer North Atlantic oscillation: Past, present, and future, 22, 1082–1103, <https://doi.org/10.1175/2008JCLI2459.1>, 2009.

Gao, Z., Liu, H., Katul, G. G., and Foken, T.: Non-closure of the surface energy balance explained by phase difference between vertical velocity and scalars of large atmospheric eddies, *Environ. Res. Lett.*, 12, 034025, <https://doi.org/10.1088/1748-9326/aa625b>, 2017.

Gastineau, G., D'Andrea, F., and Frankignoul, C.: Atmospheric response to the North Atlantic Ocean variability on seasonal to decadal time scales, *Clim Dyn*, 40, 2311–2330, <https://doi.org/10.1007/s00382-012-1333-0>, 2013.

Genuer, R., Poggi, J.-M., and Tuleau-Malot, C.: Variable selection using Random Forests, 11, 2012.

Geoffroy, O.: Relationship between drizzle rate, liquid water path and droplet concentration at the scale of a stratocumulus cloud system, 14, 2008.

Gkikas, A., Obiso, V., Pérez García-Pando, C., Jorba, O., Hatzianastassiou, N., Vendrell, L., Basart, S., Solomos, S., Gassó, S., and Baldasano, J. M.: Direct radiative effects during intense Mediterranean desert dust outbreaks, 18, 8757–8787, <https://doi.org/10.5194/acp-18-8757-2018>, 2018.

González, B. B., Santana, J. R., Cachorro, V. E., Toledano, C., Antuña-Marrero, J. C., Arredondo, R. E., and Frutos, Á. M. de: Surface shortwave cloud radiative effect of Cumulus (Cu) and Stratocumulus-Cumulus (Sc-Cu) cloud types in the Caribbean area (Camagüey Cuba, 2010-2016)., 2020.

Grömping, U.: Estimators of Relative Importance in Linear Regression Based on Variance Decomposition, *The American Statistician*, 61, 139–147, <https://doi.org/10.1198/000313007X188252>, 2007.

Haefelin, M., Barthès, L., Bock, O., Boitel, C., Bony, S., Bouniol, D., Chepfer, H., Chiriaco, M., Cuesta, J., Delanoë, J., Drobinski, P., Dufresne, J.-L., Flamant, C., Grall, M., Hodzic, A., Hourdin, F., Lapouge, F., Lemaître, Y., Mathieu, A., Morille, Y., Naud, C., Noël, V., O'Hirok, W., Pelon, J., Pietras, C., Protat, A., Romand, B., Scialom, G., and Vautard, R.: SIRTa, a ground-based atmospheric observatory for cloud and aerosol research, *Ann. Geophys.*, 23, 253–275, <https://doi.org/10.5194/angeo-23-253-2005>, 2005.

Haefelin, M., Angelini, F., Morille, Y., Martucci, G., Frey, S., Gobbi, G. P., Lolli, S., O'Dowd, C. D., Sauvage, L., Xueref-Rémy, I., Wastine, B., and Feist, D. G.: Evaluation of Mixing-Height Retrievals from Automatic Profiling Lidars and Ceilometers in View of Future Integrated Networks in Europe, *Boundary-Layer Meteorol*, 143, 49–75, <https://doi.org/10.1007/s10546-011-9643-z>, 2012.

Hakuba, M. Z., Folini, D., Sanchez-Lorenzo, A., and Wild, M.: Spatial representativeness of ground-based solar radiation measurements, 118, 8585–8597, <https://doi.org/10.1002/jgrd.50673>, 2013.

Hartmann, D. L.: Chapter 6 Radiative Effects of Clouds on Earth's Climate, in: *International Geophysics*, vol. 54, Elsevier, 151–173, [https://doi.org/10.1016/S0074-6142\(08\)60215-6](https://doi.org/10.1016/S0074-6142(08)60215-6), 1993.

Hartmann, D. L. (Ed.): Chapter 4 The Energy Balance of the Surface, in: *International Geophysics*, vol. 56, Academic Press, 81–114, [https://doi.org/10.1016/S0074-6142\(08\)60561-6](https://doi.org/10.1016/S0074-6142(08)60561-6), 1994.

Hartmann, D. L., Ockert-Bell, M. E., and Michelsen, M. L.: The Effect of Cloud Type on Earth's Energy Balance: Global Analysis., *Journal of Climate*, 5, 1281–1304, [https://doi.org/10.1175/1520-0442\(1992\)005<1281:TEOCTO>2.0.CO;2](https://doi.org/10.1175/1520-0442(1992)005<1281:TEOCTO>2.0.CO;2), 1992.

Hennemuth, B. and Lammert-Stockschlaeder, A.: Determination of the Atmospheric Boundary Layer Height from Radiosonde and Lidar Backscatter, *Boundary-Layer Meteorology*, v.120, 181–200 (2006), 120, <https://doi.org/10.1007/s10546-005-9035-3>, 2006.

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. D., 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. de, Rozum, I., Vamborg, F., Villaume, S., and Thépaut, J.-N.: The ERA5 global reanalysis, 146, 1999–2049, <https://doi.org/10.1002/qj.3803>, 2020.

Hoffmann, L., Günther, G., Li, D., Stein, O., Wu, X., Griessbach, S., Heng, Y., Konopka, P., Müller, R., Vogel, B., and Wright, J. S.: From ERA-Interim to ERA5: the considerable impact of ECMWF's next-generation reanalysis on Lagrangian transport simulations, *Atmos. Chem. Phys.*, 19, 3097–3124, <https://doi.org/10.5194/acp-19-3097-2019>, 2019.

Holtzlag, A. a. M. and Van Ulden, A. P.: A Simple Scheme for Daytime Estimates of the Surface Fluxes from Routine Weather Data, *J. Climate Appl. Meteor.*, 22, 517–529, [https://doi.org/10.1175/1520-0450\(1983\)022<0517:ASSFDE>2.0.CO;2](https://doi.org/10.1175/1520-0450(1983)022<0517:ASSFDE>2.0.CO;2), 1983.

Horányi, A.: Some aspects on the use and impact of observations in the ERA5 copernicus climate change service reanalysis, *Idojaras*, 121, 329–344, 2017.

Hurrell, J. W., Kushnir, Y., Ottersen, G., and Visbeck, M.: An overview of the North Atlantic Oscillation, in: *Geophysical Monograph Series*, vol. 134, edited by: Hurrell, J. W., Kushnir, Y., Ottersen, G., and Visbeck, M., American Geophysical Union, Washington, D. C., 1–35, <https://doi.org/10.1029/134GM01>, 2003.

Ionita, M., Boroneanț, C., and Chelcea, S.: Seasonal modes of dryness and wetness variability over Europe and their connections with large scale atmospheric circulation and global sea surface temperature, *Clim Dyn*, 45, 2803–2829, <https://doi.org/10.1007/s00382-015-2508-2>, 2015.

Jacobs, A. F. G., Heusinkveld, B. G., and Holtzlag, A. a. M.: Long-term record and analysis of soil temperatures and soil heat fluxes in a grassland area, *The Netherlands*, 151, 774–780, <https://doi.org/10.1016/j.agrformet.2011.01.002>, 2011.

James, G., Witten, D., Hastie, T., and Tibshirani, R.: An Introduction to Statistical Learning, Springer New York, New York, NY, <https://doi.org/10.1007/978-1-4614-7138-7>, 2013.

Jegade, O., Foken, T., Balogun, A., and Abimbola, O.: Bowen ratio determination of sensible and latent heat fluxes in a humid tropical environment at Ile-Ife, Nigeria, *MAU-SAM*, 52, 669–678, 2001.

Jia, R., Liu, Y., Chen, B., Zhang, Z., and Huang, J.: Source and transportation of summer dust over the Tibetan Plateau, *Atmospheric Environment*, 123, 210–219, <https://doi.org/10.1016/j.atmosenv.2015.10.038>, 2015.

Kato, S., Loeb, N. G., Rutan, D. A., Rose, F. G., Sun-Mack, S., Miller, W. F., and Chen, Y.: Uncertainty Estimate of Surface Irradiances Computed with MODIS-, CALIPSO-, and CloudSat-Derived Cloud and Aerosol Properties, *Surv Geophys*, 33, 395–412, <https://doi.org/10.1007/s10712-012-9179-x>, 2012.

Kauppinen, J., Heinonen, J., and Malmi, P.: Influence of Relative Humidity and Clouds on the Global Mean Surface Temperature, *Energy & Environment*, 25, 389–399, <https://doi.org/10.1260/0958-305X.25.2.389>, 2014.

Kienast-Sjögren, E., Rolf, C., Seifert, P., Krieger, U. K., Luo, B. P., Krämer, M., and Peter, T.: Climatological and radiative properties of midlatitude cirrus clouds derived by automatic evaluation of lidar measurements, *Atmos. Chem. Phys.*, 16, 7605–7621, <https://doi.org/10.5194/acp-16-7605-2016>, 2016.

Klavins, M. and Rodinov, V.: Influence of large-scale atmospheric circulation on climate in Latvia, 15, 11, 2010.

Korhonen, K., Giannakaki, E., Mielonen, T., Pfüller, A., Laakso, L., Vakkari, V., Baars, H., Engelmann, R., Beukes, J. P., Van Zyl, P. G., Ramandh, A., Ntsangwane, L., Josipovic, M., Tiitta, P., Fourie, G., Ngwana, I., Chiloane, K., and Komppula, M.: Atmospheric boundary layer top height in South Africa: measurements with lidar and radiosonde compared to three atmospheric models, *Atmos. Chem. Phys.*, 14, 4263–4278, <https://doi.org/10.5194/acp-14-4263-2014>, 2014.

Kotthaus, S. and Grimmond, C. S. B.: Atmospheric boundary-layer characteristics from ceilometer measurements. Part 1: A new method to track mixed layer height and classify clouds, 144, 1525–1538, <https://doi.org/10.1002/qj.3299>, 2018a.

Kotthaus, S. and Grimmond, C. S. B.: Atmospheric boundary-layer characteristics from ceilometer measurements. Part 2: Application to London's urban boundary layer, 144, 1511–1524, <https://doi.org/10.1002/qj.3298>, 2018b.

Kukla, G. and Karl, T. R.: Nighttime warming and the greenhouse effect, *Environ. Sci. Technol.*, 27, 1468–1474, <https://doi.org/10.1021/es00045a001>, 1993.

Lange, D., Rocadenbosch, F., Tiana-Alsina, J., and Frasier, S.: Atmospheric Boundary Layer Height Estimation Using a Kalman Filter and a Frequency-Modulated Continuous-Wave Radar, 53, 3338–3349, <https://doi.org/10.1109/TGRS.2014.2374233>, 2015.

Leuning, R., van Gorsel, E., Massman, W. J., and Isaac, P. R.: Reflections on the surface energy imbalance problem, *Agricultural and Forest Meteorology*, 156, 65–74, <https://doi.org/10.1016/j.agrformet.2011.12.002>, 2012.

Li, Z. and Leighton, H. G.: Global climatologies of solar radiation budgets at the surface and in the atmosphere from 5 years of ERBE data, 98, 4919–4930, <https://doi.org/10.1029/93JD00003>, 1993.

Liebenthal, C. and Foken, T.: Evaluation of six parameterization approaches for the ground heat flux, *Theor. Appl. Climatol.*, 88, 43–56, <https://doi.org/10.1007/s00704-005-0234-0>, 2007.

Lisok, J., Rozwadowska, A., Pedersen, J. G., Markowicz, K. M., Ritter, C., Kaminski, J. W., Struzewska, J., Mazzola, M., Udisti, R., Becagli, S., and Gorecka, I.: Radiative impact of an extreme Arctic biomass-burning event, 18, 8829–8848, <https://doi.org/10.5194/acp-18-8829-2018>, 2018.

Llasat, M. C. and Puigcerver, M.: Cold air pools over Europe, *Meteorol. Atmos. Phys.*, 42, 171–177, <https://doi.org/10.1007/BF01314823>, 1990.

Lolli, S., Campbell, J. R., Lewis, J. R., Welton, E. J., Girolamo, P. D., Fatkhuroyan, F., Gu, Y., and Marquis, J. W.: Assessment of cirrus cloud and aerosol radiative effect in South-East Asia by ground-based NASA MPLNET lidar network data and CALIPSO satellite measurements, in: *Remote Sensing of Clouds and the Atmosphere XXII*, Remote Sensing of Clouds and the Atmosphere XXII, 1042405, <https://doi.org/10.1117/12.2278987>, 2017.

Long, C. N. and Turner, D. D.: A method for continuous estimation of clear-sky downwelling longwave radiative flux developed using ARM surface measurements, 113, <https://doi.org/10.1029/2008JD009936>, 2008.

Long, C. S., Fujiwara, M., Davis, S., Mitchell, D. M., and Wright, C. J.: Climatology and interannual variability of dynamic variables in multiple reanalyses evaluated by the SPARC Reanalysis Intercomparison Project (S-RIP), 17, 14593–14629, <https://doi.org/10.5194/acp-17-14593-2017>, 2017.

Luo, T., Yuan, R., and Wang, Z.: Lidar-based remote sensing of atmospheric boundary layer height over land and ocean, 7, 173–182, <https://doi.org/10.5194/amt-7-173-2014>, 2014.

Malardel, S.: *Fondamentaux de Météorologie* 2e édition, 2e ed., Cépaduès, 2009.

Mallet, M., Tulet, P., Serça, D., Solmon, F., Dubovik, O., Pelon, J., Pont, V., and Thouron, O.: Impact of dust aerosols on the radiative budget, surface heat fluxes, heating rate profiles and convective activity over West Africa during March 2006, 9, 7143–7160, <https://doi.org/10.5194/acp-9-7143-2009>, 2009.

Manninen, A. J., Marke, T., Tuononen, M., and O'Connor, E. J.: Atmospheric Boundary Layer Classification With Doppler Lidar, 123, 8172–8189, <https://doi.org/10.1029/2017JD028169>, 2018.

Markowicz, K. M., Zawadzka-Manko, O., Lisok, J., Chilinski, M. T., and Xian, P.: The impact of moderately absorbing aerosol on surface sensible, latent, and net radiative fluxes during the summer of 2015 in Central Europe, *Journal of Aerosol Science*, 151, 105627, <https://doi.org/10.1016/j.jaerosci.2020.105627>, 2021.

Marks, D., Winstral, A., Flerchinger, G., Reba, M., Pomeroy, J., Link, T., and Elder, K.: Comparing Simulated and Measured Sensible and Latent Heat Fluxes over Snow under a Pine Canopy to Improve an Energy Balance Snowmelt Model, 9, 1506–1522, <https://doi.org/10.1175/2008JHM874.1>, 2008.

Maruyama, T. and Segawa, M.: Reciprocal Analysis of Sensible and Latent Heat Fluxes in a Forest Region Using Single Height Temperature and Humidity Based on the Bowen Ratio Concept, 08, 724, <https://doi.org/10.4236/jwarp.2016.87059>, 2016.

Matsueda, M. and Palmer, T. N.: Estimates of flow-dependent predictability of winter-time Euro-Atlantic weather regimes in medium-range forecasts, 144, 1012–1027, <https://doi.org/10.1002/qj.3265>, 2018.

Mauder, M., Jegede, O. O., Okogbue, E. C., Wimmer, F., and Foken, T.: Surface energy balance measurements at a tropical site in West Africa during the transition from dry to wet season, *Theor. Appl. Climatol.*, 89, 171–183, <https://doi.org/10.1007/s00704-006-0252-6>, 2007.

Mauno, P., McFarquhar, G. M., Räisänen, P., Kahnert, M., Timlin, M. S., and Nousiainen, T.: The influence of observed cirrus microphysical properties on shortwave radiation: A case study over Oklahoma, 116, <https://doi.org/10.1029/2011JD016058>, 2011.

McCaffrey, K., Wilczak, J. M., Bianco, L., Gritmit, E., Sharp, J., Banta, R., Friedrich, K., Fernando, H. J. S., Krishnamurthy, R., Leo, L. S., and Muradyan, P.: Identification and Characterization of Persistent Cold Pool Events from Temperature and Wind Profilers in the Columbia River Basin, 58, 2533–2551, <https://doi.org/10.1175/JAMC-D-19-0046.1>, 2019.

McNider, R. T.: Response and Sensitivity of the Nocturnal Boundary Layer Over Land to Added Longwave Radiative Forcing, 6, 2011.

McNider, R. T., Steeneveld, G. J., Holtslag, A. a. M., Pielke Sr., R. A., Mackaro, S., Pour-Biazar, A., Walters, J., Nair, U., and Christy, J.: Response and sensitivity of the nocturnal boundary layer over land to added longwave radiative forcing, 117, <https://doi.org/10.1029/2012JD017578>, 2012.

Ménégoz, M., Guemas, V., Salas y Melia, D., and Voldoire, A.: Winter interactions between aerosols and weather regimes in the North Atlantic European region, 115, <https://doi.org/10.1029/2009JD012480>, 2010.

Miao, S., Dou, J., Chen, F., Li, J., and Li, A.: Analysis of observations on the urban surface energy balance in Beijing, *Sci. China Earth Sci.*, 55, 1881–1890, <https://doi.org/10.1007/s11430-012-4411-6>, 2012.

Miller, N. B., Shupe, M. D., Cox, C. J., Noone, D., Persson, P. O. G., and Steffen, K.: Surface energy budget responses to radiative forcing at Summit, Greenland, 11, 497–516, <https://doi.org/10.5194/tc-11-497-2017>, 2017.

Morille, Y., Haeffelin, M., Drobinski, P., and Pelon, J.: STRAT: An Automated Algorithm to Retrieve the Vertical Structure of the Atmosphere from Single-Channel Lidar Data, 24, 761–775, <https://doi.org/10.1175/JTECH2008.1>, 2007.

Moura, D., Gomes, A., Mendes, I., and Anibal, J.: Guadiana River Estuary - Investigating the past, presente and future, 2017.

Nieuwstadt, F. T. M.: Some Aspects of the Turbulent Stable Boundary Layer, in: *Boundary Layer Structure: Modeling and Application to Air Pollution and Wind Energy*, edited by: Kaplan, H. and Dinar, N., Springer Netherlands, Dordrecht, 31–55, [https://doi.org/10.1007/978-94-009-6514-0\\_2](https://doi.org/10.1007/978-94-009-6514-0_2), 1984.

Ningombam, S. S., Bagare, S. P., Khatri, P., Sohn, B. J., and Song, H.-J.: Estimation of aerosol radiative forcing over an aged-background aerosol feature during advection and non-advection events using a ground-based data obtained from a Prede Skyradiometer observation, *Atmospheric Research*, 164–165, 76–83, <https://doi.org/10.1016/j.atmosres.2015.05.001>, 2015.

Nuijens, L., Medeiros, B., Sandu, I., and Ahlgrimm, M.: The behavior of trade-wind cloudiness in observations and models: The major cloud components and their variability, 7, 600–616, <https://doi.org/10.1002/2014MS000390>, 2015.

Obermann, A., Bastin, S., Belamari, S., Conte, D., Gaertner, M. A., Li, L., and Ahrens, B.: Mistral and Tramontane wind speed and wind direction patterns in regional climate simulations, *Clim Dyn*, 51, 1059–1076, <https://doi.org/10.1007/s00382-016-3053-3>, 2018.

Offerle, B., Eliasson, I., Grimmond, C. S. B., and Holmer, B.: Surface heating in relation to air temperature, wind and turbulence in an urban street canyon, *Boundary-Layer Meteorol*, 122, 273–292, <https://doi.org/10.1007/s10546-006-9099-8>, 2007.

Oncley, S. P., Foken, T., Vogt, R., Kohsiek, W., DeBruin, H. A. R., Bernhofer, C., Christen, A., Gorsel, E. van, Grantz, D., Feigenwinter, C., Lehner, I., Liebenthal, C., Liu, H., Mauder, M., Pitacco, A., Ribeiro, L., and Weidinger, T.: The Energy Balance Experiment EBEX-2000. Part I: overview and energy balance, *Boundary-Layer Meteorol*, 123, 1–28, <https://doi.org/10.1007/s10546-007-9161-1>, 2007.

Pal, S. and Haeffelin, M.: Forcing mechanisms governing diurnal, seasonal, and interannual variability in the boundary layer depths: Five years of continuous lidar observations over a suburban site near Paris, 120, 11,936–11,956, <https://doi.org/10.1002/2015JD023268>, 2015.

Pal, S., Haeffelin, M., and Batchvarova, E.: Exploring a geophysical process-based attribution technique for the determination of the atmospheric boundary layer depth using aerosol lidar and near-surface meteorological measurements, 118, 9277–9295, <https://doi.org/10.1002/jgrd.50710>, 2013.

Palter, J. B.: The Role of the Gulf Stream in European Climate, *Annu. Rev. Mar. Sci.*, 7, 113–137, <https://doi.org/10.1146/annurev-marine-010814-015656>, 2015.

Panofsky, H. A. and Dutton, J. A.: *Atmospheric Turbulence: Models and Methods for Engineering Applications*, John Wiley & Sons Inc, New York, 418 pp., 1984.

Parding, K., Olseth, J. A., Dagestad, K. F., and Liepert, B. G.: Decadal variability of clouds, solar radiation and temperature at a high-latitude coastal site in Norway, 66, 25897, <https://doi.org/10.3402/tellusb.v66.25897>, 2014.

Peings, Y. and Magnusdottir, G.: Forcing of the wintertime atmospheric circulation by the multidecadal fluctuations of the North Atlantic ocean, *Environ. Res. Lett.*, 9, 034018, <https://doi.org/10.1088/1748-9326/9/3/034018>, 2014.

Péré, J. C., Mallet, M., Pont, V., and Bessagnet, B.: Impact of aerosol direct radiative forcing on the radiative budget, surface heat fluxes, and atmospheric dynamics during the heat wave of summer 2003 over western Europe: A modeling study, 116, <https://doi.org/10.1029/2011JD016240>, 2011.

Pietroni, I., Argentini, S., Petenko, I., and Sozzi, R.: Measurements and Parametrizations of the Atmospheric Boundary-Layer Height at Dome C, Antarctica, *Boundary-Layer Meteorol*, 143, 189–206, <https://doi.org/10.1007/s10546-011-9675-4>, 2012.

Pinardi, N. and Masetti, E.: Variability of the large scale general circulation of the Mediterranean Sea from observations and modelling: a review, *Palaeogeography, Palaeoclimatology, Palaeoecology*, 158, 153–173, [https://doi.org/10.1016/S0031-0182\(00\)00048-1](https://doi.org/10.1016/S0031-0182(00)00048-1), 2000.

Popiel, C. O. and Wojtkowiak, J.: Temperature distributions of ground in the urban region of Poznan City, *Experimental Thermal and Fluid Science*, 51, 135–148, <https://doi.org/10.1016/j.expthermflusci.2013.07.009>, 2013.

Powell, S., Houze, R., Kumar, A., and McFarlane, S.: Comparison of Simulated and Observed Continental Tropical Anvil Clouds and Their Radiative Heating Profiles, *Journal of the Atmospheric Sciences*, 69, <https://doi.org/10.1175/JAS-D-11-0251.1>, 2012.

Raschke, E., Ohmura, A., Rossow, W. B., Carlson, B. E., Zhang, Y.-C., Stubenrauch, C., Kottek, M., and Wild, M.: Cloud effects on the radiation budget based on ISCCP data (1991 to 1995), 25, 1103–1125, <https://doi.org/10.1002/joc.1157>, 2005.

Raymond, F., Ullmann, A., Camberlin, P., Oueslati, B., and Drobinski, P.: Atmospheric conditions and weather regimes associated with extreme winter dry spells over the Mediterranean basin, *Clim Dyn*, 50, 4437–4453, <https://doi.org/10.1007/s00382-017-3884-6>, 2018.

Rebetez, M., Dupont, O., and Giroud, M.: An analysis of the July 2006 heatwave extent in Europe compared to the record year of 2003, *Theor Appl Climatol*, 95, 1–7, <https://doi.org/10.1007/s00704-007-0370-9>, 2009.

Rieutord, T., Aubert, S., and Machado, T.: Deriving boundary layer height from aerosol lidar using machine learning: KABL and ADABL algorithms, *Atmos. Meas. Tech.*, 14, 4335–4353, <https://doi.org/10.5194/amt-14-4335-2021>, 2021.

Ringard, J., Chiriaco, M., Bastin, S., and Habets, F.: Recent trends in climate variability at the local scale using 40 years of observations: the case of the Paris region of France, 19, 13129–13155, <https://doi.org/10.5194/acp-19-13129-2019>, 2019.

Rojas, M., Li, L. Z., Kanakidou, M., Hatzianastassiou, N., Seze, G., and Le Treut, H.: Winter weather regimes over the Mediterranean region: their role for the regional climate and projected changes in the twenty-first century, *Clim Dyn*, 41, 551–571, <https://doi.org/10.1007/s00382-013-1823-8>, 2013a.

Rojas, M., Li, L. Z., Kanakidou, M., Hatzianastassiou, N., Seze, G., and Le Treut, H.: Winter weather regimes over the Mediterranean region: their role for the regional climate and projected changes in the twenty-first century, *Clim Dyn*, 41, 551–571, <https://doi.org/10.1007/s00382-013-1823-8>, 2013b.

Rojas Muñoz, O. J., Chiriaco, M., Bastin, S., and Ringard, J.: Estimation of the terms acting on local 1&thinsp;h surface temperature variations in Paris region: the specific contribution of clouds, 21, 15699–15723, <https://doi.org/10.5194/acp-21-15699-2021>, 2021.

Russell, E. S., Liu, H., Gao, Z., Finn, D., and Lamb, B.: Impacts of soil heat flux calculation methods on the surface energy balance closure, *Agricultural and Forest Meteorology*, 214–215, 189–200, <https://doi.org/10.1016/j.agrformet.2015.08.255>, 2015.

SanchezGomez, E. and Terray, L.: Large-scale atmospheric dynamics and local intense precipitation episodes, 32, <https://doi.org/10.1029/2005GL023990>, 2005.

Santanello, J. A., Peters-Lidard, C. D., and Kumar, S. V.: Diagnosing the Sensitivity of Local Land–Atmosphere Coupling via the Soil Moisture–Boundary Layer Interaction, 12, 766–786, <https://doi.org/10.1175/JHM-D-10-05014.1>, 2011.

Santos, J. a., Corte-Real, J., and Leite, S. m.: Weather regimes and their connection to the winter rainfall in Portugal, 25, 33–50, <https://doi.org/10.1002/joc.1101>, 2005.

Sauter, T. and Galos, S. P.: Effects of local advection on the spatial sensible heat flux variation on a mountain glacier, 10, 2887–2905, <https://doi.org/10.5194/tc-10-2887-2016>, 2016.

Schlögl, S., Lehning, M., and Mott, R.: How Are Turbulent Sensible Heat Fluxes and Snow Melt Rates Affected by a Changing Snow Cover Fraction?, *Frontiers in Earth Science*, 6, <https://doi.org/10.3389/feart.2018.00154>, 2018.

Sedlar, J., Tjernström, M., Mauritsen, T., Shupe, M., Brooks, I., Persson, O., Birch, C., Leck, C., Sirevaag, A., and Nicolaus, M.: A transitioning Arctic surface energy budget: The impacts of solar zenith angle, surface albedo and cloud radiative forcing, *Climate Dynamics*, v.11 (2010), 37, <https://doi.org/10.1007/s00382-010-0937-5>, 2010.

Sherwood, S. C., Bony, S., and Dufresne, J.-L.: Spread in model climate sensitivity traced to atmospheric convective mixing, 505, 37–42, <https://doi.org/10.1038/nature12829>, 2014.

Shi, X., McNider, R. T., Singh, M. P., England, D. E., Friedman, M. J., Lapenta, W. M., and Norris, W. B.: On the Behavior of the Stable Boundary Layer and the Role of Initial Conditions, *Pure appl. geophys.*, 162, 1811–1829, <https://doi.org/10.1007/s00024-005-2694-7>, 2005.

Solomos, S., Kalivitis, N., Mihalopoulos, N., Amiridis, V., Kouvarakis, G., Gkikas, A., Biniotoglou, I., Tsekeri, A., Kazadzis, S., Kottas, M., Pradhan, Y., Proestakis, E., Nastos, P. T., and Marenco, F.: From Tropospheric Folding to Khamsin and Foehn Winds: How Atmospheric Dynamics Advanced a Record-Breaking Dust Episode in Crete, 9, 240, <https://doi.org/10.3390/atmos9070240>, 2018.

Sousa, P. M., Trigo, R. M., Barriopedro, D., Soares, P. M. M., and Santos, J. A.: European temperature responses to blocking and ridge regional patterns, *Clim Dyn*, 50, 457–477, <https://doi.org/10.1007/s00382-017-3620-2>, 2018a.

Sousa, P. M., Trigo, R. M., Barriopedro, D., Soares, P. M. M., and Santos, J. A.: European temperature responses to blocking and ridge regional patterns, *Clim Dyn*, 50, 457–477, <https://doi.org/10.1007/s00382-017-3620-2>, 2018b.

Stapf, J., Ehrlich, A., Jäkel, E., Lüpkes, C., and Wendisch, M.: Reassessment of shortwave surface cloud radiative forcing in the Arctic: consideration of surface-albedo–cloud interactions, 20, 9895–9914, <https://doi.org/10.5194/acp-20-9895-2020>, 2020.

Strobl, C., Boulesteix, A.-L., Zeileis, A., and Hothorn, T.: Bias in random forest variable importance measures: Illustrations, sources and a solution, *BMC Bioinformatics*, 8, 25, <https://doi.org/10.1186/1471-2105-8-25>, 2007.

Stull, R. B.: *An Introduction to Boundary Layer Meteorology*, Springer Netherlands, <https://doi.org/10.1007/978-94-009-3027-8>, 1988.

Tian, J., Dong, X., Xi, B., Williams, C. R., and Wu, P.: Estimation of liquid water path below the melting layer in stratiform precipitation systems using radar measurements during MC3E, *Atmos. Meas. Tech.*, 12, 3743–3759, <https://doi.org/10.5194/amt-12-3743-2019>, 2019.

Toledo, D., Córdoba-Jabonero, C., and Gil-Ojeda, M.: Cluster Analysis: A New Approach Applied to Lidar Measurements for Atmospheric Boundary Layer Height Estimation, 31, 422–436, <https://doi.org/10.1175/JTECH-D-12-00253.1>, 2014.

Tomczyk, A. M., Bednorz, E., Pórolniczak, M., and Kolendowicz, L.: Strong heat and cold waves in Poland in relation with the large-scale atmospheric circulation, *Theor Appl Climatol*, 137, 1909–1923, <https://doi.org/10.1007/s00704-018-2715-y>, 2019.

Trigo, R. M., Osborn, T. J., and Corte-Real, J. M.: The North Atlantic Oscillation influence on Europe: climate impacts and associated physical mechanisms, 20, 9–17, 2002.

Van Tricht, K., Lhermitte, S., Lenaerts, J. T. M., Gorodetskaya, I. V., L'Ecuyer, T. S., Noël, B., van den Broeke, M. R., Turner, D. D., and van Lipzig, N. P. M.: Clouds enhance Greenland ice sheet meltwater runoff, *Nat Commun*, 7, 10266, <https://doi.org/10.1038/ncomms10266>, 2016.

Vautard, R. and Yiou, P.: Control of recent European surface climate change by atmospheric flow, 36, <https://doi.org/10.1029/2009GL040480>, 2009.

Vial, J., Dufresne, J.-L., and Bony, S.: On the interpretation of inter-model spread in CMIP5 climate sensitivity estimates, *Clim Dyn*, 41, 3339–3362, <https://doi.org/10.1007/s00382-013-1725-9>, 2013.

Vihma, T., Graversen, R., Chen, L., Handorf, D., Skific, N., Francis, J. A., Tyrrell, N., Hall, R., Hanna, E., Uotila, P., Dethloff, K., Karpechko, A. Y., Björnsson, H., and Overland, J. E.: Effects of the tropospheric large-scale circulation on European winter temperatures during the period of amplified Arctic warming, 40, 509–529, <https://doi.org/10.1002/joc.6225>, 2020.

Vivone, G., D'Amico, G., Summa, D., Lolli, S., Amodeo, A., Bortoli, D., and Pappalardo, G.: Atmospheric boundary layer height estimation from aerosol lidar: a new approach based on morphological image processing techniques, 21, 4249–4265, <https://doi.org/10.5194/acp-21-4249-2021>, 2021.

Wallace, J. and Hobbs, P.: *Atmospheric Science - An Introduction Survey*, 2nd ed., Academic Press, 504 pp., 2006.

Walters, J. T., McNider, R. T., Shi, X., Norris, W. B., and Christy, J. R.: Positive surface temperature feedback in the stable nocturnal boundary layer, 34, <https://doi.org/10.1029/2007GL029505>, 2007.

Wang, K. and Dickinson, R. E.: Contribution of solar radiation to decadal temperature variability over land, *PNAS*, 110, 14877–14882, <https://doi.org/10.1073/pnas.1311433110>, 2013.

Wang, M. and Penner, J. E.: Cirrus clouds in a global climate model with a statistical cirrus cloud scheme, 10, 5449–5474, <https://doi.org/10.5194/acp-10-5449-2010>, 2010.

Wang, Y., Xie, S.-P., Wang, B., and Xu, H.: Large-Scale Atmospheric Forcing by Southeast Pacific Boundary Layer Clouds: A Regional Model Study, 18, 934–951, <https://doi.org/10.1175/JCLI3302.1>, 2005.

Wang, Z.-H. and Bou-Zeid, E.: A novel approach for the estimation of soil ground heat flux, *Agricultural and Forest Meteorology*, 154–155, 214–221, <https://doi.org/10.1016/j.agrformet.2011.12.001>, 2012.

Watanabe, M., Shiogama, H., Yokohata, T., Kamae, Y., Yoshimori, M., Ogura, T., Annan, J. D., Hargreaves, J. C., Emori, S., and Kimoto, M.: Using a Multiphysics Ensemble for Exploring Diversity in Cloud–Shortwave Feedback in GCMs, 25, 5416–5431, <https://doi.org/10.1175/JCLI-D-11-00564.1>, 2012.

Weigel, A. P. and Rotach, M. W.: Flow structure and turbulence characteristics of the daytime atmosphere in a steep and narrow Alpine valley, 130, 2605–2627, <https://doi.org/10.1256/qj.03.214>, 2004.

Weigel, A. P., Chow, F. K., Rotach, M. W., Street, R. L., and Xue, M.: High-Resolution Large-Eddy Simulations of Flow in a Steep Alpine Valley. Part II: Flow Structure and Heat Budgets, 45, 87–107, <https://doi.org/10.1175/JAM2323.1>, 2006.

Weiner, R. F. and Matthews, R. A. (Eds.): Chapter 18 - Meteorology and Air Pollution, in: *Environmental Engineering (Fourth Edition)*, Butterworth-Heinemann, Burlington, 351–374, <https://doi.org/10.1016/B978-075067294-8/50018-X>, 2003.

Wendisch, M., Macke, A., Ehrlich, A., Lüpkes, C., Mech, M., Chechin, D., Dethloff, K., Velasco, C. B., Bozem, H., Brückner, M., Clemen, H.-C., Crewell, S., Donth, T., Dupuy, R., Ebell, K., Egerer, U., Engelmann, R., Engler, C., Eppers, O., Gehrman, M., Gong, X., Gottschalk, M., Gourbeyre, C., Griesche, H., Hartmann, J., Hartmann, M., Heinold, B., Herber, A., Herrmann, H., Heygster, G., Hoor, P., Jafariserajehlou, S., Jäkel, E., Järvinen, E., Jourdan, O., Kästner, U., Kecorius, S., Knudsen, E. M., Köllner, F., Kretzschmar, J., Lelli, L., Leroy, D., Maturilli, M., Mei, L., Mertes, S., Mioche, G., Neuber, R., Nicolaus, M., Nomokonova, T., Notholt, J., Palm, M., Pinxteren, M. van, Quaas, J., Richter, P., Ruiz-Donoso, E., Schäfer, M., Schmieder, K., Schnaiter, M., Schneider, J., Schwarzenböck, A., Seifert, P., Shupe, M. D., Siebert, H., Spreen, G., Stapf, J., Stratmann, F., Vogl, T., Welti, A., Wex, H., Wiedensohler, A., Zanatta, M., and Zeppenfeld, S.: The Arctic Cloud Puzzle: Using ACLOUD/PASCAL Multiplatform Observations to Unravel the Role of Clouds and Aerosol Particles in Arctic Amplification, 100, 841–871, <https://doi.org/10.1175/BAMS-D-18-0072.1>, 2019.

Wiel, K. van der, Bloomfield, H. C., Lee, R. W., Stoop, L. P., Blackport, R., Screen, J. A., and Selten, F. M.: The influence of weather regimes on European renewable energy production and demand, *Environ. Res. Lett.*, 14, 094010, <https://doi.org/10.1088/1748-9326/ab38d3>, 2019.

Wild, M.: Towards Global Estimates of the Surface Energy Budget, 3, 87–97, <https://doi.org/10.1007/s40641-017-0058-x>, 2017.

- Wild, M., Folini, D., Hakuba, M. Z., Schär, C., Seneviratne, S. I., Kato, S., Rutan, D., Ammann, C., Wood, E. F., and König-Langlo, G.: The energy balance over land and oceans: an assessment based on direct observations and CMIP5 climate models, *Clim Dyn*, 44, 3393–3429, <https://doi.org/10.1007/s00382-014-2430-z>, 2015.
- Willson, R. C., Gulkis, S., Janssen, M., Hudson, H. S., and Chapman, G. A.: Observations of Solar Irradiance Variability, 211, 700–702, <https://doi.org/10.1126/science.211.4483.700>, 1981.
- Wood, R.: Stratocumulus Clouds, 140, 2373–2423, <https://doi.org/10.1175/MWR-D-11-00121.1>, 2012.
- Wulf, H., Mulder, T., Schaepman, M. E., Keller, A., Jörg, P. C., and Schaepman, M. E.: Remote sensing of soils, <https://doi.org/info:doi/10.5167/uzh-109992>, 2015.
- Xoplaki, E., González-Rouco, J., Gyalistras, D., Luterbacher, J., Rickli, R., and Wanner, H.: Interannual summer air temperature variability over Greece and its connection to the large-scale atmospheric circulation and Mediterranean SSTs 1950–1999, *Climate Dynamics*, 20, 537–554, <https://doi.org/10.1007/s00382-002-0291-3>, 2003.
- Xoplaki, E., González-Rouco, J. F., Luterbacher, J., and Wanner, H.: Wet season Mediterranean precipitation variability: influence of large-scale dynamics and trends, *Climate Dynamics*, 23, 63–78, <https://doi.org/10.1007/s00382-004-0422-0>, 2004a.
- Xoplaki, E., González-Rouco, J. F., Luterbacher, J., and Wanner, H.: Wet season Mediterranean precipitation variability: influence of large-scale dynamics and trends, *Climate Dynamics*, 23, 63–78, <https://doi.org/10.1007/s00382-004-0422-0>, 2004b.
- Yao, M.-S. and Cheng, Y.: Cloud Simulations in Response to Turbulence Parameterizations in the GISS Model E GCM, 25, 4963–4974, <https://doi.org/10.1175/JCLI-D-11-00399.1>, 2012.
- Yiou, P., Goubanova, K., Li, Z. X., and Nogaj, M.: Weather regime dependence of extreme value statistics for summer temperature and precipitation, 15, 365–378, <https://doi.org/10.5194/npg-15-365-2008>, 2008.
- Zeng, X., Gong, J., Li, X., and Wu, D. L.: Modeling the Radiative Effect on Microphysics in Cirrus Clouds Against Satellite Observations, 126, e2020JD033923, <https://doi.org/10.1029/2020JD033923>, 2021.
- Zhou, C., Yang, P., Dessler, A. E., Hu, Y., and Baum, B. A.: Study of Horizontally Oriented Ice Crystals with CALIPSO Observations and Comparison with Monte Carlo Radiative Transfer Simulations, 51, 1426–1439, <https://doi.org/10.1175/JAMC-D-11-0265.1>, 2012.



## LIST OF FIGURES

---

<b>Figure 1-1:</b> Explanatory diagram of the (a) negative and (b) positive phases of the North Atlantic Oscillation. This figures is extracted from Moura et al. (2017). .....	19
<b>Figure 1-2:</b> European weather regimes calculated from the daily potential height data at 500 mb isobaric level obtained from a combination of ERA40 and ERA-Interim for the period 1957-2014. The four regimes are the positive (upper left) and the negative (lower right) phases of the NAO, the Scandinavian blocking (upper right) and the Atlantic Ridge (lower left). This figure is extracted from Fabiano et al. (2020). .....	20
<b>Figure 1-3:</b> Vertical cross section of the Earth’s atmosphere showing the two layers in the troposphere. This figure is extracted from Wallace and Hobbs (2006). .....	22
<b>Figure 1-4:</b> Schematic atmospheric boundary layer structure over land in high pressure systems. This figure is extracted from Stull (1988). .....	23
<b>Figure 1-5:</b> Vertical mean profiles of temperature (T), potential temperature ( $\theta$ ), specific humidity (q) and wind speed (V) for (a) day and (b) nighttime. FA stands for free atmosphere, EZ for entrainment zone, ML for mixed layer, SL for surface layer, CI for capping inversion, RL for residual layer, SBL for stable boundary layer, $z_i$ is the height of the capping inversion and $V_g$ is the geostrophic wind. This figure is extracted from Wallace and Hobbs (2006). .....	25
<b>Figure 1-6:</b> Vertical potential temperature profile (left), wind speed (middle), and specific humidity (left) from radiosoundings (dots) launched at 15:08 UTC in the Riviera Valley in the Alps the 08/25/199, and two Large Eddy Simulations (dashed and solid line). This figure is extracted from Chow et al. (2006). .....	26
<b>Figure 1-7:</b> Contributions to the surface energy balance over land. $F_R$ is the radiative fluxes, $F_G$ is the ground heat conduction, $F_A$ is the surface heat fluxes with $F_{sens}$ and $F_{lat}$ (sensible and latent heat fluxes, respectively) and Adv is advection. The radiative fluxes (yellow and red arrows) are set to be positive when arriving to the surface and negative when leaving it. On the contrary, for the latent and sensible heat fluxes and the ground flux, upwards arrows represent a positive sign whereas downwards arrows are for negative signs. Adv is set as a horizontal arrow since it does not play part of the SEB but rather on estimating temperature variations (cf. Chapter 2). .....	28
<b>Figure 1-8:</b> Mean diurnal cycle of the contribution of $FSW \downarrow$ (red line), $FSW \uparrow$ (red dashed line), $FLW \downarrow$ (green line), $FLW \uparrow$ (green dashed line) and $FR$ (black line) radiative fluxes in clear-sky conditions. This figure is extracted from Wallace and Hobbs (2006). .....	30
<b>Figure 1-9:</b> Mean diurnal cycle of the turbulent sensible and latent heat fluxes in the SIRTA observatory from 2003 to 2017, split into seasons. ....	31

**Figure 1-10:** Directly measured SEB terms at Ile-Ife, Nigeria, for the 02/23/2004 with low soil moisture and low nighttime water vapor (left), 03/03/2004 with high soil moisture and low nighttime water vapor (middle), and 03/06/2004 with high soil moisture and high nighttime water vapor (right). This figure is extracted and adapted from Mauder et al. (2007)..... 33

**Figure 1-11:** Diurnal cycle of the four SEB terms ( $F_R$  in black,  $F_{lat}$  in dashed blue,  $F_{sens}$  in red and  $F_G$  in dashed brown lines) under clear-sky conditions (left). Annual cycle of the four SEB terms ( $F_R$  in black,  $F_{lat}$  in purple,  $F_{sens}$  in yellow and  $F_G$  in red) (right). The figure on the left is adapted and extracted from Wallace and Hobbs (2006) and the one on the right from Miller et al. (2017)..... 34

**Figure 1-12:** Scheme of the various types of clouds presented in the troposphere. This figure is extracted from <https://www.rochesterfirst.com/weather/weather-glossary/cloud-index-defining-types-of-clouds-and-their-unique-characteristics/>. ... 35

**Figure 1-13:** Explanatory diagram of the radiative effect of (a) low- and (b) high-level clouds. This figure is extracted and adapted from [https://www.nasa.gov/pdf/135641main\\_clouds\\_trifold21.pdf](https://www.nasa.gov/pdf/135641main_clouds_trifold21.pdf). ..... 37

**Figure 2-1:** Satellite image of (a) the Paris region and (b) the Toulouse region. The yellow star in (a) indicates the position of the SIRTA Supersite and the red one in (b) where the Météopole site is located. Bottom right in both figures: overview map. Figure extracted from the Google Earth tool. .... 41

**Figure 2-2:** Photos of some of the instruments installed in the SIRTA observatory. This figure is extracted from Chiriaco et al., (2018). ..... 42

**Figure 2-3:** Histograms of (a) hourly and (b) seasonal available data in the SIRTA ReOBS dataset, considering all the variables needed for the period of study 2009 – 2014. .... 43

**Figure 2-4:** Hourly values from January 2009 to February 2014 of (a)  $T_{2m}$ , (b)  $TG$ , (c)  $FLW \downarrow$ , (d)  $FLW \downarrow$ , |  $FLW \uparrow$  and (f)  $FSW \uparrow$ . ..... 46

**Figure 2-5:** Same as Figure 2-4 but for (a)  $FLW$ ,  $CS \downarrow$ , (b)  $FSW$ ,  $CS \downarrow$ , (c) sensible and (d) latent heat fluxes, and | MLD. .... 46

**Figure 2-6:** Lidar scattering ratio ( $SR(z)$ ) histogram obtained by cumulating all lidar observations during daytime from 2009 to 2014. The color bar is the logarithm of the percentage of occurrence (the sum of one level is equal to  $log_{10}100\%$ ); lidar data showed in each subplot start above the instrument's recovery altitude ( $z = 1$  km); the red horizontal lines represent the limits of low-, mid- and high-level clouds; and the white vertical line shows the threshold of clouds detection ( $SR(z) = 5$ ). ..... 47

**Figure 2-7:** Hourly difference between  $T_{2m}$  and  $T_{2m,ERA5}$  for (a) SIRTA and (b) Météopole sites. The horizontal black line on each subfigure is the mean of this temperature difference and is equal to  $-0.55$  °C for SIRTA and  $-0.17$  For the Météopole site. .... 50

**Figure 2-8:** Balance of the total work of the pressure forces exerted on the surface elements perpendicular to the axis  $Ox$  of a fluid particle. (a) The balance of the total

work of the pressure forces breaks down into two terms: (b) the work of the balance of the pressure forces leading to the displacement of the center of gravity and (c) the balance of the work of the pressure forces resulting in a variation in volume of the particle without displacement of the center of gravity. Figure adapted from Malardel (2009). ..... 54

**Figure 2-9:** Single non-averaged temperature profiles at (a) 11:00 and (b) 23:00 LT for a month of July of 2011. Each blue line in (a) and (b) corresponds to one day of the month whereas the thick dark red line is the average vertical temperature profile. Monthly averaged temperature profiles from 2003 to 2017 at (c) 11:00 LT and (d) 23:00 LT in Trappes. The horizontal dashed line represents the altitude above sea level where the Trappes observatory is located, which is 160 m. (a) and (c) have not the same vertical scale as (b) and (d). ..... 58

**Figure 2-10:** Averaged vertical temperature profile retrieved from the radiosoundings launched at 23:00 LT at Trappes (blue line) and averaged vertical temperature profile estimated with Eq. 2.12 (orange line) at 01:00 UTC. The averaged profiles are estimated from January 2009 to February 2014. .... 59

**Figure 2-11:** Hourly evolution of (a) observations, (b) the radiative clear-sky and (c) cloudy terms, (d) the ground heat exchange term, (e) atmospheric heat exchange term and (f) the Advection term. .... 62

**Figure 2-12:** Hourly sensible (yellow dots), latent (green dots) heat fluxes at the surface and the mixed layer depth (brown dots) retrieved from the SIRTA-ReOBS dataset. The red square indicates the common period for the three variables (from November 2008 to December 2009). .... 63

**Figure 2-13:** (a) Hourly  $T_{MLD}$  (blue dots) and  $T_{2m}$  (orange dots) values from 2009 to 2014. The  $T_{MLD}$  variable is retrieved using both the SIRTA-ReOBS and ERA5 datasets. (b)  $\tau\alpha$  Monthly-hourly mean values. Units in the color bar are in hours..... 64

**Figure 2-14:** Vertical mean soil temperature profile at SIRTA averaged for all the period of the study at (a) 13:00 UTC and (b) 23:00 UTC..... 65

**Figure 2-15:** Same as Figure 2-11 but for the Météopole site in Toulouse..... 67

**Figure 2-16:** Same as Figure 2-13b but for the Météopole site..... 69

**Figure 2-17:** Basic random forest structure for a regression problem. This figure is extracted and adapted from <https://medium.com/swlh/random-forest-and-its-implementation-71824ced454f>. .... 70

**Figure 2-18:** Simplified step-by-step explanatory diagram of the random forest method. This diagram is an example on how random forest method works, therefore it does not reproduce and takes exactly the same features shown e.g. in step 3. Refer to the text for further details on the syllabus. .... 73

**Figure 3-1:** PDFs of (a) radiation (red line), radiative clear-sky (red dashed line), radiative cloud (red dotted line), heat ground (brown line) and heat atmospheric (green

line) exchange, and advection (cyan line) terms, and (b)  $\partial T_{2m}/\partial t_{obs}$  (pink line) and the  $\partial T_{2m}/\partial t_{OTEM}$  (blue line). (c) Hourly evolution of all OTEM terms and of (d) observed temperature variations and OTEM for the period of study (same colors as in (a) and (b))...... 78

**Figure 3-2:** Scatter plot of  $\partial T_{2m}/\partial t_{obs}$  vs  $\partial T_{2m}/\partial t_{OTEM}$  for (a) all the hourly data, (b) only daytime cases and (c) only nighttime cases. The correlation coefficient "r" and the linear equation fitting the best for the two datasets are also shown and correspond to the diagonal dashed line, whereas the solid one represents the  $y=x$  equation. .... 79

**Figure 3-3:** Same as Figure 3-2 but for (a) winter, (b) spring, (c) summer and (d) fall. .... 80

**Figure 3-4:** Hourly evolution from 14<sup>th</sup> to 21<sup>th</sup> August 2011 (top panel) and from 23<sup>rd</sup> to 30<sup>th</sup> March 2012 (bottom panel) of the temperature variation for the different terms, the observations, and OTEM. .... 83

**Figure 3-5:** Mean diurnal cycle averaged from 2009 to 2014 of  $T_{MLD}$  (red line),  $T_{2m}$  (yellow line) for (a) winter, (b) spring, (c) summer and (d) fall seasons. The colored-semitransparent brown space between  $T_{MLD}$  and  $T_{2m}$  in each subfigure represents its difference, called  $\delta T$ . .... 84

**Figure 3-6:** Hourly evolution for the two days in which the lowest hourly observed temperature variations are found: (a) 2<sup>nd</sup> of July 2010 and (b) 6<sup>th</sup> of August 2009. Gaps in  $\partial T_{2m}/\partial t_{OTEM}$  (blue line) are due to the absence of  $FLW \uparrow$  for these hours that do not allow to estimate  $\Delta F_{NET}$ ,  $CS$  and thus  $RCS$  (cf. Eq. 2.22). .... 85

**Figure 3-7:** 07/02/2010 quicklooks 1-min values of (a)  $T_{2m}$ , (b) 24-h cumulated precipitation, (c) wind speed, and (d) wind direction. These quicklooks are available at <https://sirta.ipsl.fr/sirta/data/quicklooks/>. .... 86

**Figure 3-8:** 07/02/2010 sky imager photographs took every 30 min from 03:30 to 20:30 UTC. .... 87

**Figure 3-9:** 08/06/2009 quicklooks 1-min values of (a)  $T_{2m}$ , (b) surface downwelling longwave irradiance, (c) wind speed, and (d) wind direction. .... 88

**Figure 3-10:** Mean annual cycle averaged monthly from 2009 to 2014 of the five terms of OTEM, the observations, and the residual (dashed gray lines) for (a) day and (b) night. The shaded gray area in each of the panels represents the standard deviation of  $\partial T_{2m}/\partial t_{obs}$ . .... 90

**Figure 3-11:** Monthly-hourly mean values for (a)  $\partial T_{2m}/\partial t_{obs}$ , (b)  $\partial T_{2m}/\partial t_{OTEM}$ , (c) the residual (i.e. difference between OTEM and the observations), for (d)  $RCS$ , (e)  $RCL$ , (f) HG, (g) HA and (h) Adv. Units in the color bars are all in  $^{\circ}C/h$ , and their scale is different for each subfigure. The black contour lines in each figure correspond to sunrise (bottom line) and sunset (top line) approximative hours. .... 92

**Figure 3-12:** Advection distribution at SIRTA from 2009 to February 2014. .... 93

**Figure 3-13:** Predictor weight estimates obtained by the random forest method for (a) day and (b) night. The abscissa in both cases represents each predictor (or term) of OTEM, and the y-axis their weight (unitless) defined as the sum of their mean square error when permutation in the decision trees is done. .... 95

**Figure 3-14:** Diurnal cycle of the predictor weight estimate for each term of OTEM, for (a) winter, (b) spring, (c) summer and (d) fall. Vertical dashed black lines are for sunrise and sunset mean hours. .... 97

**Figure 3-15:** Scatter plot of  $\partial T2m/\partial t_{obs}$  as a function of the developed model  $\partial T2m/\partial t_{OTEM}$  before applying the random forest method (blue circles) and  $\partial T2m/\partial t_{obs}$  as a function of the model trained after the RF method is applied (orange circles). The correlation coefficient "r" for both scatterplots are also presented. .... 98

**Figure 3-16:** Out-of-bag error over the number of grown regression trees for day (blue line) and night (orange line) ..... 99

**Figure 4-1:** Scatter plot of the occurrence (in  $\text{Log}_{10}$ ) of daytime  $\partial T2m/\partial t_{obs}$  as a function of  $R_{CS}$  (firs row),  $R_{CL}$  (second row) and HA (third row) terms for all-sky (first column), clear-sky (second column) and cloudy (third column) conditions. .... 104

**Figure 4-2:** Same as Figure 4-1 but for nighttime. .... 105

**Figure 4-3:** (a) Daytime histogram of  $R_{CL}/R_{CS}$  and (b) nighttime histogram of  $R_{CL}$ . The red line in both figures represents the PDF and the rectangular semi-transparent brown boxes the different bins created to analyze clouds influence. These histograms are built by considering only cloudy hours. Negative and close-to-zero values are removed for daytime hours (see text for further information). .... 107

**Figure 4-4:** Histograms of total cloudy and lidar percentage sampling availability for each bin created, with respect to the total daytime cloudy cases of each season. .... 108

**Figure 4-5:** Cloudy daytime (a) percentage occurrence of precipitations and (b) mean precipitation rates for each bin split into seasons. .... 109

**Figure 4-6:** Seasonal cycle of occurrence of clouds estimated from SR values for (a) bin 1:  $-1.0 \leq R_{CL}/R_{CS} < -0.5$  (b) bin 2:  $-0.5 \leq R_{CL}/R_{CS} < 0$  , and (c) bin 3:  $0 < R_{CL}/R_{CS} \leq 0.5$ . The occurrence is defined as the number of cloudy profiles ( $SR(z) \geq 5$ ) over the number of profiles for which  $SR(z) \geq 0.01$ . November appears as a white column because there were no data at this period. .... 110

**Figure 4-7:**  $\partial T2m/\partial t_{obs} R_{CS}$  daytime values for the four seasons for three  $R_{CL}/R_{CS}$  bins:  $-1.0 \leq R_{CL}/R_{CS} < -0.5$  in blue,  $-0.5 \leq R_{CL}/R_{CS} < 0$  in green and  $0 < R_{CL}/R_{CS} \leq 0.5$  in red. Dark-colors box plots with the mean represented as '\*' correspond to cases when all meteorological variables are available at the same time (without considering lidar availability), whereas light-colors box plots with 'o' as their mean value represent the sample when both all meteorological variables and Lidar profiles are available simultaneously. Distributions are represented by box-and-whiskers plots, where the boxes indicate the 25<sup>th</sup> and the 75<sup>th</sup> data percentiles, the

whiskers indicate 5<sup>th</sup> and 95<sup>th</sup> percentiles, the middle line represents the median, and the \* or ° is the mean. .... 111

**Figure 4-8:** Daytime values for (a) temperature, (b) relative humidity, (c) upward sensible heat flux, (d) cloud cover retrieved from a sky imager, (e) shortwave and (f) longwave cloud radiative effect, (g) liquid water path and (h) surface pressure. Colors and boxplots follow the same definition as in Figure 4-7. .... 113

**Figure 4-9:** Lidar scattering ratio (SR(z)) histogram obtained by cumulating all lidar observations during daytime for bin 1 (a)–(d), bin 2 (e)–(h) and bin 3 (i)–(l) for winter (first column), spring (second column), summer (third column) and fall (fourth column). The color bar is the logarithm of the percentage of occurrence (the sum of one level is equal to  $\log_{10}100\%$ ); lidar data showed in each subplot start above the instrument’s recovery altitude ( $z = 1$  km); the red horizontal lines represent the limits of low-, mid- and high-level clouds; and the white vertical line shows the threshold of clouds detection ( $SR(z) = 5$ ). A single “Noise bin” is presented here, which is simply the sum of the -999 and -777 bins that correspond to noisy and non-normalized profiles (cf. Section 2.1.1.2.)..... 114

**Figure 4-10:** Same as Figure 4-9 but for bin 3 ( $0 < RCL/RCS \leq 0.5$ ) for (a) 02/07/2011 at 10:00 UTC and (b) 02/03/2012 at 15:00 UTC..... 116

**Figure 4-11:**  $\partial T_{2m}/\partial t_{obs}$  nighttime values for two *RCL* bins:  $0 < RCL < 0.75^\circ C/h$  in maroon and  $0.75 \leq RCL \leq 1.5^\circ C/h$  in purple. Distributions are represented by box-and-whiskers plots, where the boxes indicate the 25<sup>th</sup> and the 75<sup>th</sup> data percentiles, the whiskers indicate 5<sup>th</sup> and 95<sup>th</sup> percentiles, the middle line represents the median, and the \* is the mean..... 117

**Figure 4-12:** Same as Figure 4-5 but for nighttime cloudy cases..... 117

**Figure 4-13:** Nighttime values for (a) temperature, (b) relative humidity, (c) upward sensible heat flux at 2 m, (d) longwave cloud radiative effect, (e) liquid water path and (f) surface pressure. Colors and boxplots follow the same definition as in Figure 4-7. .... 118

**Figure 5-1:** Weather-hourly mean values for winter of (a) *RCS*, (b) *RCL*, (b) HG, (d) HA and (e) Adv and (f) the residual (i.e. difference between OTEM and the observations). Units in the color bars are all in  $^\circ C/h$ , and their scale is different for each subfigure. .... 123

**Figure 5-2:** Same as Figure 5-1 but for summer..... 124

**Figure 5-3:** Predictor weight estimates obtained by the random forest method for (a) NAO<sup>-</sup>, (b) Atl. Ridge, (c) Blocking and (d) NAO<sup>+</sup> during winter. .... 126

**Figure 5-4:** Same as Figure 5-3 but during summer season. .... 127

**Figure 5-5:** PDFs of (a) radiation (red line), radiative clear sky (red dashed line), radiative cloud (red dotted line), heat ground (brown line) and heat atmospheric (green line) exchange, and Adv (cyan line) terms, and (b)  $\partial T_{2m}/\partial t_{obs} TLS$  (pink line) and the

$\partial T_{2m}/\partial t_{O TEM TLS}$  (blue line). (c) is the scatter plot of  $\partial T_{2m}/\partial t_{obs TLS}$  vs  $\partial T_{2m}/\partial t_{O TEM TLS}$ . The sloping solid line represents the 1:1 line and the dashed one correspond to the least-square best fit linear regression line between the two datasets. The correlation coefficient "r" and the linear equation fitting the best for the two datasets are also indicated. (d) Hourly evolution for some days of August 2017 of the temperature variation for the different terms, the observations, and OTEM (same colors as in (a) and (b)). ..... 129

**Figure 5-6:** PDF of (a)  $\partial T_{2m}/\partial t_{obs}$ , (b)  $R_{CS}$ , (c)  $R_{CL}$ , (d) HG, (e) HA and (f) Adv for SIRTA (light colors) and the Météopole site (dark colors). X- and y-axis limits are different for each figure. .... 131

**Figure 5-7:** Monthly-hourly mean values for (a)  $\partial T_{2m}/\partial t_{obs TLS}$ , (b)  $\partial T_{2m}/\partial t_{O TEM TLS}$  (c) the residual (i.e. difference between OTEM and the observations), for (d)  $R_{CS}$ , (e)  $R_{CL}$ , (f) HG, (g) HA and (h) Adv. Units in the color bars are all in  $^{\circ}C/h$ , and their scale is different for each subfigure. The black contour lines in each figure correspond to sunrise (bottom line) and sunset (top line) approximative hours. .... 133

**Figure 5-8:** Diurnal cycle of the predictor weight estimate for each term of OTEM split into seasons at SIRTA (left column, exactly similar than Figure 3-13) and at the Météopole site (right column). Vertical dashed black lines are for sunrise and sunset mean hours. .... 136

**Figure 6-i:** Predictor weight estimates of the main variables acting on 2 m surface temperatures for (a) hourly, (b) daily, (c) weekly and (d) monthly temporal scales. ... 142

**Figure A-1:** Monthly mean values of MLD for SIRTA (blue line) and the Météopole site (red line). .... 177

**Figure A-2:** Map of soil types in France. The red circle indicates the location of SIRTA and the red one the location of the Météopole site. This map is extracted from Wulf et al. (2015). .... 178

**Figure A-3:** Diurnal cycle of the sensible (red line) and latent (blue line) heat fluxes. Solid lines are for SIRTA and dashed lines for Toulouse mean diurnal values. As a reminder, SIRTA period of study is from January 2009 to February 2014, and for the Météopole site is from January 2017 to December 2018. .... 179

**Figure A-4:** Pressure at the MLD for (a) SIRTA and (b) Toulouse. .... 180

**Figure A-5:** Scheme representing the calculation of the ERA5 variables in the grid point where SIRTA is located and the very next others in both horizontal axes. Note that the centered position of SIRTA is idealized since it could happen that this point is not at the center of the grid that corresponds to the one where the observatory is located. .... 181

**Figure B-6:** Diurnal cycle of the five terms in OTEM ( $R_{CS}$  in dashed red,  $R_{CL}$  in dotted red, HG in brown, HA in green and Adv in cyan), the observed temperature variations

(black solid line) and the residual (gray dashed line), for (a) winter, (b) spring, (c) summer and (d) fall seasons..... 182

**Figure C-7:** Daytime PDF of (a)  $\partial T_{2m}/\partial t_{obs}$ , (b)  $R_{CS}$ , (c)  $R_{CL}$ , (d) HG, (e) HA and (f) Adv for SIRTA (light colors) and the Météopole site (dark colors). X- and y-axis limits are different for each figure. .... 183

**Figure C-8:** Same as Figure C-7 but for nighttime. .... 184

**Figure C-9:** Monthly mean values of the sensible and latent heat fluxes and MLD in the Météopole site..... 185

## LIST OF TABLES

---

<b>Table 2-1:</b> Variables available in the SIRTA-ReOBS dataset and the Météopole site used as inputs in the current study. The percentage of availability of SIRTA-ReOBS is from January 2009 to February 2014, whereas for the Météopole site is from January 2017 to December 2018. * Variables not directly available in the dataset but estimated otherwise (see text for further information). .....	45
<b>Table 2-2:</b> Variables retrieved from ERA5 dataset and used as inputs in the current study.....	49
<b>Table 2-3:</b> Nighttime averaged values of MLD for each season of the year. ....	60
<b>Table 2-4:</b> Same as Table 2-3 but for nighttime MLD mean values at the Météopole site.....	68
<b>Table 3-1:</b> Statistics values between $\partial T_{2m}/\partial t_{obs}$ and $\partial T_{2m}/\partial t_{OTEM}$ for the four seasons of the year. The values in brackets correspond to the statistics within the 5th and 95th percentiles.....	81
<b>Table 3-2:</b> Mean values of the observations (first line), and OTEM (second line), and the relative bias (in %, third line) split into day and nighttime for every season of the year. ....	81
<b>Table 3-3:</b> Mean, maximum and minimum value contribution of each term in OTEM split in day and nighttime.....	101
<b>Table 6-i:</b> First line: Predominance of clouds detected by the lidar during daytime for each season of the year. Second to sixth line: mean contribution values of each term of OTEM for each season. Highlighted values correspond to both the maximum of positive and negative values for the season.....	143
<b>Table 6-ii:</b> Importance of each term of OTEM along the day at SIRTA. Five cases mean the term is the most important at the time considered, whereas one case means is the less important.....	144

## LIST OF ABBREVIATIONS

---

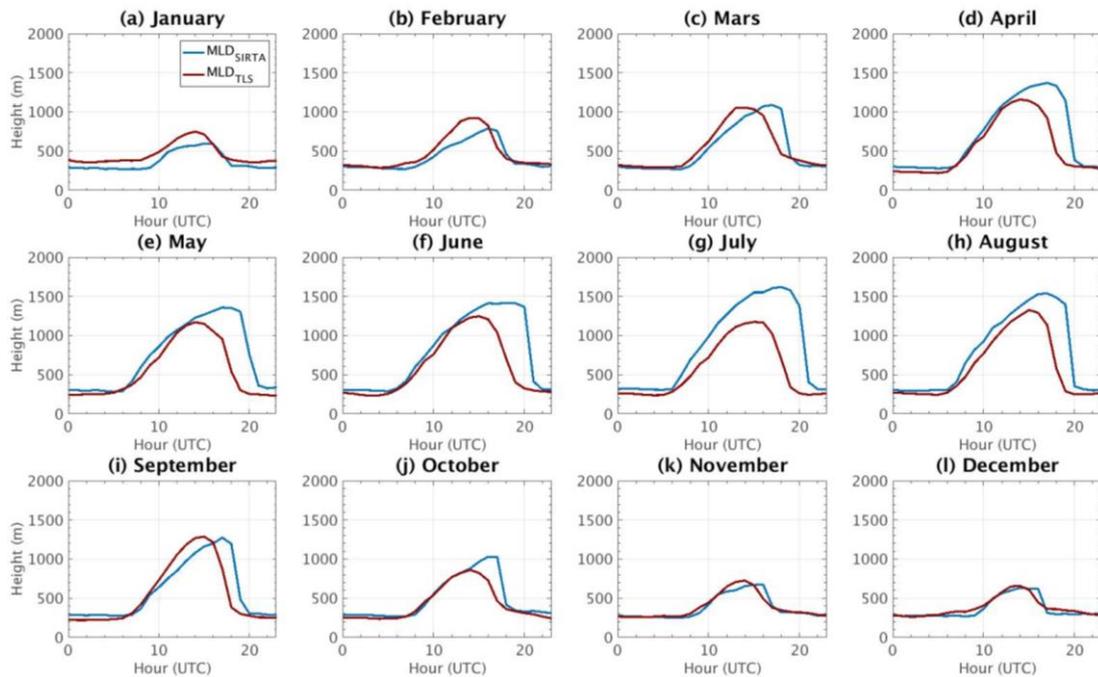
ABL	Atmospheric Boundary Layer
Adv	Advection
ATB	Attenuated Backscattered signal
CF	Cloud Fraction
CI	Capping Inversion
CL	Cloudy conditions
CRE	Cloud Radiative Effect
CS	Clear Sky
ECMWF	European Centre for Medium-Range Weather Forecast
ERA5	ECMWF Reanalysis 5
ERA5-Land	ECMWF Land Reanalysis 5
EZ	Entrainment Zone
FA	Free Atmosphere
GCM	General Circulation Models
HA	Atmospheric Heat exchange
<i>HG</i>	Ground Heat exchange term
Lidar	Light detection and ranging
LT	Local Time
LW	Long Wave
LWP	Liqui Water Path
ML	Mixed Layer
MLD	Mixing Layer Depth

MSE	Mean Squared Error
NAO	North Atlantic Oscillation
OOB	Out-Of-Bag
Obs	Observations
OTEM	Observation-based Temperature Evolution Model
PDF	Probability Density Function
R	Radiation term
RF	Random Forest
RH	Relative Humidity
RL	Residual Layer
SBL	Surface Boundary Layer
SEB	Surface Energy Balance
SIRTA	Site Instrumental de Recherche par Télédétection Atmosphérique
SR	Scattering Ratio
SW	Short Wave
SZA	Solar Zenith Angle
TLS	Toulouse
UTC	Universal Time Coordinated

## APPENDICES

### APPENDIX A.1: MLD AT SIRTA AND TOULOUSE

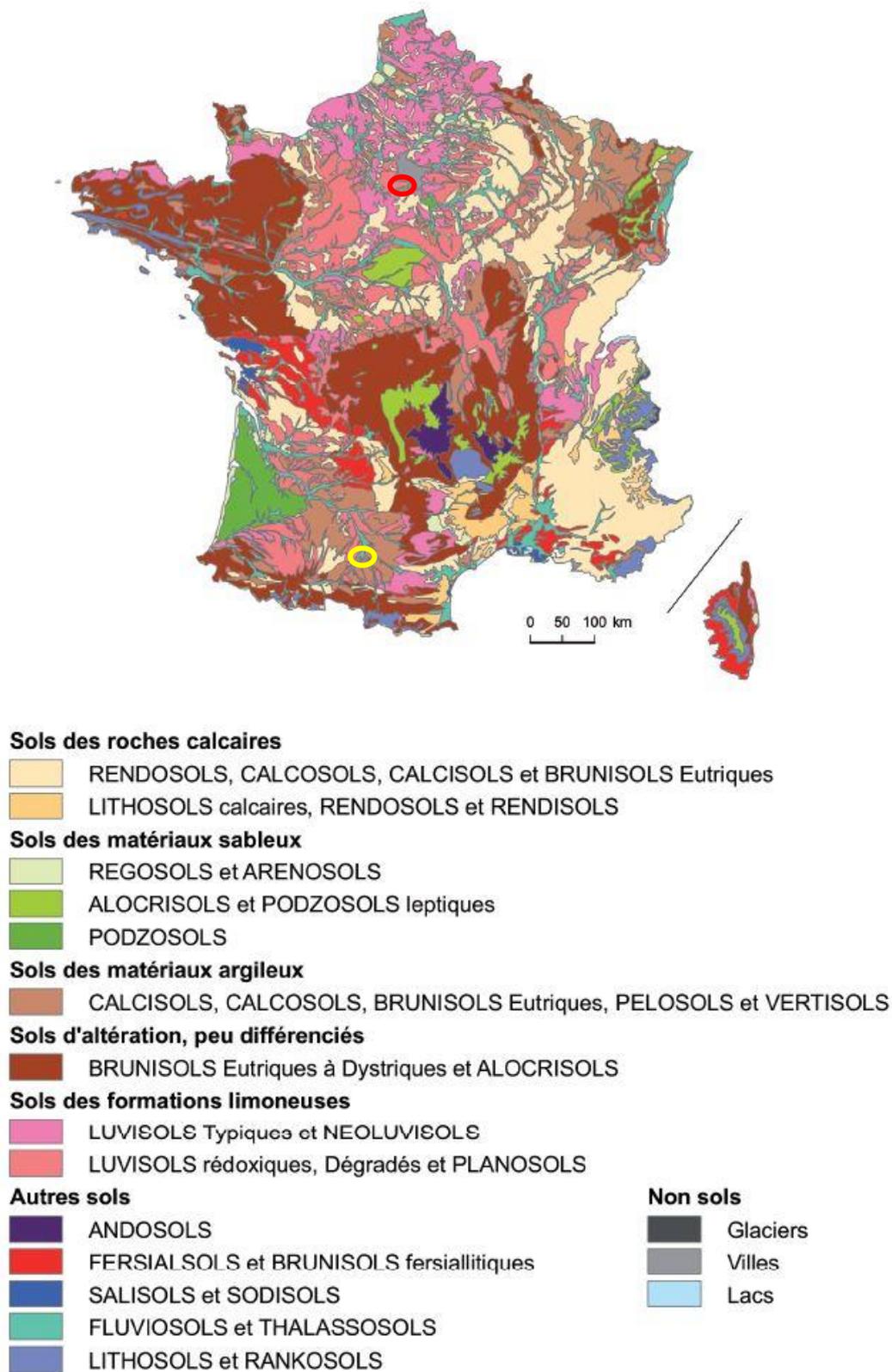
Figure A-1 shows the mean hourly values for each month for the MLD at SIRTA (blue line) and at Toulouse (red line). Both curves follow a similar behavior during all the months but their values cannot be comparable because the methods to estimate this height are different for each site.



**Figure A-1:** Monthly mean values of MLD for SIRTA (blue line) and the Météopole site (red line).

### APPENDIX A.2: TYPE OF SOILS IN FRANCE

In order to know the type of soil in SIRTA and in the Météopole to retrieve their properties and calculate the  $HG$  term in Section 2.2.1.2, a map of soil types in France done by INRA (Institut National de la Recherche Agronomique) based on their formation factors (geology, climate, relief, vegetation) is here used to know the dominant soil present on each of these observatories. This map is shown in Figure A-2, and the red circle represents the area where SIRTA is located, and the yellow one is where the Météopole is installed.

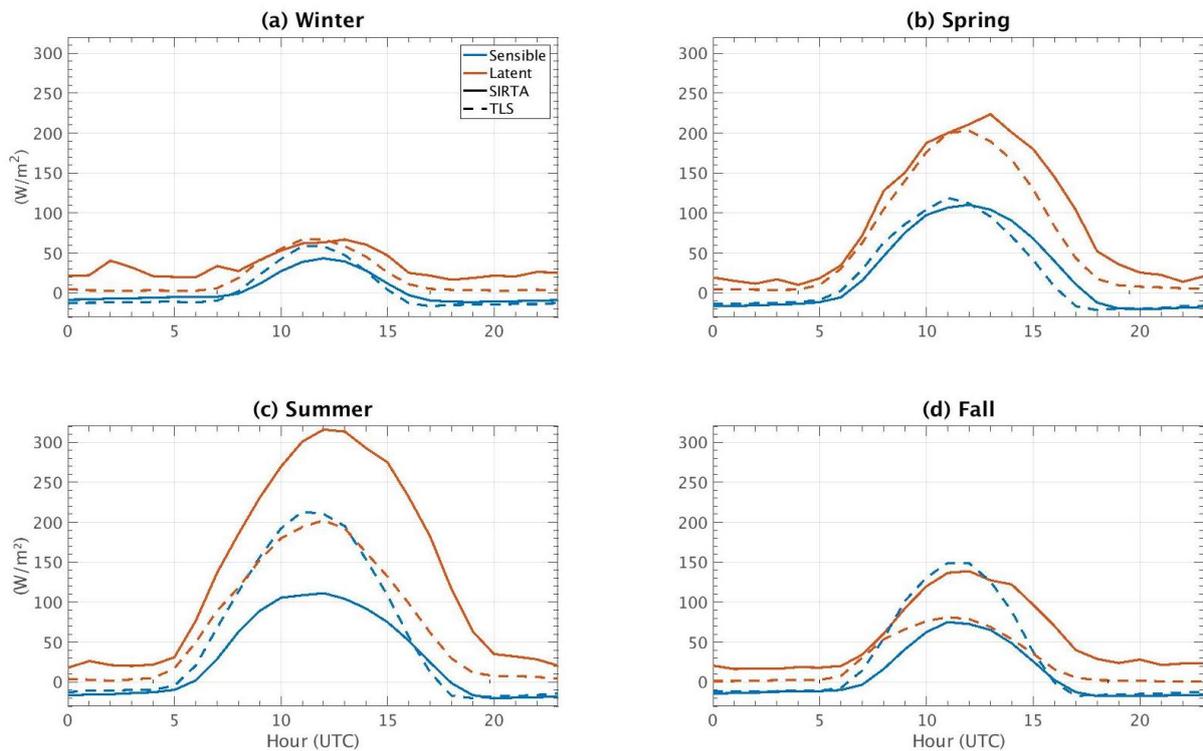


**Figure A-2:** Map of soil types in France. The red circle indicates the location of SIRTA and the red one the location of the Météopole site. This map is extracted from Wulf et al.

(2015).

### APPENDIX A.3: SENSIBLE AND LATENT HEAT FLUXES AT SIRTA AND TOULOUSE

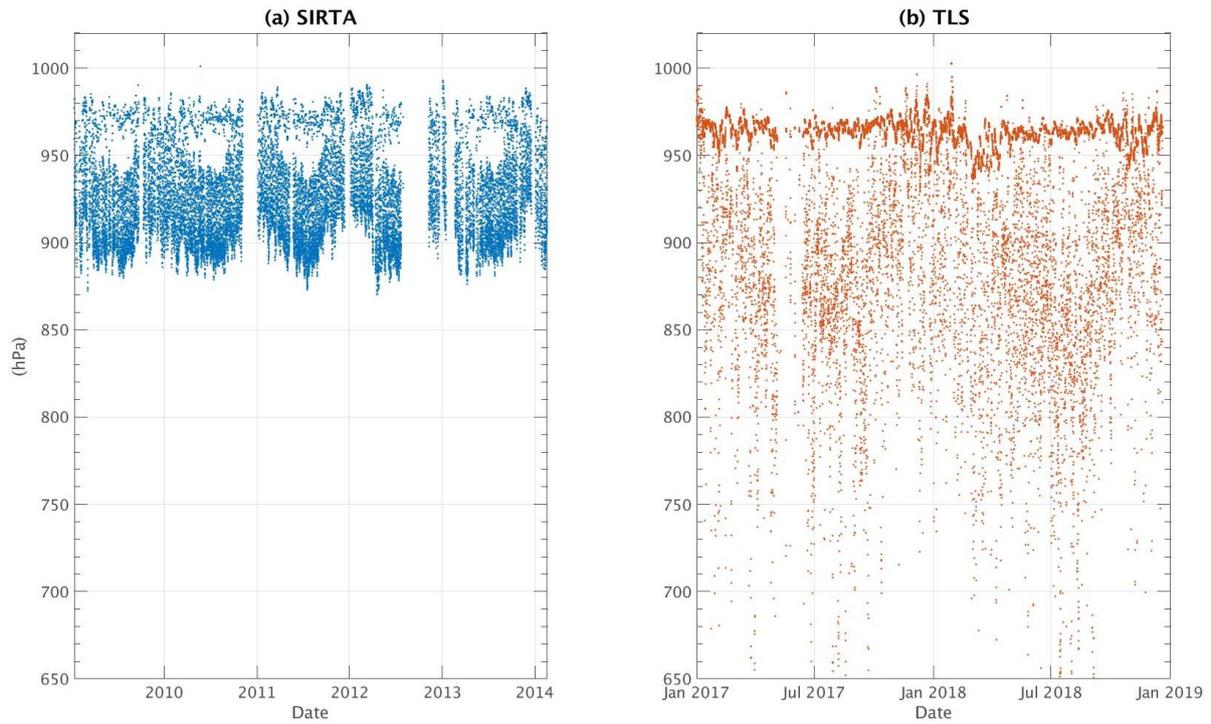
Figure A-3 shows the diurnal cycle split into seasons for the sensible and latent heat fluxes for SIRTA and the Météopole site. At SIRTA, the latent heat flux is always greater than at the Météopole in summer and fall, whereas it presents similar mean values in spring and winter (except during nighttime for the latter season). As for the sensible heat flux, it is to be expected that the highest average values are found in the Météopole because the surface is warmer than in SIRTA.



**Figure A-3:** Diurnal cycle of the sensible (red line) and latent (blue line) heat fluxes. Solid lines are for SIRTA and dashed lines for Toulouse mean diurnal values. As a reminder, SIRTA period of study is from January 2009 to February 2014, and for the Météopole site is from January 2017 to December 2018.

### APPENDIX A.4: PRESSURE AT THE MLD

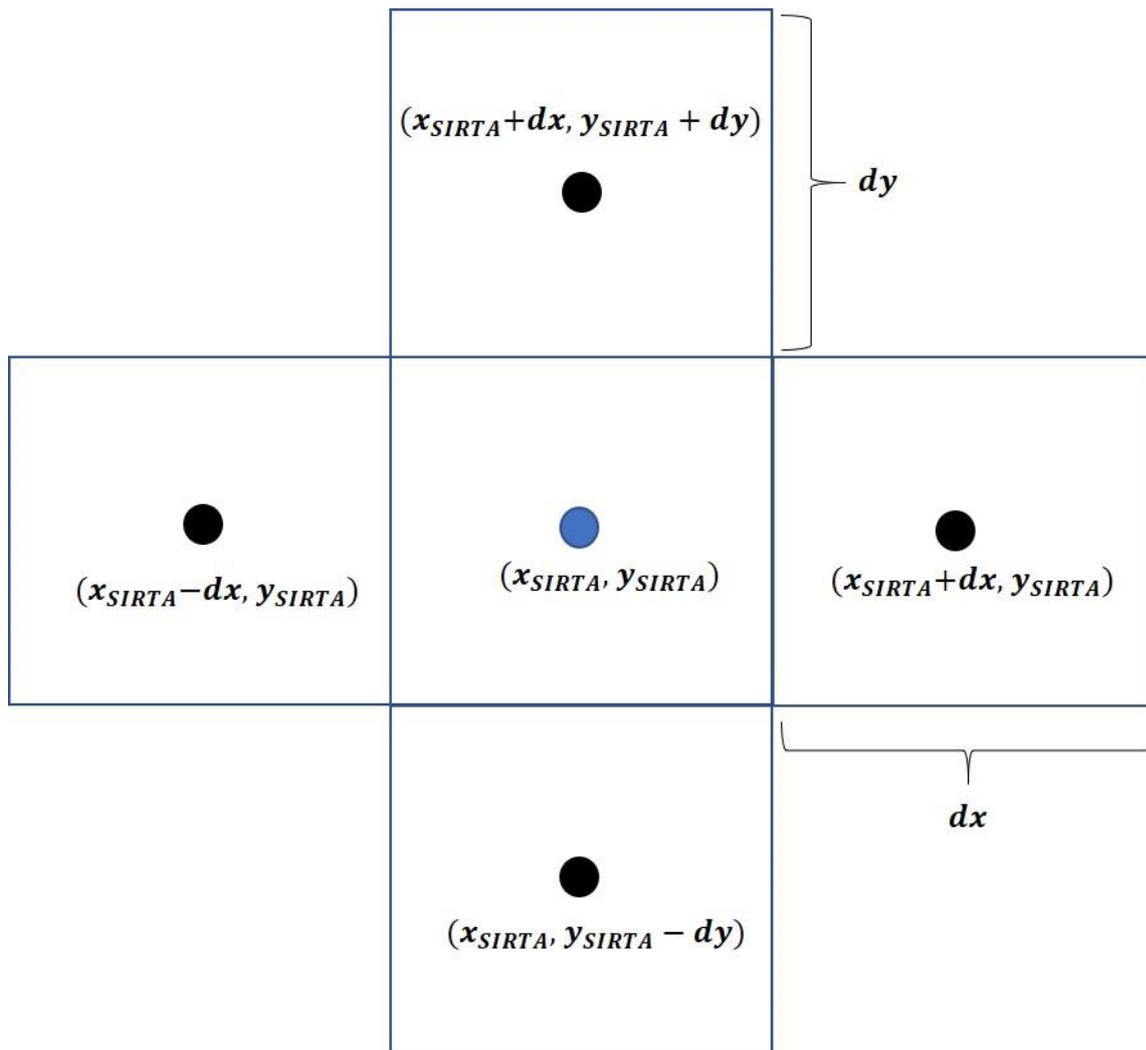
The pressure at the MLD is retrieved from the radiosoundings launched twice a day at Trappes for SIRTA and then used to the calculation of  $T_{MLD}$  with the ERA5 Reanalysis and their values are presented in Figure A-4a. For the Météopole, radiosoundings are not available and thus an approximation (cf. Eq. 2.33) is used to estimate the pressure at MLD and their values are shown in Figure A-4b. Note that for the Météopole lower values of pressure are found than for SIRTA due to the approximation of considering a constant temperature lapse rate (even during nighttime).



**Figure A-4:** Pressure at the MLD for (a) SIRT A and (b) Toulouse.

## APPENDIX A.5: ERA5-LAND

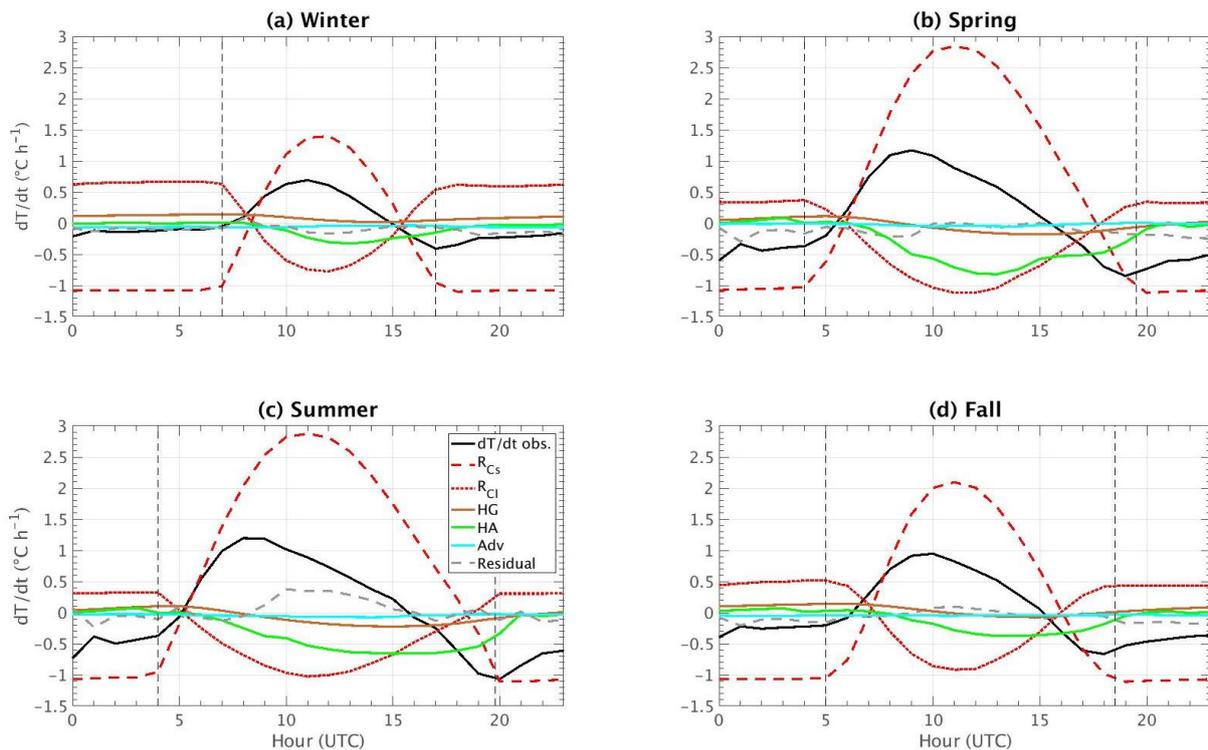
Figure A-5 presents a simple scheme on how the  $Adv$  term is calculated from the ERA5-Land Reanalysis. First, the grid where SIRT A is located is identified, in order to calculate the difference of temperature between this grid and the very next one (as it is required for the  $Adv$  term, see Eq. 2.31). We calculate the  $Adv$  term depending on the wind direction, so if for instance a wind is coming from the S-W, the temperature taken is the one at the coordinates  $(x_{SIRT A} - dx, y_{SIRT A} - dy)$ , in other words, we took for this case the grid just west of the one at SIRT A for the abscissa and the first grid just south of SIRT A for the ordinate and then we estimate  $\frac{\partial T_{2m}}{\partial x} = \frac{T_{(x,SIRT A)} - T_{(x,SIRT A-dx)}}{\partial x}$  and  $\frac{\partial T_{2m}}{\partial y} = \frac{T_{(y,SIRT A)} - T_{(y,SIRT A-dy)}}{\partial y}$ . A similar procedure is done when winds from other direction are detected.



**Figure A-5:** Scheme representing the calculation of the ERA5 variables in the grid point where SIRTA is located and the very next others in both horizontal axes. Note that the centered position of SIRTA is idealized since it could happen that this point is not at the center of the grid that corresponds to the one where the observatory is located.

## APPENDIX B: DIURNAL CYCLE OF THE OTEM TERMS

The diurnal cycle of each term split into seasons is calculated and shown in Figure B-6. This figure confirms the good accuracy of OTEM when the diurnal cycle of each of the terms is well represented. Furthermore, observed temperature variations have a peak around 09:00 to 10:00 UTC for all the seasons, except in winter when this peak is shifted two more hours. Regarding the residual, it is on average close to zero for all the seasons except in summer during daytime. As previously explained in the manuscript,  $HA$  contribution increase as the day passes by, finding a maximum just few hours before the sunset.



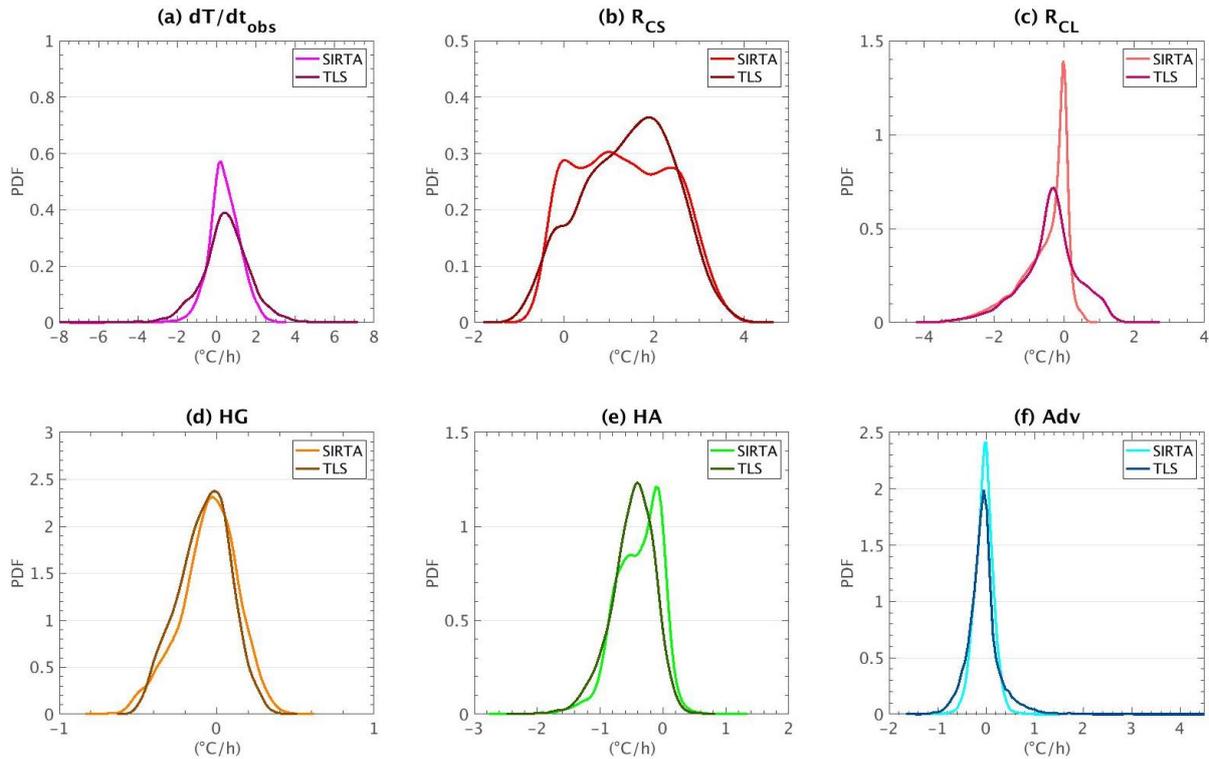
**Figure B-6:** Diurnal cycle of the five terms in OTEM ( $R_{CS}$  in dashed red,  $R_{CL}$  in dotted red, HG in brown, HA in green and Adv in cyan), the observed temperature variations (black solid line) and the residual (gray dashed line), for (a) winter, (b) spring, (c) summer and (d) fall seasons.

## APPENDIX C.1: DAYTIME AND NIGHTTIME PDFS OF OTEM TERMS AT BOTH SITES

Daytime PDFs of the observed temperature variations and the OTEM terms for SIRTA and the Météopole are plotted in Figure C-7, showing us that:

- At SIRTA, maximum observed temperature variations reach a value of  $3.9^{\circ}\text{C}/\text{h}$  whereas at Toulouse they are largely higher with more values greater than  $3.9^{\circ}\text{C}/\text{h}$  reaching up to  $7.5^{\circ}\text{C}/\text{h}$  (Figure C-7a). A similar pattern is found for the negative values when at Toulouse more negative values are found higher than  $-1.5^{\circ}\text{C}/\text{h}$  than at SIRTA. At Toulouse, temperature then both increase and decrease more than at SIRTA in an hour.
- $R_{CS}$  positive contribution presents more positive values at Toulouse than at SIRTA (Figure C-7b), but both minimum and maximum values are similar for the two sites.
- $R_{CL}$  contribute further to warm the surface during daytime at Toulouse than at SIRTA, with a maximum value of  $2.7^{\circ}\text{C}/\text{h}$  (Figure C-7c).
- HG has a more important positive contribution compared than that found at Toulouse (Figure C-7d).

- *HA* cools more the surface at Toulouse than at SIRTA during daytime (Figure C-7e).
- The *Adv* term contributes with important positive values to warm the surface at Toulouse with a maximum one around 4.5 °C/h, whereas at SIRTA its maximum warming effect is 1.57 °C/h (Figure C-7f).



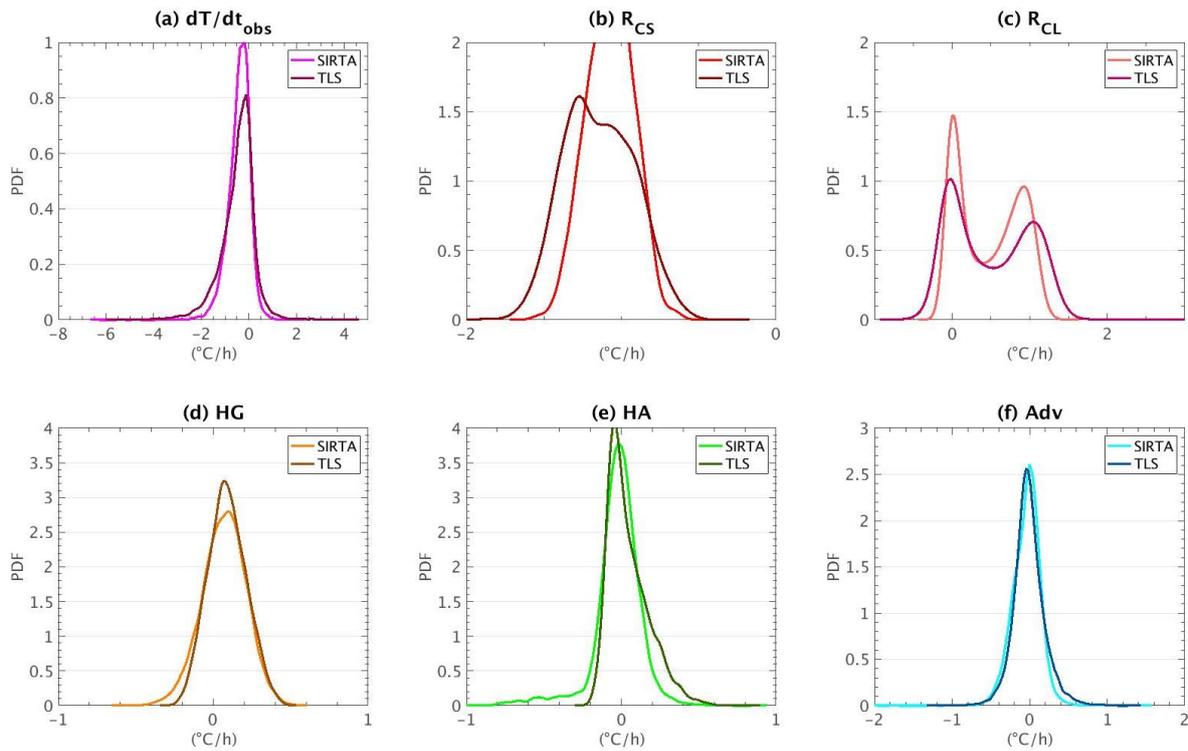
**Figure C-7:** Daytime PDF of (a)  $\frac{\partial T_{2m}}{\partial t}_{obs}$ , (b)  $R_{CS}$ , (c)  $R_{CL}$ , (d)  $HG$ , (e)  $HA$  and (f)  $Adv$  for SIRTA (light colors) and the Météopole site (dark colors). X- and y-axis limits are different for each figure.

Nighttime PDFs are also calculated for each of the terms and  $\frac{\partial T_{2m}}{\partial t}_{obs}$  and are presented in Figure C-8. The main difference between the two sites are:

- $\frac{\partial T_{2m}}{\partial t}_{obs}^{TLS}$  PDF shows higher positive values than  $\frac{\partial T_{2m}}{\partial t}_{obs}^{SIRTA}$  (Figure C-8a), indicating that at the former observatory even during the night observed temperature variations increases more rapidly than at SIRTA.
- $R_{CS}$  at Toulouse cools the surface down to -2 °C/h with an occurrence of negative values more significant than at SIRTA (Figure C-8b).
- Both negative and positive contributions of clouds are higher at Toulouse than at SIRTA during nighttime (Figure C-8c), as they are the main modulator of the clear sky radiative term which itself has high values at Toulouse.
- Regarding the ground heat exchange, the PDFs are quite similar between the two sites when this term contributes to warm the surface (Figure C-8d). Main

difference is for its cooling effect with SIRTA having lower negative values than those at Toulouse.

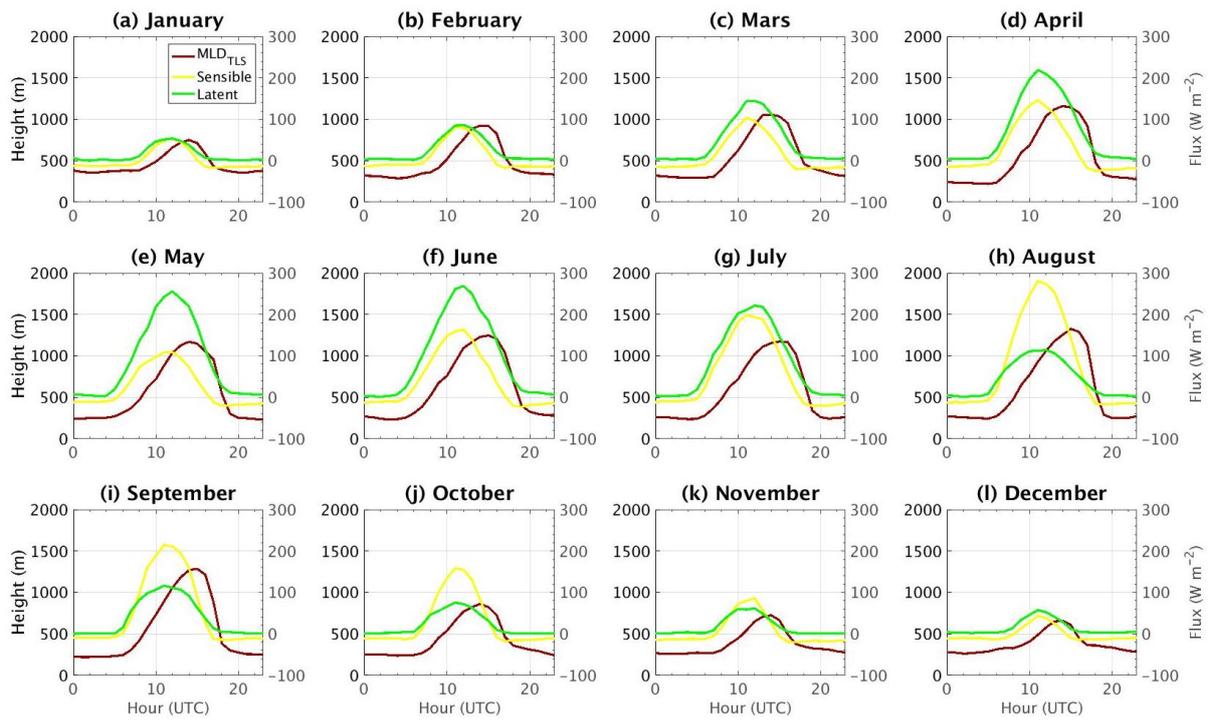
- At SIRTA, *HA* cools the surface down to  $-1.0$  °C/h which is not seen at Toulouse when a minimum negative value of  $-0.3$  °C/h is found (Figure C-8e).
- *Adv* term has a similar distribution in the two sites of study (Figure C-8f), with only difference on their minimum values which are the lowest at SIRTA.



**Figure C-8:** Same as Figure C-7 but for nighttime.

## APPENDIX C.2: SENSIBLE AND LATENT HEAT FLUXES, AND MLD AT TOULOUSE

Figure C-9 shows the mean hourly values for each month for the sensible and latent heat fluxes along with the MLD at Toulouse. The increase of both turbulent heat fluxes is linked to the increase of MLD for all the months, and these fluxes start to decrease just a couple of hours before MLD, which agrees with what is found by Pal et al. (2013) at SIRTA for these surface heat fluxes and MLD mean diurnal cycles.



**Figure C-9:** Monthly mean values of the sensible and latent heat fluxes and MLD in the Météopole site.





# Estimation of the terms acting on local 1 h surface temperature variations in Paris region: the specific contribution of clouds

Oscar Javier Rojas Muñoz<sup>1</sup>, Marjolaine Chiriaco<sup>1</sup>, Sophie Bastin<sup>1</sup>, and Justine Ringard<sup>1,†</sup>

<sup>1</sup>LATMOS/IPSL, UVSQ, Université Paris-Saclay, Sorbonne Université, CNRS, 78280, Guyancourt, France

<sup>†</sup>deceased, 21 May 2021

**Correspondence:** Oscar Javier Rojas Muñoz (oscar.rojas@latmos.ipsl.fr)

Received: 1 May 2021 – Discussion started: 18 May 2021

Revised: 9 September 2021 – Accepted: 28 September 2021 – Published: 21 October 2021

**Abstract.** Local short-term temperature variations at the surface are mainly dominated by small-scale processes coupled through the surface energy balance terms, which are well known but whose specific contribution and importance on the hourly scale still need to be further analyzed. A method to determine each of these terms based almost exclusively on observations is presented in this paper, with the main objective being to estimate their importance in hourly near-surface temperature variations at the SIRTa observatory, near Paris. Almost all terms are estimated from the multi-year dataset SIRTa-ReOBS, following a few parametrizations. The four main terms acting on temperature variations are radiative forcing (separated into clear-sky and cloudy-sky radiation), atmospheric heat exchange, ground heat exchange, and advection. Compared to direct measurements of hourly temperature variations, it is shown that the sum of the four terms gives a good estimate of the hourly temperature variations, allowing a better assessment of the contribution of each term to the variation, with an accurate diurnal and annual cycle representation, especially for the radiative terms. A random forest analysis shows that whatever the season, clouds are the main modulator of the clear-sky radiation for 1 h temperature variations during the day and mainly drive these 1 h temperature variations during the night. Then, the specific role of clouds is analyzed exclusively in cloudy conditions considering the behavior of some classical meteorological variables along with lidar profiles. Cloud radiative effect in shortwave and longwave and lidar profiles show a consistent seasonality during the daytime, with a dominance of mid- and high-level clouds detected at the SIRTa observatory, which also affects near-surface temperatures and upward sensible heat flux. During the nighttime, despite cloudy conditions and

having a strong cloud longwave radiative effect, temperatures are the lowest and are therefore mostly controlled by larger-scale processes at this time.

## 1 Introduction

Regional climate variability is to the first order driven by large-scale atmospheric conditions. In western Europe, the North Atlantic Oscillation (NAO; Trigo et al., 2002), which is associated with the locations and intensities of the centers of the Iceland low and the Azores high, controls the air mass advection over western Europe and explains a large part of weather variability. Temperature and pressure conditions are then modulated by the complex terrain (Mediterranean sea, topography, surface heterogeneities): extreme events and temperature anomalies are generally not exclusively explained by the presence of these large-scale air mass circulations (Vautard and Yiou, 2009). Indeed, synoptic and mesoscale atmospheric processes have been previously studied to explain interannual temperature changes in some parts of Europe (Efthymiadis et al., 2011; Xoplaki et al., 2003), or even precipitation occurrence (Xoplaki et al., 2004; Bartolini et al., 2009). It is, therefore, necessary to consider small-scale processes such as surface–atmosphere interactions and cloud feedbacks to better explain near-surface temperature variations (e.g., Chiriaco et al., 2014).

Temperature variations at the surface are related to the surface energy balance (SEB) following surface–atmosphere interactions and solar radiation (Wang and Dickinson, 2013) that are separated into different components: latent and sensible heat fluxes ( $F_{\text{lat}}$  and  $F_{\text{sens}}$ , respectively), ground heat

flux, temperature advection, and atmosphere radiation. Studies have been made to parametrize these terms when direct measurements are not available (Miller et al., 2017; Arnold et al., 1996), but their exact contribution is uncertain. Depending on the timescale considered, the importance of each SEB term for temperature variations will change (Bastin et al., 2018; Ionita et al., 2015).

The first objective of the current paper is to quantify the specific local contribution of the primary SEB terms acting on short-term (i.e., hourly) temperature variations in a western European location and to determine their importance and the conditions when a certain term predominates over the others, based on a simple linear model. The current study is inspired by Bennartz et al. (2013), who implemented a temperature variation model to study the influence of low-level liquid clouds over the Arctic on the ice surface melt period in July 2012, by estimating all of these SEB terms for a case study. Here, the same approach is used for a different location (mid-latitude) and on a long time period to get robust statistics. To do so, the study is based on direct measurements from the SIRTA-ReOBS dataset (Chiriaco et al., 2018), which includes many variables collected since 2002 at SIRTA (Site Instrumental de Recherche en Télédétection Active; Haeffelin et al. 2005), an observatory located in a semi-urban area in the southwest of Paris, France. This dataset is well suited to the current study objective because (i) it allows the use of a multi-variable synergistic compilation to study and compare different characteristics of the atmosphere or the surface (e.g., Bastin et al., 2018; Dione et al., 2017; Cheruy et al., 2013; Chiriaco et al., 2014), and (ii) it is located in western Europe and allows access to the hourly timescale, i.e., the scales of the local processes of the current study. The SIRTA-ReOBS dataset, along with some variables retrieved from ERA5, enables the development of a model to estimate all the terms involved in the near-surface temperature variations using predominantly real observations. The use of the model developed in the current study considers all the variables acting within the atmospheric boundary layer (ABL) and controlling surface temperature variations, all of them estimated almost exclusively from surface-based observations. Thus, it allows us to separately study the influence of each SEB term in a local scale. This indeed allows a realistic and reliable estimation of the contribution of each term (radiative fluxes, turbulent heat fluxes, etc.) on hourly temperature variations, and it would be possible to have that at different sites since each term will present a different behavior and importance. These estimations could help improve the parametrizations of the SEB terms that already exist and better understand their spatial evolution as a function of local conditions. Furthermore, a comparison between multi-model regional climate simulations and these estimations can be performed to evaluate whether the simulations are able to well reproduce these behaviors, in particular in a warming climate where these processes are expected to change. The weight of each term of the SEB on hourly temperature

variations is analyzed using a powerful random forest analysis, one of whose attributes is its capacity to handle up to thousands of input variables and identify the most significant ones.

Clouds are well known to directly modify near-surface temperatures and other near-surface variables on multiple timescales (Parding et al., 2014; Broeke et al., 2006; Kauppinen et al., 2014). Hence the second objective of the current study is to understand the specific role of clouds and their associated characteristics in hourly temperature variations. Indeed, the influence of clouds on the temperature at the surface can vary depending on their physical properties and the altitude where they are formed (Hartmann et al., 1992; Chen et al., 2000). In general, low-level clouds tend to cool the surface, whereas high-level clouds, such as cirrus clouds, tend to warm it by absorbing a significant amount of Earth's outgoing radiation. The average contribution of clouds is to decrease the near-surface temperature, combined with the damping effects of soil moisture, from a global point of view, by reflecting the solar radiation to space. But their damping effects vary depending on the season and by time of day. For instance, the reduction in near-surface temperature is the highest in fall in the northern mid-latitudes for specific cases when precipitation is not significant (Dai et al., 1999). The local contribution of clouds to temperature variations is thus an important topic to assess how this intake affects local climate variability and extreme local events, such as the extreme heatwave and drought during summer in the EU in 2006 (Chiriaco et al., 2014; Rebetez et al., 2009) or the sudden melt of the ice sheet in Greenland on July 2012 (Bennartz et al., 2013). Several studies were made primarily focusing on the large-scale effects of clouds on the radiative energy balance on a global scale, either at the top of the atmosphere (Arkin and Meisner, 1987; Raschke et al., 2005; Dewitte and Clerbaux, 2017; Willson et al., 1981; Allan et al., 2014; Cherviakov, 2016) or at the surface (Wild et al., 2015; Hakuba et al., 2013; Wild, 2017) or both at the same time (Hartmann, 1993; Kato et al., 2012; Li and Leighton, 1993), but a lack of studies investigating their impact on a smaller scale is noted due to limited reliable ground-based measurement availability. In this study, to understand how clouds influence the 1 h temperature variations, cases with a particular cloud effect on temperature variations are more deeply analyzed using lidar profiles.

To achieve these two objectives, the current paper is organized as follows: the dataset is presented in Sect. 2. Section 3 (and Appendix A) describes how the different terms acting on hourly temperature variations are estimated and evaluates how well the model fits the hourly observations based on statistics. Section 4 presents an assessment to determine which term dominates at different times by performing a random forest analysis firstly for day and night cases and then separated in a diurnal cycle perspective, along with a mean monthly–hourly and annual cycle analysis of the contribution of each term. Section 5 focuses on a discussion to study

the specific role of clouds and the atmospheric conditions under which they develop by assessing only the cloudy cases, which gives an overview of the type of clouds and surface conditions damping or enhancing near-surface temperature for both daytime and nighttime. Section 6 draws conclusions and provides perspectives opened by this work.

## 2 Data

### 2.1 Data used for the temperature variation estimation model

This study is mainly based on the SIRTAREOBS dataset analysis. The SIRTAREOBS observatory (Haefelin et al., 2005) is located in a semi-urban area 20 km southwest of Paris (48.71° N, 2.2° E) and has collected long-term meteorological variables since 2002. The ReOBS project aims to synthesize all observations available at a single observatory at an hourly timescale with exhaustive data-quality control, calibration, and rigorous treatment into a single NetCDF file. The SIRTAREOBS dataset (Chiriaco et al., 2018, <https://doi.org/10.5194/essd-10-919-2018>) contains more than 60 variables.

All the necessary hourly variables requested by the current study are available for the period from January 2009 to February 2014 and are listed in Table 1, allowing a multi-year analysis. Some variables that were retrieved from measurements at the SIRTAREOBS observatory sometimes present gaps due to instrumental issues (Chiriaco et al., 2018; Pal and Haefelin, 2015). In particular  $F_{\text{lat}}$  and  $F_{\text{sens}}$  variables are limited (with only 5 % and 73 % availability, respectively): Sect. 3 and Appendix A show how this issue is handled.

Figure 1 shows the complete dataset availability in the SIRTAREOBS dataset for the 5-year study period. The availability of data is quasi-homogeneous around 60 % for all hours (Fig. 1a). Most of the time gaps are due to the absence of the mixing layer depth (MLD) variable for some complete days, extended sometimes for more than 2 months (not shown), restricting 71 % of MLD data availability. Other gaps are caused by the absence of radiative variables (9 % of missing data; see Table 1). Summer is the season with the best data coverage (70 %), and winter has the least data coverage (56 % – Fig. 1b), whose absences are due precisely to gaps in the MLD variable.

To complete this dataset, hourly ERA5-Land (horizontal resolution of  $0.1^\circ \times 0.1^\circ$ ; Copernicus Climate Change Service, 2019) is used to estimate the horizontal wind at 10 m and the temperature near the surface  $T_{2\text{m}}$ , as required for the advection term (see Sect. 3 and Appendix A). Furthermore, the temperature at the mixing layer depth ( $T_{\text{MLD}}$ ) is also retrieved from the ERA5 (spatial resolution of 31 km and 137 levels up to 1 hPa; ERA5, 2020).

### 2.2 Variables used for the cloud contribution analysis

In order to get vertical information about clouds (see Sect. 5), SIRTAREOBS lidar profiles are also used, retrieved from a LNA lidar (532 and 1064 nm) whose vertical resolution is 15 m (for further details, see Chiriaco et al., 2018), and their analysis is based on hourly lidar scattering ratio (SR) altitude–intensity histograms, calculated as follows:

$$\text{SR}(z) = \frac{\text{ATB}(z)}{\text{ATB}(z)_{\text{mol}}}, \quad (1)$$

where  $\text{ATB}(z)$  is defined as the total attenuated backscatter lidar signal, and  $\text{ATB}(z)_{\text{mol}}$  is the signal in clear-sky conditions. These altitude–intensity histograms are used to estimate the mean cloud fraction percentage at a given altitude level  $z$ . The intensity axis which contains  $\text{SR}(z)$  thresholds as well as the three vertical atmospheric layers (i.e., low, middle, and high layers) for cloud detection and characterization are defined in Chepfer et al. (2010). The value of  $\text{SR}(z) = -999$  corresponds to non-normalized profiles, the value  $-777$  represents the profiles that cannot be normalized due to the presence of a very low opaque cloud, and the value of  $-666$  is set as invalid data. Then, bins located in the range  $0.01 < \text{SR}(z) < 1$  are for clear-sky conditions and  $1.2 < \text{SR}(z) < 5$  is defined as unclassified data. For cloud detection, a threshold of  $\text{SR}(z) > 5$  is set. Details are given in Chiriaco et al. (2018). Even though the instrument does not operate uninterrupted (it does not operate when it is raining, when it is nighttime, and on the weekends), it is very powerful to get information on the vertical structure of the atmosphere.

## 3 Estimation of the terms acting on near-surface temperature variations

The exchange of energy between the surface and the overlying atmosphere involves four terms: radiation, heat exchange with ground, heat exchange with the free atmosphere, and advection. In this section, a model which estimates these four terms is presented, based on several meteorological observations retrieved from the SIRTAREOBS dataset. Each term involved in this model is described (further descriptions are presented in Appendix), and then a statistical evaluation is performed to assess how well the model follows the real observations of temperature variations. Finally, the mean monthly–hourly and annual cycles of the averaged contribution of each term and the same for 1 h temperature variations ( $\frac{\partial T_{2\text{m}}}{\partial t}_{\text{obs}}$ ) are presented (split into day and night).

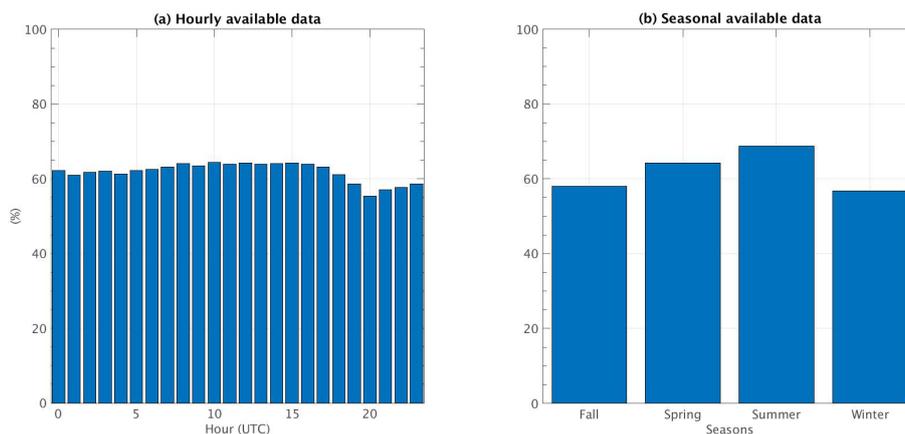
### 3.1 Model description

The temperature variation at the surface is estimated from the sum of four terms:

$$\frac{\partial T_{2\text{m}}}{\partial t} = R + \text{HG} + \text{HA} + \text{Adv}, \quad (2)$$

**Table 1.** Variables available in the SIRTReOBS dataset used as inputs in the temperature variation model. \*  $T_{\text{MLD}}$  is also estimated using the ERA5 dataset.

Variable, unit	Notation	SIRTReOBS available data from 2009 to 2014, in %
2 m air temperature, K	$T_{2\text{m}}$	97
Soil temperature below the ground, K	$T_{\text{s}}$	99
Temperature at the mixing layer height*, K	$T_{\text{MLD}}$	71
Surface downwelling LW radiation, $\text{W m}^{-2}$	$F_{\text{LW}}^{\downarrow}$	100
Surface downwelling SW radiation, $\text{W m}^{-2}$	$F_{\text{SW}}^{\downarrow}$	100
Surface upwelling LW radiation, $\text{W m}^{-2}$	$F_{\text{LW}}^{\uparrow}$	91
Surface upwelling SW radiation, $\text{W m}^{-2}$	$F_{\text{SW}}^{\uparrow}$	94
Surface downwelling LW radiation for clear sky, $\text{W m}^{-2}$	$F_{\text{LW,CS}}^{\downarrow}$	97
Surface downwelling SW radiation for clear sky, $\text{W m}^{-2}$	$F_{\text{SW,CS}}^{\downarrow}$	97
Surface upward sensible heat flux, $\text{W m}^{-2}$	$F_{\text{sens}}$	73
Surface upward latent heat flux, $\text{W m}^{-2}$	$F_{\text{lat}}$	5
Mixing layer depth, m	MLD	71

**Figure 1.** Histograms of (a) hourly and (b) seasonal available data in the SIRTReOBS dataset, considering all the variables needed for the period of study 2009–2014 for the temperature variation model.

where  $R$  is the net radiative flux at the surface,  $HG$  is the ground heat exchange,  $HA$  is the atmospheric heat exchange, and  $Adv$  is the air advection. These four terms control the changes in temperature over the surface, but depending on the temporal scale, one will dominate over the others (Ionita et al., 2012). Following Appendix A and partly based on Benartz et al. (2013), these four terms are expressed as

$$R = \frac{\alpha + 1}{\rho c_p \text{MLD}} \Delta F_{\text{NET}}, \quad (3)$$

$$HG = \frac{T_{\text{s}} - T_{2\text{m}}}{\tau_{\text{s}}}, \quad (4)$$

$$HA = \frac{T_{\text{MLD}} - T_{2\text{m}}}{\tau_{\text{a}}}, \quad (5)$$

$$Adv = \left( u_{10} \frac{\partial T_{2\text{m}}}{\partial x} + v_{10} \frac{\partial T_{2\text{m}}}{\partial y} \right). \quad (6)$$

Then

$$\frac{\partial T_{2\text{m}}}{\partial t} = \frac{\alpha + 1}{\rho c_p \text{MLD}} \Delta F_{\text{NET}} + \frac{T_{\text{s}} - T_{2\text{m}}}{\tau_{\text{s}}} + \frac{T_{\text{MLD}} - T_{2\text{m}}}{\tau_{\text{a}}} + \left( u_{10} \frac{\partial T_{2\text{m}}}{\partial x} + v_{10} \frac{\partial T_{2\text{m}}}{\partial y} \right), \quad (7)$$

where  $T_{2\text{m}}$  is the near-surface temperature,  $t$  is the time (in hours),  $\alpha$  is a coefficient characterizing the form of the temperature vertical profile in the boundary layer,  $\rho$  is the average air density of the boundary layer,  $c_p$  is the specific heat of air,  $\text{MLD}$  is the mixing layer depth,  $\Delta F_{\text{NET}}$  is the net radiative flux at the surface,  $T_{\text{s}}$  is the temperature in the ground at 20 cm depth,  $T_{\text{MLD}}$  is the temperature at the top of the boundary layer (or mixed layer depth),  $u_{10}$  and  $v_{10}$  are the zonal and meridional wind components, respectively, at 10 m above the ground,  $x$  is the zonal wind component towards the east and  $y$  is the meridional component wind towards the north, and  $\tau_{\text{s}}$

and  $\tau_a$  are defined as relaxation timescales for heat exchange processes in the ground and the atmosphere, respectively.

If the model is realistic, then the sum of the four terms on the left side of Eq. (2), denoted as  $\frac{\partial T_{2m}}{\partial t}_{\text{mod}}$ , is as close as possible to the observed temperature variations denoted as  $\frac{\partial T_{2m}}{\partial t}_{\text{obs}}$ .

The net radiative flux at the surface is calculated as the difference between the radiative longwave (LW) and shortwave (SW) fluxes leaving and arriving at the surface as follows:

$$\Delta F_{\text{NET}} = F_{\text{SW}}^{\downarrow} - F_{\text{SW}}^{\uparrow} + F_{\text{LW}}^{\downarrow} - F_{\text{LW}}^{\uparrow}, \quad (8)$$

where the upward and downward arrows represent the upwelling and downwelling radiation, respectively. It is possible to add and subtract the clear-sky (CS) downwelling radiation fluxes to Eq. (8), to obtain the flux radiative components in CS and cloudy (CL) conditions (i.e., the specific contribution of clouds), as follows.

$$\begin{aligned} \Delta F_{\text{NET}} &= F_{\text{SW}}^{\downarrow} - F_{\text{SW}}^{\uparrow} + F_{\text{LW}}^{\downarrow} - F_{\text{LW}}^{\uparrow} + F_{\text{SW,CS}}^{\downarrow} \\ &\quad - F_{\text{SW,CS}}^{\downarrow} + F_{\text{LW,CS}}^{\downarrow} - F_{\text{LW,CS}}^{\downarrow} \\ \Delta F_{\text{NET}} &= \underbrace{F_{\text{SW}}^{\downarrow} - F_{\text{SW,CS}}^{\downarrow} + F_{\text{LW}}^{\downarrow} - F_{\text{LW,CS}}^{\downarrow}}_{\Delta F_{\text{NET,CL}}} \\ &\quad + \underbrace{F_{\text{SW,CS}}^{\downarrow} - F_{\text{SW}}^{\uparrow} + F_{\text{LW,CS}}^{\downarrow} - F_{\text{LW}}^{\uparrow}}_{\Delta F_{\text{NET,CS}}} \end{aligned} \quad (9)$$

Here, the  $F_{\text{LW}}^{\uparrow}$  flux is the same for clear-sky and cloudy conditions, because it mostly depends on the near-surface temperature (i.e., the Stefan–Boltzmann law,  $F^{\uparrow} = \sigma T_{2m}^4$ , where  $\sigma$  is the Stefan–Boltzmann constant), but note that in an annual global mean,  $F_{\text{LW}}^{\uparrow} > F_{\text{LW,CS}}^{\uparrow}$  (around  $0.5 \text{ W m}^{-2}$ ; Allan, 2011) due to the increase in longwave radiation emitted toward the surface by clouds, where a small proportion of this radiation is reflected by the surface.

Hence Eq. (2) becomes

$$\frac{\partial T_{2m}}{\partial t} = R_{\text{CL}} + R_{\text{CS}} + \text{HG} + \text{HA} + \text{Adv}, \quad (10)$$

with  $R_{\text{CL}} = \frac{\alpha+1}{\rho c_p \text{MLD}} \Delta F_{\text{NET,CL}}$  and  $R_{\text{CS}} = \frac{\alpha+1}{\rho c_p \text{MLD}} \Delta F_{\text{NET,CS}}$ .

This formulation specifically allows estimation of the role of clouds ( $R_{\text{CL}}$ ) in temperature variations, compared to the other terms. Details on how each term is estimated along with the assumptions that have been made are stated in the Appendix.

As shown by Eqs. (3) to (6) and in Eqs. (A10) to (A17) in Appendix A, temperature variations are estimated from variables described in Sect. 2, i.e., predominantly based on observations retrieved from SIRTAREOBS, but also on ERA5 datasets.

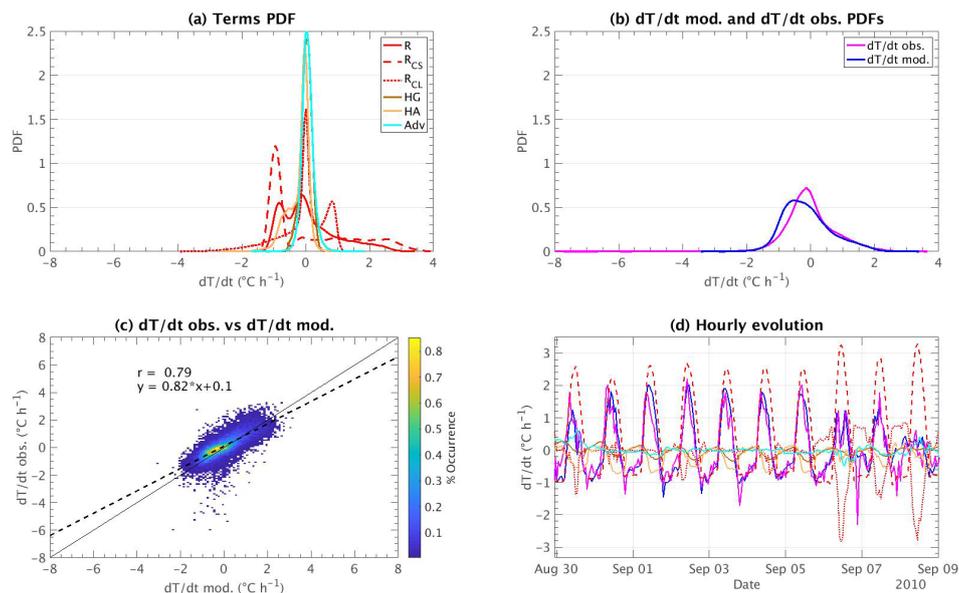
Concerning the advection term, the computation of this term requires the extraction of the temperature at 2 m ( $T_{2m}$ )

and northward ( $v_{10m}$ ) and eastward ( $u_{10m}$ ) wind components at 10 m, at the SIRTAREOBS grid point and the surrounding grid points. The temperature at the mixing layer depth  $T_{\text{MLD}}$  is retrieved using SIRTAREOBS combined with ERA5 datasets. First, radiosoundings available in SIRTAREOBS (twice a day) are used to get the pressure at the MLD level, and then  $T_{\text{MLD}}$  is retrieved from the vertical temperature profile in ERA5 at the nearest grid box from SIRTAREOBS (48.7° N and 2.2° E) and at this pressure level but at the closest time. Note that uncertainty remains in  $T_{\text{MLD}}$  because it is the temperature at a certain pressure level, which is only available twice a day.

### 3.2 Statistical evaluation of the model

Statistics are considered to assess how close the total model ( $\frac{\partial T_{2m}}{\partial t}_{\text{mod}}$ ) is to the observed temperature variations ( $\frac{\partial T_{2m}}{\partial t}_{\text{obs}}$ ). The PDFs of each term in Eq. (2) are shown in Fig. 2a. Firstly, the radiative term  $R$  dominates, and contributes the most in the temperature variations at an hourly scale since its distribution presents different significant peaks for negative and positive values of temperature variations. In clear-sky conditions, the radiative term contributes the most to warm the surface, whereas the radiative effect of clouds has an opposite effect. A significant negative contribution by the HA term is also observed (light brown line), meaning that the mixing with an atmosphere of higher levels contributes to decrease surface temperature variations, even if surface temperatures continue increasing along the day. The two other terms (HG and Adv) have a similar but weak impact on the temperature variation model at this timescale. In addition, the PDF of the  $\frac{\partial T_{2m}}{\partial t}_{\text{mod}}$  is compared to  $\frac{\partial T_{2m}}{\partial t}_{\text{obs}}$  in 1 h in Fig. 2b: differences occur for cases where the temperature decreases during the hour, but these differences correspond to some cases where the model presents more negative values than the observations, around  $-1 \text{ }^\circ\text{C h}^{-1}$ . The modeled PDF fits very well with the observed PDF for cases where temperature increases during the hour (a smaller number of cases as shown by the PDFs).

Figure 2c shows the scatter plot of the observations versus the model. The correlation between the two datasets is good (0.79) and the bias remains small ( $-0.20 \text{ }^\circ\text{C h}^{-1}$ ). However, the model has difficulties in reproducing the extremes of temperature change within an hour, e.g.,  $-6 \text{ }^\circ\text{C h}^{-1}$ , which corresponds to the 5 July 2011 at 20:00 UTC, or  $-8 \text{ }^\circ\text{C h}^{-1}$ , which corresponds to the 2 July 2010 at 15:00 UTC. This high decrease in temperature could be associated with a cold pool event, which is more detectable in summer thanks to higher near-surface temperatures than those in winter and bring along with it storms and heavy precipitations, as detected for these two cases (not shown). The hourly values in the observations (as well as in ERA5) do not allow the capture of this cold pool event properly in the model developed, since it is related to rapid cloud formation ( $< 1 \text{ h}$ ) that is not well captured by the  $R_{\text{CL}}$  term. However since this type



**Figure 2.** PDFs of (a) radiation (red line), radiative clear sky (red dashed line), radiative cloud (red dotted line), ground heat (dark brown line) and atmospheric heat (light brown line) exchange, and advection (cyan line) terms. (b)  $\frac{\partial T_{2m}}{\partial t}_{\text{obs}}$  (pink line) and the  $\frac{\partial T_{2m}}{\partial t}_{\text{mod}}$  (blue line). (c) Scatter plot of  $\frac{\partial T_{2m}}{\partial t}_{\text{obs}}$  vs.  $\frac{\partial T_{2m}}{\partial t}_{\text{mod}}$ . The sloping solid line represents the 1 : 1 line, and the dashed line corresponds to the least-square best fit linear regression line between the two datasets. The correlation coefficient  $r$  and the linear equation fitting the best for the two datasets are also indicated. (d) Hourly evolution for September 2010 of the temperature variation for the different terms, the observations, and the model (same colors as in a and b).

of event is rare and ephemeral on a local scale (Llasat and Puigcerver, 1990; Conangla et al., 2018), its presence does not significantly bias the general performance of the model.

These statistics are estimated for each season as presented in Table 2. Correlation is high (0.82) and bias is low in summer. Nevertheless, the standard deviation remains high for this season, probably due to the higher values of temperature variation and the contribution of all terms involved for the summer in comparison to other seasons. For the spring and fall seasons, the correlation coefficient is still high (0.80). However, in winter the correlation coefficient is the lowest (0.67) with a higher bias ( $-0.31^{\circ}\text{C h}^{-1}$ ). Removing the extreme values of the observations and the model (i.e., taking the values that are within the 5th and 95th percentiles) gives better statistics (values in brackets in Table 2), confirming that the model encounters difficulties in reproducing these extreme events.

Figure 2d presents the hourly evolution of the five terms on the right side of Eq. (10), their sum (i.e.,  $\frac{\partial T_{2m}}{\partial t}_{\text{mod}}$  as blue solid line), and  $\frac{\partial T_{2m}}{\partial t}_{\text{obs}}$  (pink solid line) for the first days of September 2010. As expected, a diurnal cycle is identified for the radiative (both cloud and clear-sky terms) and heat atmosphere exchange terms. For the radiative terms, it is quite expected to have a positive (negative) contribution during daytime (nighttime) due to the presence (absence) of solar radiation. In addition, there are some days when  $R_{CL}$  (dotted red lines) plays an important modulating role in tempera-

**Table 2.** Statistics values between  $\frac{\partial T_{2m}}{\partial t}_{\text{obs}}$  and  $\frac{\partial T_{2m}}{\partial t}_{\text{mod}}$  for the four seasons of the year. The values in brackets correspond to the statistics within the 5th and 95th percentiles.

Season	Correlation coefficient	Bias ( $^{\circ}\text{C h}^{-1}$ )	Standard deviation ( $^{\circ}\text{C h}^{-1}$ )
Winter	0.67 (0.68)	$-0.31$ ( $-0.23$ )	0.43 (0.40)
Spring	0.80 (0.81)	$-0.15$ ( $-0.11$ )	0.52 (0.45)
Summer	0.82 (0.85)	$-0.18$ ( $-0.10$ )	0.55 (0.46)
Fall	0.79 (0.80)	$-0.21$ ( $-0.10$ )	0.46 (0.41)

ture variations, noticed during the day on 6 and 8 September, when clouds manage to provide a maximum cooling effect of  $-2.9^{\circ}\text{C h}^{-1}$ .

Evaporation and thermal conduction at the surface increase as the day goes on, and the larger these terms are the more they will prevent the increase in temperature linked to the sensible and latent turbulent fluxes. But overall the temperature increases during the day when these fluxes increase. However, the atmospheric heat exchange term (HA) is on average a negative contribution to temperature variations in 1 h when  $F_{\text{lat}}$  and  $F_{\text{sens}}$  increase in the afternoons. This negative contribution could be mostly associated with the mixing of masses of air at the top of MLD, where depending on the state of temperature inversion,  $T_{\text{MLD}}$  is likely to be very negative and this cold air could cool the surface. When  $R_{CL}$

presents an important contribution, HA becomes weaker and its diurnal cycle is attenuated due to the absence of solar radiation, and hence fewer surface fluxes are developed. The ground heat exchange and the advection terms play a minor role at this timescale and contribute negligibly to the model without a significant diurnal cycle detected.

Finally, the model follows the observations well, with a better agreement for daytime than for nighttime and with a better correlation for the first case (not shown). Reasons that may explain the bias are (1) the limited availability of  $T_{\text{MLD}}$  data, which are estimated using only two radiosoundings per day and not continuous hourly values, and (2) the hypothesis and assumptions in some variables and atmospheric conditions outlined in Eqs. (A3) and (A4). Further, the temperature variations at night might be hard to quantify accurately due to the minor contribution of each of the non-radiative terms, for which contributions are close to zero, especially for the HA term, and where  $F_{\text{lat}}$  and  $F_{\text{sens}}$  are very low most of the time (Sect. 3.3).

### 3.3 Annual and monthly–hourly cycles of the different terms

The mean contributions of each term are averaged monthly for both day and night (Fig. 3), whereas Fig. 4 presents the magnitudes of the monthly–hourly mean values from January 2009 to February 2014 of each term of the model. A residual term is calculated on these figures, which is calculated as the difference between the model (sum of all terms) and the observations (i.e.,  $\frac{\partial T_{2\text{m}}}{\partial t}_{\text{mod}} - \frac{\partial T_{2\text{m}}}{\partial t}_{\text{obs}}$ ).

According to Figs. 3b and 4b for nighttime, the clear-sky and cloud radiative forcing terms dominate the hourly temperature variations in magnitude at the surface, and the other terms remain close to zero. Still, during the night,  $\frac{\partial T_{2\text{m}}}{\partial t}_{\text{obs}}$  monthly mean is negative throughout the entire year, with higher negative values during warm months than during winter. Indeed,  $R_{\text{CL}}$  is always positive during the night and has an annual cycle more pronounced than the other terms, having a stronger effect in the cold months, due to the increase in cloud cover, especially low-level clouds, which enhance an increase in the downwards flux of LW radiation and whose effect gets weaker while approaching summer, due to the modification of the radiative effect of clouds for this season. This decrease in  $R_{\text{CL}}$  in summer explains the  $\frac{\partial T_{2\text{m}}}{\partial t}_{\text{obs}}$  annual cycle during nighttime.

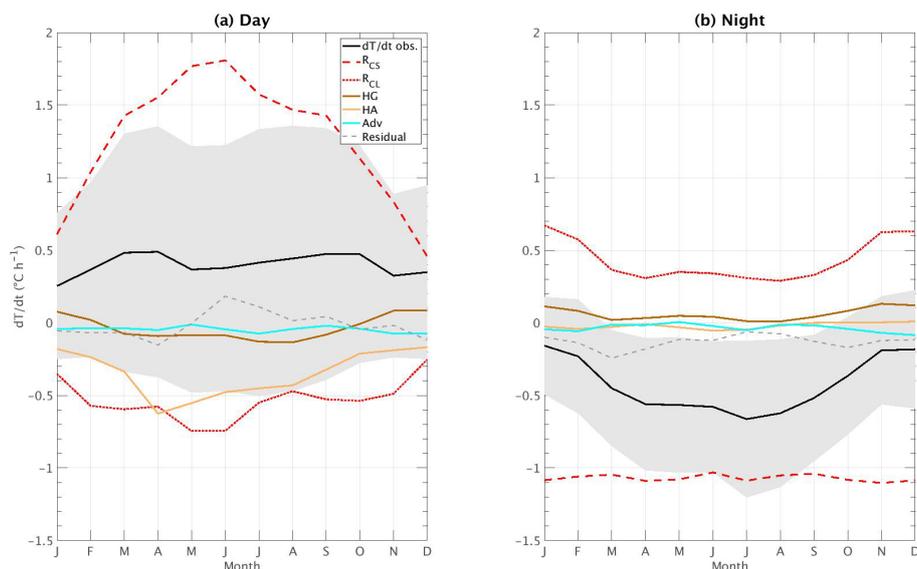
A diurnal cycle is pronounced for all the terms (Fig. 4).  $R_{\text{CS}}$  contributes the most in magnitude to local temperature variations during daytime (Fig. 3a), and the other terms damp its effect by providing a negative contribution. Furthermore, all the terms are mainly driven by the solar radiation intensity during the day. The diurnal and annual cycles of the  $R_{\text{CS}}$  ( $R_{\text{CL}}$ ) term are important, by having a positive (negative) contribution during the day, mostly dominating the temperature variations at this scale (as stated in Gaevskaya et al., 1962). The contribution of  $R_{\text{CL}}$  on  $\frac{\partial T_{2\text{m}}}{\partial t}_{\text{obs}}$  during daytime

is more important during May and June and finds its minimum for December and January, when the solar radiation is weak, thus preventing clouds from strongly reducing it. Further, a maximum negative (positive) contribution to temperature variations in  $R_{\text{CL}}$  ( $R_{\text{CS}}$ ) is found during the late mornings for all the months, cooling (warming) the surface with a maximum mean value of  $-1.3\text{ °C h}^{-1}$  ( $3.2\text{ °C h}^{-1}$ ) in May and June at 10:00 and 11:00 UTC.

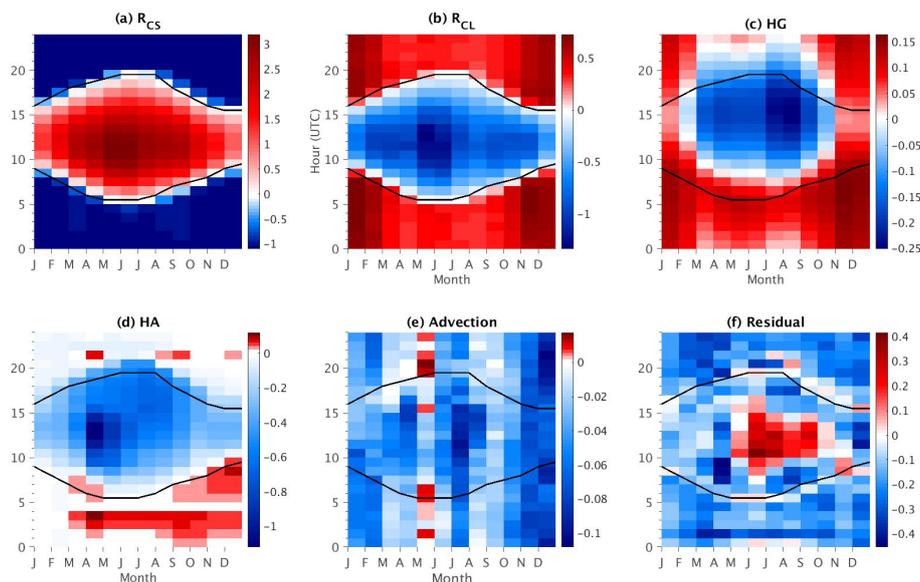
In addition, the HA term plays an important role in temperature variations for spring and fall seasons, when there is an increase in hourly solar radiation (Fig. 3a and 4d). These two seasons are characterized by an important contribution of this term in the late morning and the afternoon, in particular during spring, with an hourly maximum averaged contribution of  $-1.1\text{ °C h}^{-1}$ . The difference between day and night for the HA term is the largest during these months when the MLD can reach high values in the afternoon due to increased turbulence. For winter, its contribution is minimal due to the low development of the MLD. This result is in contrast with the negative (and very low) contribution of the HG term, which is maximal in summer in the afternoon with an average smaller value of  $-0.25\text{ °C h}^{-1}$  when the near-surface temperature is the highest and thus strengthens its cooling action during daytime. An opposite effect is found in winter when the ground is usually warmer than the surface, yielding to a positive contribution (Figs. 3a and 4c). The advection term does not present a strong monthly–hourly cycle compared to the other terms, although one can distinguish a mean negative action (still very low) to local temperature variations at all seasons with a mean minimum in July in the afternoon of  $-0.12\text{ °C h}^{-1}$ , as shown in Fig. 4e.

Lastly, the residual, defined as the difference between the model (sum of all terms) and the observations (i.e.,  $\frac{\partial T_{2\text{m}}}{\partial t}_{\text{mod}} - \frac{\partial T_{2\text{m}}}{\partial t}_{\text{obs}}$ ), also has a seasonal variability and is mainly negative, but a minimum difference of  $-0.45\text{ °C h}^{-1}$  between the two datasets is found in April between 07:00 and 09:00 UTC, which is partly related to the negative increase in the HA term at that time. This is due to the important absence of latent heat flux data especially for this month, which implies an increase in the bias when  $\tau_{\text{a}}$  is estimated (see Table 1 and Appendix A). Nevertheless, during the months when solar radiation is strong, the residual reaches positive values (with a maximum value around  $0.4\text{ °C h}^{-1}$  in June in late morning) but remains a low overestimation. These values for the residual are low compared to the magnitudes of  $R_{\text{CS}}$ , but it remains almost at the same order of magnitude as the  $R_{\text{CL}}$  and HA terms, and a good correlation in a diurnal cycle approach is found. Generally, this hourly mean residual could also be due to the simplifications and assumptions made in the model, undersampling, and energy imbalance.

Focusing on the transition periods (sunrise and sunset, black lines in Fig. 4), the residual presents low values at these times. Indeed, there is a slight underestimation of the model of about  $-0.13\text{ °C h}^{-1}$  for some months (e.g., February) at



**Figure 3.** Mean annual cycle averaged monthly from 2009 to 2014 of the five terms of the model, the observations, and the residual (dashed gray lines) for (a) day and (b) night. Same colors as in Fig. 2a and b. The shaded gray area in each of the panels represents the standard deviation of  $\frac{\partial T_{2m}}{\partial t}_{\text{obs}}$ .



**Figure 4.** Monthly–hourly mean values for (a)  $R_{CS}$ , (b)  $R_{CL}$ , (c) HG, (d) HA, (e) Adv, and (f) the residual (i.e., difference between the model and the observations). Units on the color bars are all in  $^{\circ}\text{C h}^{-1}$ , and their scale is different for each subfigure. The black contour line on each figure corresponds to sunrise (bottom line) and sunset (top line) approximate hours.

sunrise hours, whereas a low overestimation with close-to-zero residual mean values is found for May and June. For the sunset, a similar behavior is found (with very similar values for the residual term). Therefore, a good agreement is found between the model and the observations for these specific hours.

#### 4 Weight of the different terms acting on temperature variations: the random forest method

In this section, a random forest (RF) evaluation (James et al., 2013; Manish, 2016; Brownlee, 2016; Loh, 2002) is carried out at different timescales to estimate the relative weight (i.e., importance) of each term in temperature changes according to the hour, month, or season. This machine learning method

consists of bootstrapped–aggregated decision trees, a method that combines and gathers together the results of these trees to construct more powerful prediction models.

One of the most impressive features of RF is used here, which consists of the ability to provide a fully nonparametric estimation of the importance of each term (or predictor) for the model. One of the main advantages of this method is that it allows us to cover not only the impact of each term individually in the model but also the multivariate interactions with other predictors. Here, the model (i.e.,  $\frac{\partial T_{2m}}{\partial t}_{\text{mod}}$ ) has already been developed and defined as the sum of five terms. Therefore, to determine the importance of each term, the input data (the five terms) in the RF method are trained to predict the modeled temperature changes ( $\frac{\partial T_{2m}}{\partial t}_{\text{mod}}$ ), and so here the output of the RF method is still  $\frac{\partial T_{2m}}{\partial t}_{\text{mod}}$ . To know more on how the hyperparameters are tuned, how the data are split up into training and testing, and further information on the RF method, please refer to Appendix B.

Other approaches to estimate the predictor importance (e.g., simple squared marginal correlations method, squared standardized coefficients) do not give reliable results when the problem involves correlated predictors (Grömping, 2007). Additionally, this method is used as a result of the nonlinear relationship between each separated term and the model (not shown), which does not yield to an estimation and quantification on how each process at the surface affects the temperature variations, an approach suggested and used by Miller et al. (2017), who estimated the response and importance of some surface processes (such as  $F_{\text{lat}}$  and  $F_{\text{sens}}$ ) to the forcing radiative terms in Summit, Greenland.

#### 4.1 General behavior of the weight of the different terms

Firstly, the random forest method is used to establish which term dominates the temperature changes in the model ( $\frac{\partial T_{2m}}{\partial t}_{\text{mod}}$ ) during day and nighttime. The predictor importance estimate value is defined as the sum of the mean square error (MSE) of each term averaged over all decision trees used and normalized by the standard deviation taken over the trees. This principle consists of permuting the values of a term – predictor – in the decision trees where this term was left out-of-bag and assessing how much worse the MSE becomes after the permutation (James et al., 2013, Chap. 8). Thus, the larger this value, the more important the term. This importance estimation feature from the random forest method has been previously used to calculate the variable importance in different datasets for many applied problems in sciences or health fields for both regression and classification studies (e.g., Archer and Kimes, 2008; Genuer et al., 2012; Strobl et al., 2007).

Figure 5a and b present the predictor (i.e., each term involved in  $\frac{\partial T_{2m}}{\partial t}_{\text{mod}}$ ; see Eq. 2) importance estimate value for daytime and nighttime, respectively. Figure 5a corroborates that  $R_{\text{CS}}$  is the most important term, followed by  $R_{\text{CL}}$  and

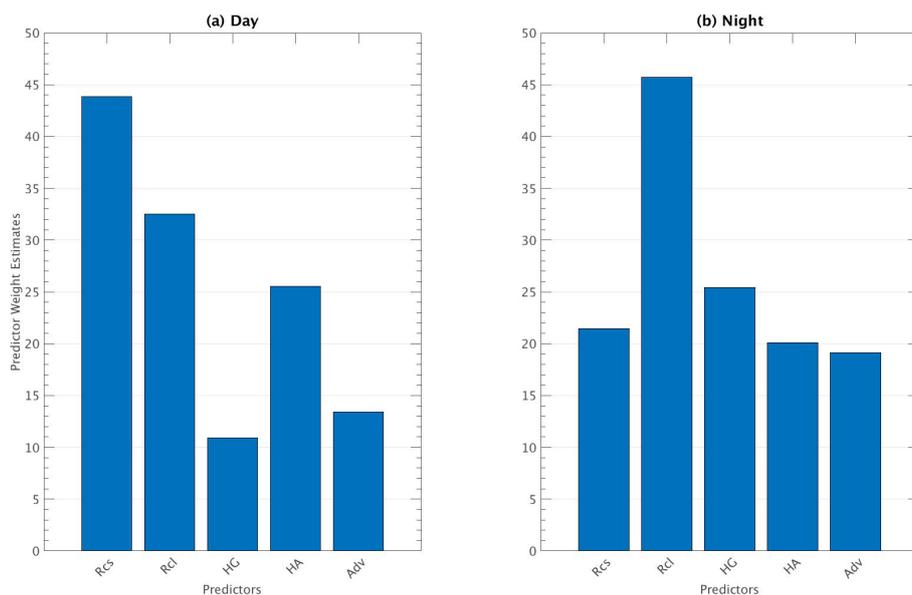
then HA, whereas HG is the least important term in the model developed for the timescale considered. Next, Fig. 5b illustrates that  $R_{\text{CL}}$  dominates hourly temperature variations during nighttime, followed by HG and  $R_{\text{CS}}$ . These results agree with Kukla and Karl (1993), who suggested that the key regulator of nighttime warming temperatures is downward thermal radiation when cloudy cases are presented, an effect that is damped most of the time by the radiative cooling effect of the ground.

#### 4.2 Diurnal cycle of the weights

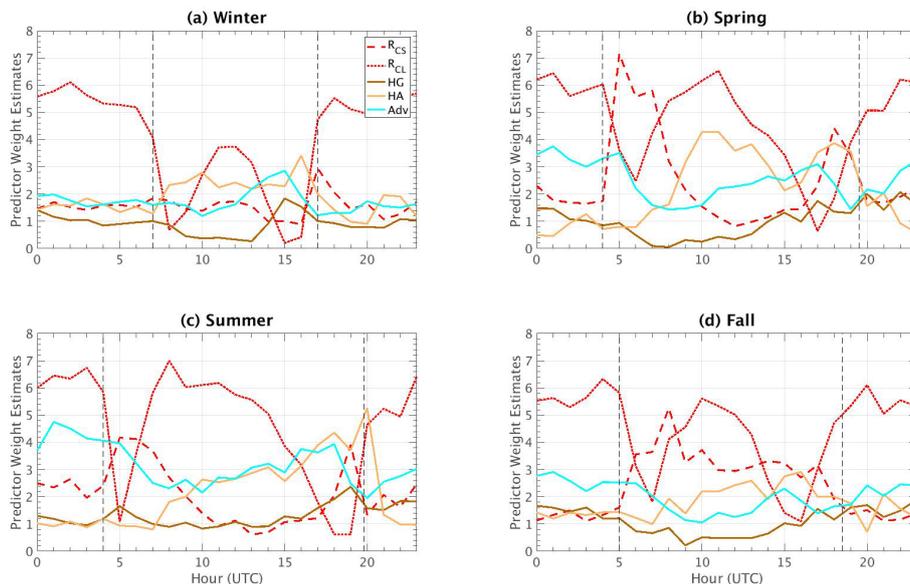
The importance estimation previously calculated for both day and nighttime periods considers all the processes occurring during each case and thus gives a global importance estimation. In order to separate the influence of each term on hourly temperature variations, an importance estimation value is performed for each hour independently. Figure 6 presents the results of this method for each season. This diurnal cycle is estimated by applying the random forest to each hour separately. As expected (and previously exposed), for all the seasons  $R_{\text{CL}}$  is the term dominating during nighttime, just after sunset, and before sunrise (indicated by vertical dashed lines in black).

After sunrise, the surface heating produced by the sun in the early morning enhances a high growth rate in the temperature variations, whose effect makes  $R_{\text{CS}}$  the dominant term driving  $\frac{\partial T_{2m}}{\partial t}_{\text{mod}}$  at those hours of the day for all seasons, except for winter. For this latter season (Fig. 6a), the growth rate of  $R_{\text{CS}}$  is almost the same as that of  $R_{\text{CL}}$ , and thus it does not expose an important estimation value. This effect is due to the weak mean solar zenith angle (SZA) for this season, and the surface heating by the sun in clear-sky conditions is not strong enough to dominate temperature variations. On the contrary,  $R_{\text{CL}}$  importance reaches its minimum just after sunrise and before sunset, where, depending on the season, either  $R_{\text{CS}}$  or HA is the main term controlling temperature variations.

A shift in importance between  $R_{\text{CL}}$  and  $R_{\text{CS}}$  occurs during the rest of the day for the four seasons of the year. This effect is explained by the variation in the data for these two terms:  $R_{\text{CS}}$  standard deviation for a given hour or season is weak, and thus its influence to explain the difference between one day and another at a specific hour remains minimal; it is not a strong predictor at the diurnal cycle scale, especially in the summer and spring (Fig. 6b and c, respectively) when other variables will modulate temperature variations. On the other hand,  $R_{\text{CL}}$  turns into the main modulator of  $\frac{\partial T_{2m}}{\partial t}_{\text{mod}}$  for all the seasons due to its strong standard deviation for a given hour or season, reaching its maximum importance in summer (Fig. 6c). Therefore, the hourly temperature variations are more sensitive to cloud changes rather than solar radiation, which does not vary significantly for a specific hour from day to day.



**Figure 5.** Predictor importance estimates obtained by the random forest method for (a) day and (b) night. The abscissa in both cases represents each predictor (or term) of the model, and the y axis represents their importance (unitless) defined as the sum of their mean square error when permutation in the decision trees is done.



**Figure 6.** Diurnal cycle of the predictor importance estimate for each term of the model, for (a) winter, (b) spring, (c) summer, and (d) fall. Same colors as in Fig. 2a and b. Vertical dashed black lines are for sunrise and sunset mean hours.

Concerning the other terms, HA importance grows along the day. Its variation is greater than that of the other terms, making it the second most important modulator most of the time (except for autumn, Fig. 6d). Its contribution is weak in winter due to the lack of solar radiation and vegetation, whose absence will diminish the turbulent heat fluxes measured at the surface, along with a weak MLD. Note that it sometimes becomes the most important term in the late af-

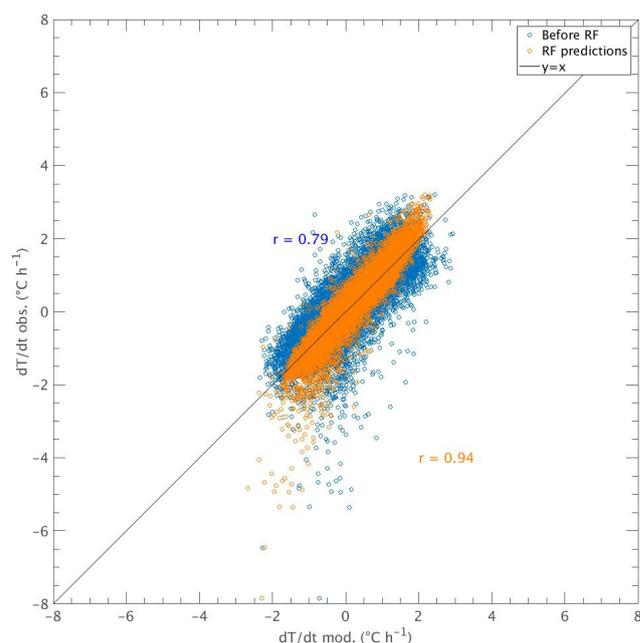
ternoon just before sunset. This is explained by an increase in instability of the ABL enhanced by a large boundary layer and near-surface temperature gradient (i.e., the difference between  $T_{MLD} - T_{2m}$  reaches its maximum; see the third term of the right side of Eq. 7) in the late afternoon (not shown), which is generated by the increase in turbulence fluxes near the surface.

Regarding the Adv term, it shows an important weight in some hours in the late afternoon in winter, which makes it the term controlling average hourly temperature variations at that time (then it is HA that becomes more important). In summer (Fig. 6c), it presents an important increase as the day goes on, similar to HA after 10:00 UTC, but HA is even more important thanks to a development of turbulent heat fluxes at the surface in the late afternoon. Adv becomes the second most important modulator during nighttime before sunrise for all seasons except summer. A thermally induced circulation linked to the urban heat island set up in Paris is probably at the origin of this importance: this circulation is likely to create a greater variation in temperature some nights when cooler air from rural areas is advected to warmer zones towards the city center. The SIRTAs observatory, located in a suburban area 20 km from the center of Paris, could be regularly affected by this modulation of circulation. Indeed, in this area the two predominant winds come from the S-E regime (the Siberian High) bringing mostly cold air temperatures and the S-W regime (air masses coming from the Atlantic Ocean) with warmer and more humid air (not shown).

Figure 6 supports a clear and reliable estimation of the importance of each term split into seasons for every hour of the day, which is not seen by estimating the diurnal and annual cycle contribution of each term to temperature variations (see Sect. 3). Indeed, depending on the hour of the day (and thus the state of the atmosphere), one term will become important over the rest, and temperature variations will be more sensitive to its change even if its contribution in terms of magnitude remains smaller compared to other terms. For instance,  $R_{CS}$  absolute contribution during nighttime is higher than that due to  $R_{CL}$  (Fig. 3b), and yet the latter is more important during nighttime due to its seasonal and hourly (not shown) contribution variability to  $\frac{\partial T_{2m}}{\partial t}_{obs}$ .

### 4.3 Validation of the random forest method

However, this machine learning method is generally used in other studies to train and to have better estimations of a particular model. In order to validate the random forest method skill on predicting new observed temperature variations, the method is used to predict  $\frac{\partial T_{2m}}{\partial t}_{obs}$  (rather than  $\frac{\partial T_{2m}}{\partial t}_{mod}$  as done to estimate the weight of each term in Sect. 4.1). The output for this case is called  $\frac{\partial T_{2m}}{\partial t}_{obs,RF}$ . A comparison between  $\frac{\partial T_{2m}}{\partial t}_{mod}$  (i.e., the linear sum of the five terms) and the new  $\frac{\partial T_{2m}}{\partial t}_{obs,RF}$  is done; the results of this validation are shown in Fig. 7. Indeed, the scatterplot before performing the random forest method ( $\frac{\partial T_{2m}}{\partial t}_{mod}$ ) shows the distribution of values between the observations and the model (i.e.,  $\frac{\partial T_{2m}}{\partial t}_{obs}$  vs.  $\frac{\partial T_{2m}}{\partial t}_{mod}$ ; blue points) as found in Fig. 2c. Then, when the random forest method is performed and the data are trained based on  $\frac{\partial T_{2m}}{\partial t}_{obs}$  (instead of  $\frac{\partial T_{2m}}{\partial t}_{mod}$ ), better predictions are obtained between  $\frac{\partial T_{2m}}{\partial t}_{obs}$  and  $\frac{\partial T_{2m}}{\partial t}_{obs,RF}$  (orange points), and the correlation coefficient has a higher value (0.94). In



**Figure 7.** Scatter plot of  $\frac{\partial T_{2m}}{\partial t}_{obs}$  as a function of the developed model  $\frac{\partial T_{2m}}{\partial t}_{mod}$  before applying the random forest method (blue circles) and  $\frac{\partial T_{2m}}{\partial t}_{obs}$  as a function of the model trained after the RF method is applied (orange circles).

such a case, the RF method gives better estimations of temperature variations, but the retrieval of the function used to obtain these results is not available. Nevertheless, this result validates considering temperature variations as the sum of the five terms to estimate their importance using the RF method (when it is used to predict the modeled temperature variations).

## 5 Discussion on the specific role of clouds in temperature variations

Section 4 shows that clouds are the main modulator of solar radiation on hourly temperature variations during the day (and the main contributor during the night). Knowing how and in what measure each term contributes to temperature variations, a deeper analysis is performed in this section in order to better understand the role of clouds. This analysis is performed by considering other variables available in the SIRTAs-ReOBS dataset to characterize both the atmosphere and the clouds.

In the following, only cloudy cases are considered. These cases are identified based on a criterion on the absolute value of cloud radiative effect (CRE) that must be higher than  $5 W m^{-2}$  in SW and LW (Chiriaco et al., 2018) during daytime, whereas for nighttime only LW is considered for this criterion. The SWCRE (and LWCRE) is calculating following Eq. (11):

$$\text{SWCRE} = F_{\text{SW}}^{\downarrow} - F_{\text{SW,CS}}^{\downarrow} \quad (11)$$

According to this threshold, cloudy conditions correspond to 82 % of the total cases from January 2009 to February 2014, and unique-clear-sky conditions represent the remaining 18 %.

## 5.1 Daytime analysis

Since the  $R_{\text{CS}}$  term dominates during daytime, the ratio of  $\frac{\partial T_{2\text{m}}}{\partial t}_{\text{obs}}$  divided by  $R_{\text{CS}}$  is created to estimate how much these two terms are driven by the solar radiation: when this ratio is close to 1, it means that the variations in temperature are the ones that would be expected in clear-sky conditions with no other modulators, and when this ratio deviates from 1 it means that the temperature variations are damped or enhanced by the other terms. The  $R_{\text{CL}}/R_{\text{CS}}$  ratio is also estimated, of which the distribution (not shown) has two peaks, one between  $-0.5$  and  $-1$  and one slightly negative with a tail in positive values. Thus, three bins are created from the highest cooling effect to the warming effect: (i)  $-1.0 \leq \frac{R_{\text{CL}}}{R_{\text{CS}}} < -0.5$  (bin 1), (ii)  $-0.5 \leq \frac{R_{\text{CL}}}{R_{\text{CS}}} < 0$  (bin 2), and (iii)  $0 < \frac{R_{\text{CL}}}{R_{\text{CS}}} \leq 0.5$  (bin 3). The cases with negatives or values of  $R_{\text{CS}}$  very close to zero (10 % of the cases), which occurred in early morning and late afternoon, are excluded since they affect the sign of the distribution or give divergent values. This histogram along with the one used in Sect. 5.2 are presented in Appendix C.

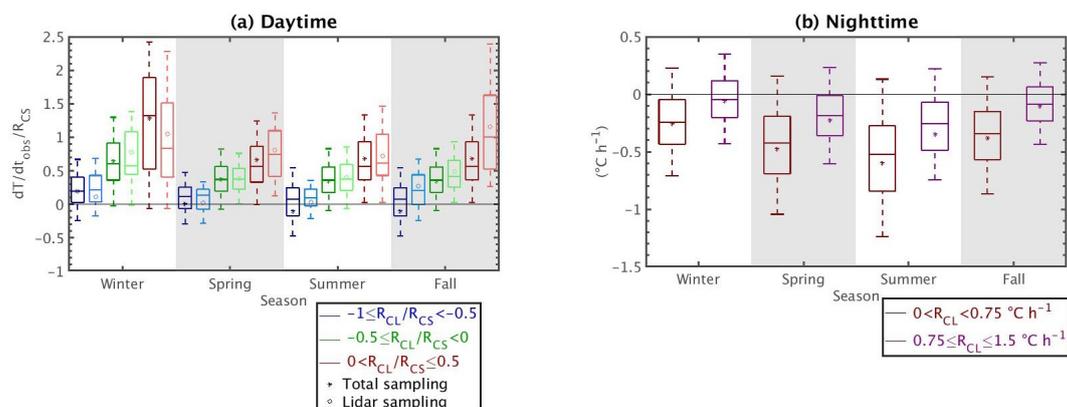
Figure 8a shows the values of the observed temperature variations divided by  $R_{\text{CS}}$ , and Fig. 9 shows the distribution (represented as box-and-whisker plots) of different relevant meteorological observations available in the SIRTAREOBS dataset, for each bin created, split into seasons for daytime. To add cloud information, lidar data are also considered, but because they are not available all the time, and in particular when it is raining (see Sect. 2.2), they are presented as additional boxplots (light color ones) that correspond to the lidar sampling (whereas the dark-color boxplots are for the total cloudy sampling). This difference of sampling mostly affects the first two bins when clouds have a cooling effect, for which both the occurrence and the amount of precipitation are the highest (not shown). Lidar SR( $z$ ) histograms presented in Fig. 10 are estimated by cumulating all lidar SR( $z$ ) observations available for one bin and one season in particular. The red horizontal lines on each histogram correspond to low- ( $P > 680$  hPa), mid- ( $440 < P < 680$  hPa), and high-level ( $P \geq 440$  hPa) cloud limits (Chepfer et al., 2010; Chiriaco et al., 2018). Note that a noise bin is presented here, which is simply the sum of the  $-999$  and  $-777$  bins that correspond to noisy and non-normalized profiles (see Sect. 2.2).

### 5.1.1 Case with strong cloud cooling effect – bin 1

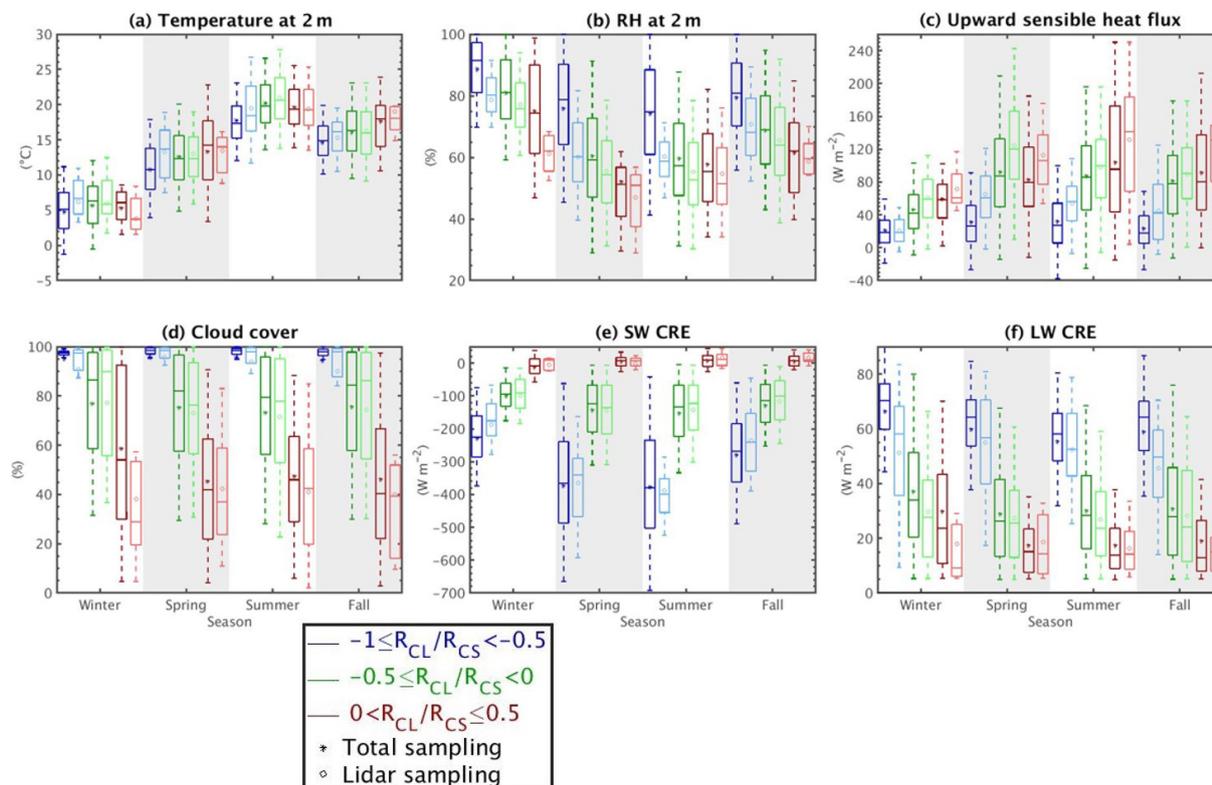
Figure 8a shows that the values of temperature variations (normalized by  $R_{\text{CS}}$ ) are the lowest for clouds having the most cooling effect (bin 1 in blue) for all the seasons. In winter this ratio stays mostly positive (temperature increase in 1 h) but less than 0.5 (less than half the increase that could be reached if the sky were clear), and it can become negative during summer and fall, i.e., the warming induced by  $R_{\text{CS}}$  can be counterbalanced by the clouds in these seasons. In addition, the cloud cover is almost total, and the radiative effects are strong, as seen in Fig. 9d, e, and f. In addition, the seasonal variability of SWCRE is directly related to the seasonal variability of  $F_{\text{SW,CS}}^{\downarrow}$ , and not (or only slightly) to the seasonal variability of cloud properties, since the ratio  $\text{SWCRE}/F_{\text{SW,CS}}^{\downarrow}$  remains almost constant during the entire year (not shown) and cloud cover presents the highest mean values (Fig. 9d).

The light blue bin is the same as the dark one except for lidar sampling, i.e., for non-rainy cases exclusively. Indeed, lower RH values (Fig. 9b) favor lower precipitation occurrence (not shown), higher amount of  $F_{\text{sens}}$  (Fig. 9c), and higher (i.e., less negative) values of SW CRE (Fig. 9e) for the lidar sampling compared to the total one. Logically, lower LWCRE is found for the lidar sampling. As expected, temperatures are higher for the lidar sampling for all the seasons except summer, as shown in Fig. 8a, along with a more frequent negative temperature variation (Fig. 8a). For this sampling that excludes rainy cases, both higher and lower values of SWCRE are found for spring rather than for summer (Fig. 9e): the minimal value is around  $\sim -600 \text{ W m}^{-2}$  for spring, whereas for summer it is  $\sim -500 \text{ W m}^{-2}$  (a situation not captured by the total sampling where summer has a lower value of SWCRE than spring). Figure 10b shows that low- and high-level clouds are more frequent for spring than for summer (Fig. 10c), and thus stronger negative values of SWCRE are found for spring. Then, the important presence of mid-level clouds with a high value of lidar SR( $z$ ) ( $> 80$ ) spotted in summer in Fig. 10c is potentially the cause of the strong negative values of SWCRE despite the very low presence of other clouds. These strong and negative SWCRE could be associated with a presence of nimbostratus clouds, due to the high SR detected for lidar, clouds which are more likely to form in summer because of the strong convective systems developed during that time due to higher surface temperatures. In addition, despite the strong negative values of SWCRE for this bin for all seasons, near-surface temperatures are not so low (Fig. 9a), partly due to the high LWCRE values (Fig. 9f) that slightly dampen the strong SWCRE.

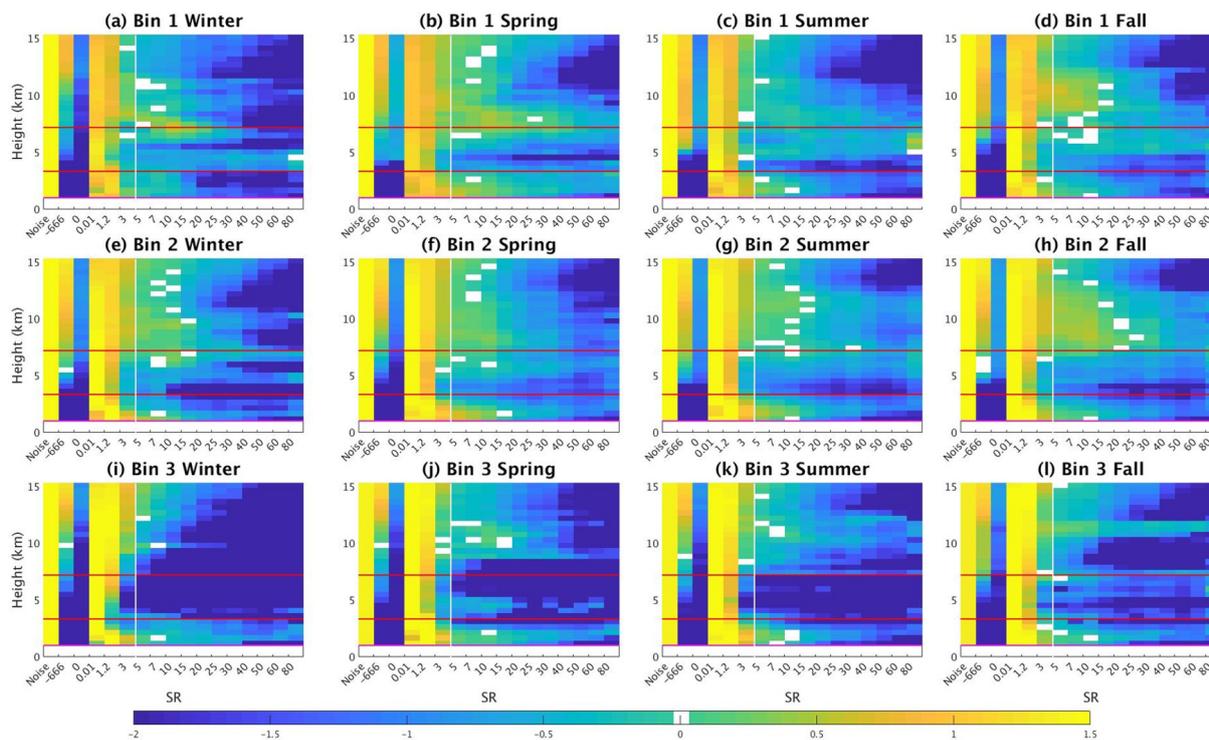
Some differences in cloud presence can be detected in Fig. 10a–d for the first bin. For the fall season, the SR( $z$ ) histogram in Fig. 10c exhibits an important presence of high-level thin clouds (especially above 8 km), along with mid-level thick clouds. Figure 10b also presents an important presence of high-level clouds in spring but within a smaller



**Figure 8.** (a)  $\frac{\partial T_{2\text{m}}/\partial t_{\text{obs}}}{R_{\text{CS}}}$  daytime values for the four seasons for three  $R_{\text{CL}}/R_{\text{CS}}$  bins:  $-1.0 \leq \frac{R_{\text{CL}}}{R_{\text{CS}}} < -0.5$  in blue,  $-0.5 \leq \frac{R_{\text{CL}}}{R_{\text{CS}}} < 0$  in green, and  $0 < \frac{R_{\text{CL}}}{R_{\text{CS}}} \leq 0.5$  in red. Dark-colored boxplots with the mean represented as "\*" correspond to cases when all meteorological variables are available at the same time (without considering lidar availability), whereas light-colored boxplots with "o" as their mean value represent the sample when both all meteorological variables and lidar profiles are available simultaneously. (b)  $\partial T_{2\text{m}}/\partial t_{\text{obs}}$  nighttime values for two  $R_{\text{CL}}$  bins:  $0 < R_{\text{CL}} < 0.75 \text{ } ^\circ\text{C h}^{-1}$  in maroon and  $0.75 \leq R_{\text{CL}} \leq 1.5 \text{ } ^\circ\text{C h}^{-1}$  in purple. Distributions are represented by box-and-whisker plots, where the boxes indicate the 25th and the 75th data percentiles, the whiskers indicate the 5th and 95th percentiles, the middle line represents the median, and the \* or o is the mean. Negative and close-to-zero values are removed (see text for further information).



**Figure 9.** Daytime values for (a) temperature, (b) relative humidity, (c) upward sensible heat flux at 2 m, (d) cloud cover retrieved from a sky imager, and (e) shortwave and (f) longwave cloud radiative effect. Colors and boxplots follow the same definition as in Fig. 8a.



**Figure 10.** Lidar scattering ratio ( $SR(z)$ ) histogram obtained by cumulating all lidar observations during daytime for bin 1 (a–d), bin 2 (e–h), and bin 3 (i–l) for winter (first column), spring (second column), summer (third column), and fall (fourth column). The color bar is the logarithm of the percentage of occurrence (the sum of one level is equal to  $\log_{10}100\%$ ); lidar data showed in each subplot start above the instrument’s recovery altitude ( $z = 1$  km); the red horizontal lines represent the limits of low-, mid-, and high-level clouds, and the white vertical line shows the threshold of cloud detection ( $SR(z) = 5$ ).

vertical range ( $7 < z < 10$  km) compared to fall. Indeed, one reason explaining the presence of these high-level clouds at these two transition seasons could be the convergence of a warm air with a cold air mass (which occurs more often in spring and fall), where the lighter warm air rises up to several kilometers from the ground and could form some cirrus clouds. In addition, spring exhibits high amounts of low-level thick clouds ( $SR > 40$ ) that are not detected the rest of the year, which could correspond to the moments just before a storm is set. Indeed, the highest precipitation rates are found for this season (not shown), and the cloud cover minimum in Fig. 9d for the lidar sample is 90 % (the highest among the other seasons), and fewer clear-sky conditions are found ( $0.01 < SR(z) < 1.2$ ).

### 5.1.2 Cases with weak cloud radiative effect: cooling or warming

When  $R_{CS}$  becomes dominant with respect to  $R_{CL}$  (bins 2 and 3), temperature variations are positive most of the time, especially in winter (Fig. 8a). In comparison with the first bin, the air becomes drier (higher sensitive heat flux and lower RH, Fig. 9b and c), which agrees with less cloud cover (Fig. 9d), and lower LWCRE and less negative SWCRE (Fig. 9e and f). Figure 10 shows fewer differences between

bin 1 and bin 2 ( $R_{CL} < 0$ ) than between bin 2 and bin 3 ( $R_{CL} > 0$ ) because this figure only considers non-rainy situations, and in Sect. 5.1.1. it is noted that there are important differences between the two samplings (total and non-rainy) for bin 1 for most of the variables. This difference between the two samplings is weaker for bins 2 and 3, despite its existence in winter and fall. Figure 10 is thus more representative of the full sampling for these two bins than it is for the first bin.

For the lidar sampling, LWCRE in winter and fall is slightly higher than for the two other seasons for the second bin (Fig. 9f). This is partly due to the presence of high-level thin clouds ( $5 > SR(z) > 20$ ), such as cirrus, detected in the histograms of  $SR(z)$  for this bin in Fig. 10e and h for these two seasons, which is slightly more than for spring and summer.

In bin 3, the absence of mid-level clouds and the lower occurrence of high clouds compared to bin 2 is obvious, while the difference for low-level clouds is not clear. This is consistent with lower LWCRE (Fig. 9f) but not necessarily with the very low values (i.e., close to zero) of SWCRE observed for bin 3 (Fig. 9e). The SWCRE values close to zero are explained by the fact that most of the hours corresponding to this bin 3 coincide just after sunrise (not shown), when solar

radiation is still weak, thus explaining that  $F_{\text{SW}}^{\downarrow} \approx F_{\text{SW,CS}}^{\downarrow}$ . These low SW and LWCRE values for the lidar sampling also correspond to low mean cloud cover (Fig. 9d) and thus a lower number of clouds are detected by the SR( $z$ ) histograms (Fig. 10i–l). Note that some of the SWCRE values are even positive, revealing that  $F_{\text{SW}}^{\downarrow} > F_{\text{SW,CS}}^{\downarrow}$ , an effect observed when the direct part of solar radiation is not fully attenuated and when some diffuse radiation reaches the surface in addition to the direct flux, radiation which is strengthened by scattering processes in clouds and atmospheric particles. This effect thus increases observed radiation beyond a respective clear-sky scenario, and it mostly happens in winter (but could occur the rest of the year) in early morning and/or late afternoon hours when the solar elevation angle is weak (Stapf et al., 2020; Wendisch et al., 2019). This distribution towards early morning hours explains that despite the high values of the ratio between temperature variations and  $R_{\text{CS}}$  (Fig. 8a) and lower cloud cover and cloud effects, near-surface temperatures are not always maximal for this bin (these maximal temperatures happen only for spring and fall).

## 5.2 Nighttime analysis

During nighttime, a similar procedure is followed to analyze and characterize different meteorological variables under different cloudy conditions. Since  $R_{\text{CL}}$  is the term dominating and controlling temperature variations (Sect. 4) and since  $R_{\text{CS}}$  is constant, it is not necessary to divide it by  $R_{\text{CS}}$ . Thus, two bins are created from the PDF of  $R_{\text{CL}}$ :  $0 < R_{\text{CL}} < 0.75 \text{ }^{\circ}\text{C h}^{-1}$  and  $0.75 \leq R_{\text{CL}} \leq 1.5 \text{ }^{\circ}\text{C h}^{-1}$ . During nighttime there is no SW radiation, so clouds always have a warming effect on near-surface temperature. The distribution of  $\frac{\partial T_{2\text{m}}}{\partial t}_{\text{obs}}$  for the two bins created for all the seasons is presented in Fig. 8b. Figure 11 shows the same meteorological variables shown in Fig. 9 but for nighttime cases. Cloud fraction profiles are not shown because the lidar does not operate during nighttime.

Figure 8b shows that the temperature variations are generally stronger for the first bin than for the second bin; i.e., the negative temperature variations induced by longwave cooling during nighttime are damped when the effect of clouds is stronger. This agrees with what is expected since, at hourly timescales, clouds are the most important factor in modulation of temperature variations during nighttime (see Sect. 4). As expected, the second bin has stronger LWCRE values (Fig. 11e). The effect of clouds seems enhanced when cool and moist air is present over SIRTa (Fig. 11a, b). More details for each bin are given in the following subsection.

### 5.2.1 Case with weak cloud warming effect

For the first bin ( $0 < R_{\text{CL}} < 0.75 \text{ }^{\circ}\text{C h}^{-1}$ ), LWCRE presents a wide range of values going from 5 up to  $75 \text{ W m}^{-2}$  with no seasonal variability detected (Fig. 11d), and yet temperatures

are the highest for this bin (Fig. 11a) except in winter. As expected, upward sensible heat flux is negative due to surface cooling most of the time, as values of  $\frac{\partial T_{2\text{m}}}{\partial t}_{\text{obs}}$  are predominantly negative (Figs. 11c and 7b, respectively), with drier air conditions since RH has the lowest values for all the seasons (Fig. 11b).

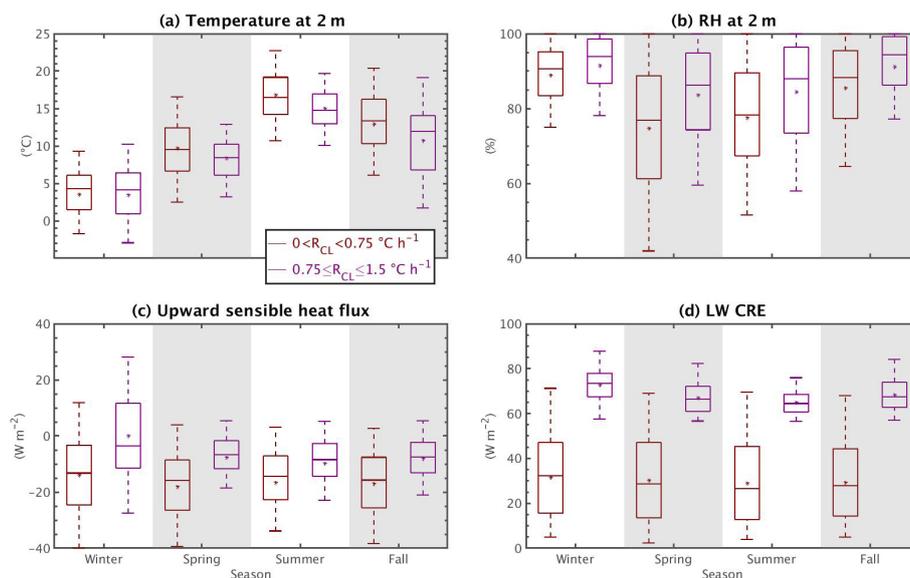
### 5.2.2 Case with strong cloud warming effect

The temperature distribution has the lowest values for all seasons (except spring) for the category of clouds that have the strongest warming effects (second bin, in purple in Fig. 11a). This category of clouds is tightly linked to strong LWCRE ranging from 55 up to  $90 \text{ W m}^{-2}$  (Fig. 11d). These low temperatures could be associated with situations when the incoming air masses are colder, which is linked to large-scale situations (Pinaridi and Masetti, 2000; Wang et al., 2005). Note that here hourly temperature variations are studied, and indeed on multi-day timescales, it is the large-scale atmospheric processes that determine the daily temperature value to the first order.

Finally, the distribution of sensible heat flux is positive in winter, with values reaching up to  $40 \text{ W m}^{-2}$  (purple boxplot, Fig. 11c). These unusual and positive values during nighttime are explained by the presence of a strong stable atmospheric layer, where near-surface temperatures are the lowest except for spring (Fig. 11a), and the atmosphere is warmer at higher altitudes since clouds contribute to warming the most at this time, and therefore a positive upward sensible heat flux develops. Similar behavior was also previously found by Miao et al. (2012) in Beijing during nighttime in cloudy conditions, where sensible and latent heat fluxes were close to zero due to the presence of a stable atmospheric layer, especially in winter when the sensible heat flux is found to be slightly greater than the latent heat flux. Indeed, almost half of temperature variation values are found to be positive for this season, as illustrated in Fig. 8b. Along with the positive upward sensible heat flux found, higher temperatures are found for this bin, which is not seen in the other seasons where the sensible heat flux is close to zero or mostly negative and temperatures are higher for the first bin (cloud with a low warming effect).

## 6 Conclusions

The European climate and most temperature anomalies are affected not only by the circulation of large-scale air masses (such as the NAO or blocking regimes) but also by smaller-scale processes such as cloud radiation and surface–atmosphere interactions located within the atmospheric boundary layer. In this paper, a method is developed and evaluated to quantify each process that affects hourly 2 m temperature variations on a local scale (near Paris), based almost exclusively on observations. The method exhibits good accuracy and is able to quantify the realistic diurnal and



**Figure 11.** Nighttime values for (a) temperature, (b) relative humidity, (c) upward sensible heat flux at 2 m, and (d) longwave cloud radiative effect. Colors and boxplots follow the same definition as in Fig. 8b.

monthly–hourly cycles of each term involved well, especially in summer when statistics are the best. The clear-sky radiative term represents the biggest positive (negative) contribution during daytime (nighttime), with values reaching up to  $4\text{ °C h}^{-1}$  ( $-1.7\text{ °C h}^{-1}$ ), whereas clouds cool (warm) up to  $-3.7\text{ °C h}^{-1}$  ( $1.4\text{ °C h}^{-1}$ ). The atmospheric heat exchange becomes predominant in the late afternoons when turbulent fluxes develop due to the increase in atmospheric instability. Some of the biases found are due to the difficulty of the model to reproduce smaller-scale processes such as the cold pool events well.

The use of a random forest analysis makes it possible to identify which term dominates over the others. Clear-sky radiation most influences the 1 h temperature variations during the day, whereas the cloud radiation has the most influence during the night followed by the ground heat exchange term. Nevertheless, separating the dataset into hours and seasons shows that the cloud radiation effect becomes predominant in several hours for all the seasons because it presents a greater variation in its hourly values than the clear-sky radiative effect, especially in spring and summer during daytime when more local cloud cover variability occurs. This cloud dominance is still found despite the fact that the clear-sky radiation magnitudes are the highest on a monthly–hourly scale. Temperature variation is then more sensitive to cloud changes within an hour rather than the large contribution of clear-sky radiation which mostly dominates the early mornings when surface heat fluxes are not yet developed. The other terms remain important as there are times when some of them modulate temperature variations, especially the turbulent heat fluxes in the late afternoons. These terms also remain necessary to obtain the best coefficient estimator be-

tween the directly measured observations and the method developed.

To better understand the importance of clouds on temperature variations at this scale, a deeper analysis is performed partly based on lidar profile observations. The atmosphere presents a high amount of cloud fraction detected by both the lidar and the sky imager during the daytime, which correlates well with the strong cooling contribution found for the cases when clouds have a strong cooling effect. Indeed, an important presence of mid-level thin and thick clouds is spotted for all the seasons (except spring) for these cases. These thick clouds could correspond to nimbostratus, which are very opaque and often produce continuous moderate rain, but the lidar detects them just before the rain starts. Indeed, temperatures at this time are very low, which contributes to these clouds forming since they usually form ahead of a warm front. On the contrary, a weak cooling effect is predominantly associated with low- and high-level thin clouds for all seasons. Situations with daytime positive cloud radiative effect occur when the atmosphere is close to clear-sky conditions and corresponds to low- and high-level thin clouds mostly in early morning and late afternoon. In addition, situations with weak cloud effect (either negative or positive) coincide with an important amount of high-level thick clouds for all the seasons (except winter) whose LWCRE is high, but SW clear-sky radiation controls temperature variations (Fig. 8a, bins 2 and 3). These high-level clouds are more present in weak cloud cooling and warming effects (bins 2 and 3) than the times of strong cooling effect (bin 1). Overall, a dominance of mid- and high-level clouds at the SIRTa observatory is detected for all cases. The dominant presence of this specific type of cloud is linked to the geographical po-

sition of the SIRTA observatory since it is located in a zone where warm and wet air coming from the Atlantic Ocean can nudge against the cold and dry masses of air coming from the Siberian region. This encounter of air masses (whether it is the warm air that encounters cold air or the opposite) tends to form mostly high-level clouds. Similar behavior has been found by Chakroun et al. (2018) during nighttime and Mariotti et al. (2015) over the Euro-Mediterranean area where the variability of cloud fraction (CF) is driven mostly by the encounter of two different air masses coming from northern Europe and southern Africa. During nighttime, even when clouds have a positive important radiative effect, temperatures are low and they are more controlled by large-scale processes since surface turbulent fluxes are low and there is no shortwave radiation most of the time. Since neither cloud fraction variable nor lidar data are available during nighttime, no information about the percentage of cloud presence can be known, yet variability on LWCRE is still found between clouds whose warming effect is weak and those whose warming effect is strong. This LWCRE variation is controlled by the presence of clouds to the first order, but it could also be due to the sensitivity of LW radiation arriving at the surface to the presence of atmospheric gases and aerosols and not only to cloud radiative properties, as shown previously by Dufresne et al. (2002), Satheesh and Krishna Moorthy (2005), and Kushta et al. (2014). In addition, the strong stability of the atmosphere at night creates positive sensible heat flux when clouds have the warmest effect on temperature variations.

The approach developed in this study is innovative because it is predominantly based on observations. It opens several perspectives on the possibility for continuing work: (i) use of the approach in combination with weather regimes to better relate the small-scale processes to the large-scale atmospheric dynamics, (ii) application of the approach to other locations to understand the spatial variability of the results and how specific local conditions affect each of the terms involved in near-surface temperature variations, (iii) identification of the moments when one term becomes predominant over the rest by integrating the time into larger scales (such as weekly and monthly), and (iv) estimation of how well these estimations are represented in the present and future climate simulations to evaluate models and/or understand the evolution of the different terms in a warming climate. Further, the ground heat exchange contribution is very site-dependent (more than the other surface terms), with different behavior in an urban environment; therefore it will be interesting to study its impact on the attenuation–amplification of maximum near-surface temperatures in periods of heatwaves.

## Appendix A: Detailed description of each term of the prognostic variations' temperature model at SIRTA observatory

The prognostic model used to study the temperature change at the surface, considering all the components driving the surface energy balance, can be simply written as the sum of four processes:

$$\frac{\partial T_{2m}}{\partial t} = R + HG + HA + Adv, \quad (A1)$$

where  $R$  is defined as the radiative forcing term,  $HG$  as the ground heat exchange term,  $HA$  as the atmospheric heat exchange term, and  $Adv$  as advection. Each of these terms represents a process involved in the variation in near-surface temperature  $\frac{\partial T_{2m}}{\partial t}$  and can be calculated using different measured and available variables as follows:

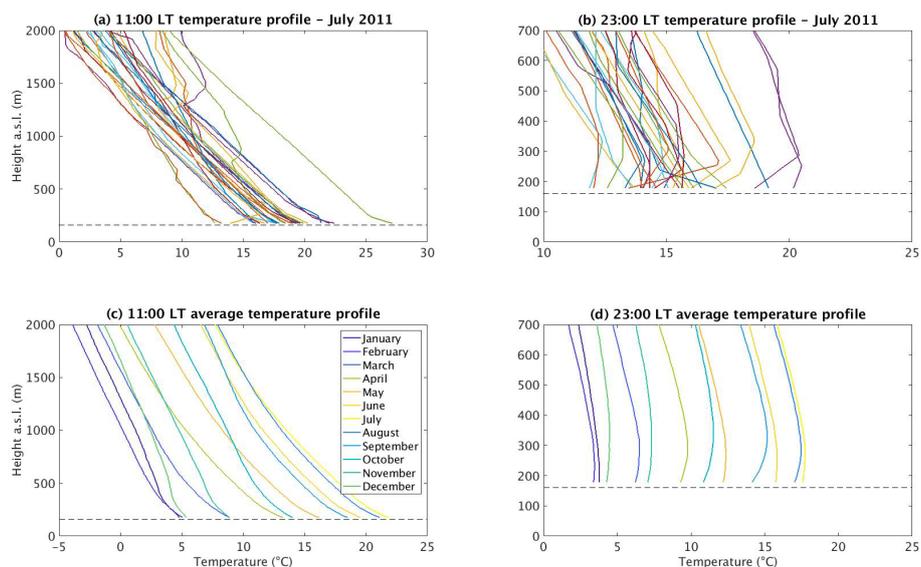
$$\frac{\partial T_{2m}}{\partial t} = \frac{\alpha + 1}{\rho c_p MLH} \Delta F_{NET} + \frac{T_s - T_{2m}}{\tau_s} + \frac{T_{MLD} - T_{2m}}{\tau_a} + \left( u_{10} \frac{\partial T_{2m}}{\partial x} + v_{10} \frac{\partial T_{2m}}{\partial y} \right), \quad (A2)$$

where  $T_{2m}$  is the near-surface temperature,  $t$  is the time,  $\alpha$  is a coefficient characterizing the form of the temperature profile in the boundary layer,  $\rho$  is the average air density of the boundary layer,  $c_p$  is the specific heat of air,  $MLH$  is the height of the boundary layer,  $\Delta F_{NET}$  is the net radiative flux at the surface,  $T_s$  is the temperature in the ground at 20 cm depth,  $T_{MLD}$  is the temperature at the top of the boundary layer,  $u_{10}$  and  $v_{10}$  are the zonal and meridional wind components, respectively, at 10 m above the ground,  $x$  is the zonal wind component towards the east and  $y$  is the meridional component wind towards the north, and  $\tau_s$  and  $\tau_a$  are defined as relaxation timescales for heat exchange processes in the ground and the atmosphere, respectively.

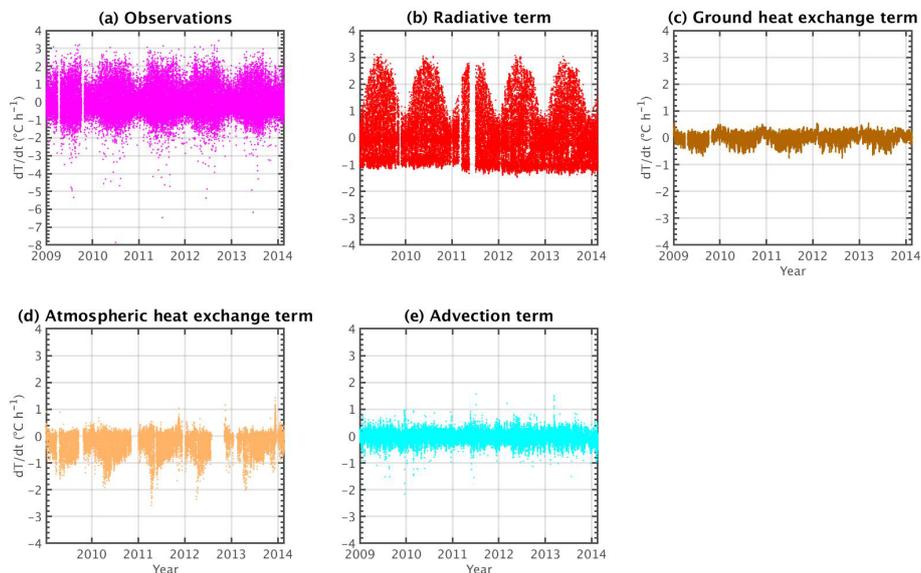
Before explaining how each term is estimated, considerations are necessary about the stability of the atmosphere and related temperature profiles. Note that these assumptions will not affect the physical behavior of the developed method; they are made in order to have a more quantitative treatment of the study.

Temperature near the surface and its behavior in lower layers are mostly driven by the quantity of net radiative flux arriving on it, which also depends on what moment of the day the temperature is observed; thus, it is important to differentiate day and night. During the first case mentioned, radiative flux is significant as a result of solar incoming flux radiation (shortwave radiation), and the vertical temperature profile will depend on the amount of this radiative flux (among other components). In the absence of any incoming shortwave radiation, which corresponds to nighttime, clouds and other processes control temperature variations at the surface and its vertical profile.

Figure A1 confirms that the planetary boundary layer (PBL) is on average unstable during daytime and stable



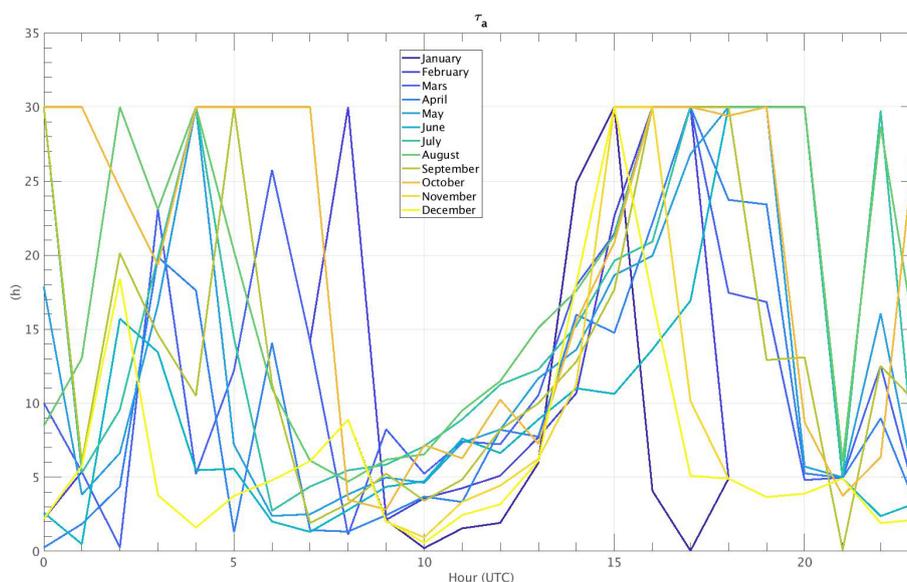
**Figure A1.** Temperature profiles at 11:00 (a) and 23:00 LT (b) for July of 2011, and monthly averaged temperature profiles from 2003 to 2017 at 11:00 (c) and 23:00 LT (d) in Trappes. The temperatures at each altitude are retrieved from radiosoundings launched twice a day at the mentioned hours. The horizontal dashed line represents the altitude above sea level where the Trappes observatory is located, which is 160 m.



**Figure A2.** Hourly evolution of observations (a),  $R$  term (b), HG term (c), HA term (d), and Adv term (e). Gaps in (d) are caused by the absence of MLD data at that moment.

during nighttime in the area of study. This figure is built from twice-daily radiosounding (at 11:00 and 23:00 local time) observations available in the SIRTA-ReOBS file from a METEO-France station located in Trappes (48.77° N, 2.01° E), 16 km away from the SIRTA observatory, to retrieve (every 15 m) the temperature and pressure up to 15 km above the ground. Figure A1a and b provide an overview of temperature profiles retrieved from these radiosoundings at 11:00 and 23:00 LT, respectively, for July of 2011 (each color rep-

resents a day of the month). Then, the monthly mean temperature profiles from 2003 to 2017, each month represented as one color, are plotted in Fig. A1c and d. As expected, the temperature decreases from the surface to the atmosphere in both daytime panels (Fig. A1a and c), and a temperature inversion occurs at lower layers in nighttime for the night cases (Fig. A1b and d) of the monthly mean.



**Figure A3.** Mean-hourly  $\tau_a$  value for each month of the year.

This said, a daytime and a nighttime equation are used to parametrize the temperature profile in the surface layer (SFL).

For the daytime approach, a linear temperature profile is established as follows:

$$T(z) = T_{2m} + \beta \cdot z, \quad (\text{A3})$$

where  $\beta$  is the temperature gradient, negative during the day.

For the stable nighttime case, the temperature profile can be defined by a polynomial approach (Stull, 1988):

$$T(z) = T_{\text{MLD}} - \left(1 - \frac{z}{\text{MLD}}\right)^\alpha (T_{\text{MLD}} - T_{2m}). \quad (\text{A4})$$

A shape parameter  $\alpha = 1.5$  shows a quasi-linear behavior of the temperature profile with a weak positive gradient near the surface, fitting well with the profiles in Fig. A1c and d (not shown).

The assumption of working with an ideal gas in an isobar environment, and applying the second law of thermodynamics, yields an expression of the enthalpy as described below (Malardel, 2009; Wallace and Hobbs, 2006):

$$h = c_p T. \quad (\text{A5})$$

Knowing the vertical temperature profile, Eq. (A5) can be integrated all over the MLD to find the total enthalpy per square meter at a given level (assuming that the density of the air is approximately constant at lower layers).

It is found that for a nighttime case, the total enthalpy can be expressed as

$$h_{\text{tot}} = \rho c_p \text{MLD} \left[ \frac{\alpha}{\alpha + 1} T_{\text{MLD}} + \frac{1}{\alpha + 1} T_{2m} \right]. \quad (\text{A6})$$

By deriving this equation with the near-surface temperature and total enthalpy, the ability to assess the change in near-surface temperature per unit change in enthalpy is estimated as

$$\frac{\partial T_{2m}}{\partial h_{\text{tot}}} = \frac{\alpha + 1}{\rho c_p \text{MLD}}. \quad (\text{A7})$$

For the daytime, proceeding as above, the same equation is found but for the nighttime, with the only exception that  $\alpha = 0$ .

As explained before, distinguishing between night and daytime is very important for the study due to the differences in the radiative flux arriving at the surface, which condition the behavior of the PBL (i.e., MLD) and SEB terms.

### A1 Radiative term

This term is calculated as

$$R = \frac{\alpha + 1}{\rho c_p \text{MLD}} \Delta F_{\text{NET}}. \quad (\text{A8})$$

The contribution of the radiative forcing to the temperature variations at SIRTa can be estimated by using the second law of thermodynamics, which states that the only energy a particle exchanges with its surroundings is heat. That said, it is assumed that the only heat exchanged for a particle of air in the low atmosphere with its surroundings is the net radiative flux divergence, yielding to

$$\frac{\partial h_{\text{tot}}}{\partial t} = F^\downarrow - F^\uparrow, \quad (\text{A9})$$

where  $F^\downarrow - F^\uparrow$  is the only source of energy of the particle. By replacing  $\partial h_{\text{tot}}$  in Eq. (A7) by Eq. (A9), it is found that

$$\frac{\partial T_{2\text{m}}}{\partial t} = \frac{\alpha + 1}{\rho c_p \text{MLD}} (F^\downarrow - F^\uparrow), \quad (\text{A10})$$

$$\frac{\partial T_{2\text{m}}}{\partial t} = \frac{1}{\rho c_p \text{MLD}} (F^\downarrow - F^\uparrow), \quad (\text{A11})$$

with Eqs. (A10) and (A11) corresponding to night and daytime cases, respectively. These two equations described above will allow estimation of the first term of the temperature variation model at SIRTa, the radiative term. MLD is a scale of height corresponding here to the height of the PBL retrieved from SIRTa-ReOBS. This value is set at this threshold because all the turbulent processes affecting the temperature variations are found within this layer. The PBL thickness varies depending, among other processes, on the amount of solar radiation, which enhances thermal and turbulent processes, making this thickness range from tens of meters to 2 km or more. Thus, an average PBL thickness value (i.e., MLD) at an hourly or monthly scale is set (an assumption that does not affect the physical behavior of the model), as seen in Fig. A1c. Since during nighttime the boundary layer depth is difficult to estimate because of the weak turbulence due to the absence of solar radiation and its very complex dynamical system (Shi et al., 2005; Walters et al., 2007; McNider et al., 2011), a fixed value of 350 m is set (Fig. A1d). Figure A2b shows the behavior of the radiative term calculated by using Eq. (A8) assuming  $\rho = 1 \text{ kg m}^{-3}$  and  $c_p = 1006 \text{ J kg}^{-1} \text{ K}^{-1}$ . A seasonal cycle is marked for this term, where during summer a peak of maximum contribution is found, whereas in winter its impact on temperature variations decreases significantly.

## A2 Atmospheric heat exchange term

This term is estimated as

$$\text{HA} = \frac{T_{\text{MLD}} - T_{2\text{m}}}{\tau_a}. \quad (\text{A12})$$

In this equation,  $\tau_a$  can be calculated at equilibrium by setting the left side of the Eq. (A2) to zero and neglecting the ground heat exchange and advection terms (i.e., for the temperature not to vary, a balance between these two latter terms must be set), leading to

$$\tau_a = \frac{\rho c_p \text{MLD}}{(\alpha + 1)} \cdot \frac{(T_{\text{MLD}} - T_{2\text{m}})}{\Delta F_{\text{NET}}}. \quad (\text{A13})$$

For the atmospheric heat exchange,  $\Delta F_{\text{NET}} = F_{\text{lat}} + F_{\text{sens}}$ . However,  $F_{\text{lat}}$  has lots of missing values in the SIRTa-ReOBS dataset (material defection) and therefore does not allow us to perform a complete analysis for the 5 years of study. Hence, an hourly and monthly look-up table (LUT) is created to have an estimation of  $\tau_a$ . Moreover,  $T_{\text{MLD}}$  is

not directly available in the SIRTa-ReOBS dataset. To retrieve that temperature, the mixing layer depth is first retrieved from the SIRTa-ReOBS dataset, by looking for the nearest radiosounding in time (two per day) and finding the pressure corresponding to this altitude; then it is possible to obtain the temperature at that pressure level by looking at them in the ERA5 dataset. Figure A3 shows an average  $\tau_a$  for all the months of the year. During night, the relaxation timescale shows a higher month-to-month variability mainly due to the absence and gaps of the latent and sensible heat fluxes, which does not allow the reflection of a marked tendency of  $\tau_a$ , whereas for daytime a more clear behavior of  $\tau_a$  is spotted for all the months thanks to an increase in the availability of the data for these hours (not shown). Sometimes during nighttime, this relaxation time exceeds 30 h, probably because the turbulence is so weak at that time that the heat does not completely reach the surface boundary layer (SBL) established. Stull (1988) suggested that for the night cases, values of  $\tau_a$  are on the order of 7 to 30 h, depending on the state of the SBL. Thus, values are limited to 30 h. For daytime,  $\tau_a$  remains smaller because the turbulence is stronger thanks to convection occurring near the surface, leading to a faster communication of surface information across the lower layers.

The atmospheric heat exchange term hourly values are shown in Fig. A2d. Most of the values are negative, indicating that this term will modulate the positive contribution made by the radiative forcing term.

## A3 Ground heat exchange term

The ground heat exchange term is defined as the energy lost by heat conduction through the lower boundary. This term can then be determined as follows:

$$\text{HG} = \frac{T_s - T_{2\text{m}}}{\tau_s}. \quad (\text{A14})$$

On average, the near-surface temperature during daytime (nighttime) is higher (lower) than that at low depth, as has been shown previously (Al-Hinti et al., 2017; Popiel and Wojtkowiak, 2013). Hence, the assumption that the vertical profile of temperature within the ground at low depths follows approximately the same behavior as the atmospheric temperature profile can be established, and therefore the approach adopted for the atmospheric heat exchange term (see Appendix A2) can here be implemented, to evaluate the relaxation timescale for the ground heat exchange at a depth of  $H_s = 0.2 \text{ m}$  below the ground. The surface is considered to be warmer than the temperature at 20 cm below the ground during the day thanks to the solar radiation arriving. The opposite happens during night, when the ground loses important longwave radiation, and it gets cold faster than  $T_s$ . Knowing the temperatures at the top of the ground ( $T_{2\text{m}}$ ) and at 20 cm

below it ( $T_s$ ), the relaxation timescale results in

$$\tau_s = \frac{\rho_s c_{p,s} H_s}{(\alpha + 1)} \cdot \frac{(T_s - T_{2m})}{\Delta F_{\text{NET}}}. \quad (\text{A15})$$

The soil type at the SIRTa observatory is a mix of clay and limestone (obtainable from a regionally predominant type of soil map in France; Wulf et al., 2015). The approximative values for ground density  $\rho_s$  and heat capacity  $c_{p,s}$  for this type of soil are then

$$\rho_s = 1300 \text{ kg m}^{-3},$$

$$c_{p,s} = 1140 \text{ J kg}^{-1} \text{ K}^{-1}.$$

A mean value of  $\tau_s = 20$  h for the daytime and nighttime cases is found. Figure A2c presents the contribution of this term to the temperature variations, which is very minor but remains important to have a better agreement between the developed model and the directly measured observations, by the fact that adding or removing some units could make a difference in estimating the correlation between the two datasets.

#### A4 Advection term

$$\text{Adv} = \left( u_{10} \frac{\partial T_{2m}}{\partial x} + v_{10} \frac{\partial T_{2m}}{\partial y} \right) \quad (\text{A16})$$

The advection term is estimated as presented in Eq. (A16) using the ERA5 dataset of horizontal wind components at 10 m and temperature at 2 m above the ground. The purpose of using this dataset is to have another point in each horizontal axis to calculate the transport of the mass of air in zonal and meridional directions, between the SIRTa observatory and the immediately following grid box. For the period and timescale considered, the advection term mostly plays a minor role, as shown in Fig. A2e, compared to the other terms.

#### A5 Observed hourly temperature variation term

The left side of Eq. (A2) is defined as

$$\frac{\partial T_{2m}}{\partial t} = \frac{T_h - T_{h-1}}{h}, \quad (\text{A17})$$

where  $T_h$  is the temperature at 2 m above the surface at hour  $h$ , and  $T_{h-1}$  is the temperature at the previously considered hour.

Figure A2a presents the evolution of this term, denoting, as expected, a seasonal cycle. Temperature variations have stronger negative than positive values, reaching a rate of  $-6 \text{ }^\circ\text{C h}^{-1}$  or even  $-8 \text{ }^\circ\text{C h}^{-1}$  a few times in summer 2010.

The sum of all terms on the right side of Eq. (A2) is hereafter called  $\frac{\partial T_{2m}}{\partial t}_{\text{mod}}$ .

### Appendix B: Basic settings on the random forest method to study the weight of each term

In Sect. 4, the random forest method is used to determine the importance of the predictors (terms) for the near-surface tem-

perature variations. Detailed information on how this method works is given here.

The training algorithm for random forests applies the general technique of bagging, where the key to bagging is that trees are repeatedly fit to bootstrapped subsets of the observations. The bootstrapped term refers here to randomly choosing data that can be chosen several times to build decision trees (no selection restrictions). A training dataset is chosen randomly with replacement (bootstrapping) from the original dataset to create a decision tree. In regression techniques, the training dataset correspond to about two-thirds of the total of the sample. The remaining one-third of data not used to train that decision tree are used later as testing data but also to determine the importance of a specific term (James et al., 2013). This procedure is repeated for all the decision trees used in the random forest. Finally, it is not necessary to have a validation dataset in this study because the main interest of using this machine learning is to determine the importance of the terms for the model developed, as it is known that random forest protects against overfitting by constructing training samples through bootstrapping.

Some hyperparameters are tuned in order to optimize the analysis.

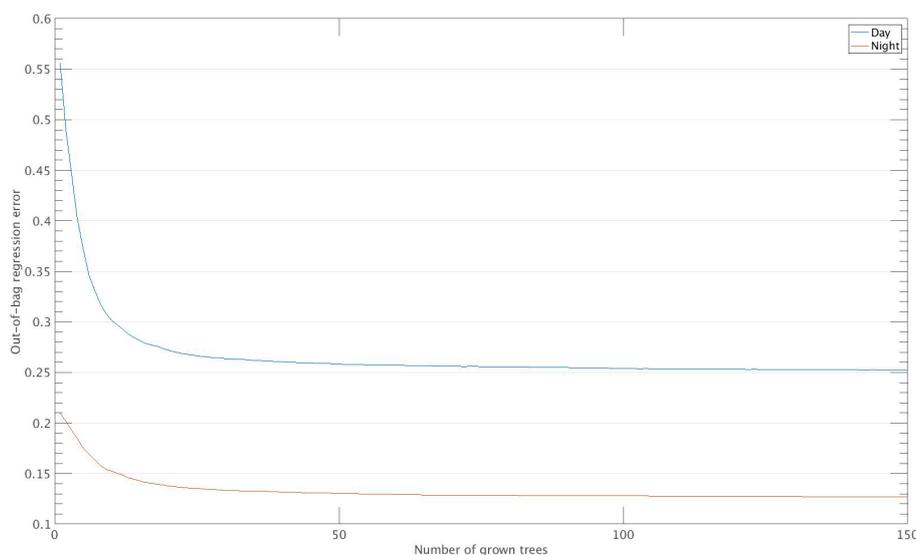
- The random forest method is set to have 150 decision trees, because at that number the error converges to a small value, as seen in Fig. B1 (converging value of 0.25 during daytime and 0.12 during nighttime).
- For the split criteria, since our model is a simple  $1^\circ$  regression, the method is set to use the mean square error (MSE) to do the split at each leaf.

The number of random features to consider at each split and the number of bootstrapped datasets used to train each decision tree in the random forest method is approximately two-thirds of the total predictors and two-thirds of the total samples, respectively (James et al., 2013).

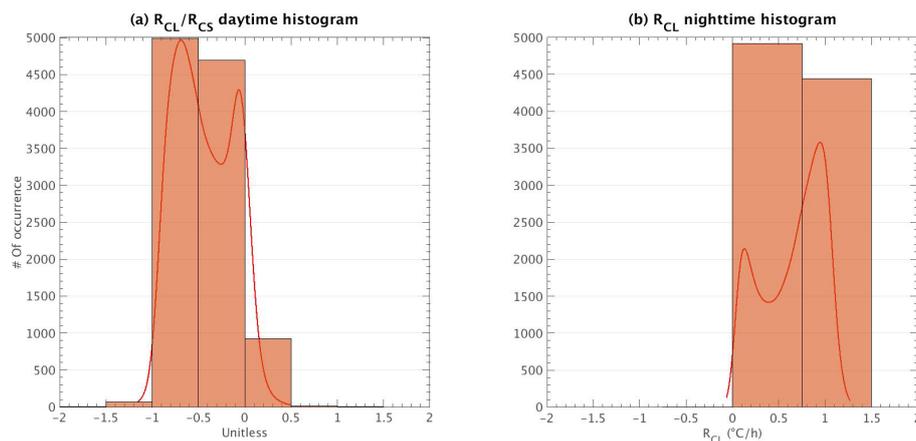
### Appendix C: Histograms for creation of the bins

In Sect. 5, different bins are created to analyze the specific contribution of clouds for both day and nighttime. Figure C1a shows the distribution of the ratio  $R_{\text{CL}}/R_{\text{CS}}$  for daytime hours, where two peaks can be easily identified: one between  $-0.5$  and  $-1$  and one slightly negative with a tail in positive values. From this distribution, three different bins can be created in order to separate the effect of clouds on observed near-surface temperature variations, going from the highest cooling effect to the warming effect (rectangular semi-transparent brown boxes in Fig. C1a): (i)  $-1.0 \leq \frac{R_{\text{CL}}}{R_{\text{CS}}} < -0.5$  (bin 1), (ii)  $-0.5 \leq \frac{R_{\text{CL}}}{R_{\text{CS}}} < 0$  (bin 2), and (iii)  $0 < \frac{R_{\text{CL}}}{R_{\text{CS}}} \leq 0.5$  (bin 3).

As for nighttime cases, bins are created only considering the distribution of  $R_{\text{CL}}$ , which is presented in Fig. C1b. Two



**Figure B1.** Out-of-bag error over the number of grown regression trees for day (blue line) and night (orange line).



**Figure C1.** (a) Daytime histogram of  $\frac{R_{CL}}{R_{CS}}$  and (b) nighttime histogram of  $R_{CL}$ . The red line in both figures represents the PDF, and the rectangular semi-transparent brown boxes represent the different bins created to analyze cloud influence. These histograms are built by considering only cloudy hours. Negative and close-to-zero values are removed for daytime hours (see text for further information).

peaks are spotted in this figure when  $R_{CL} > 0$ , and thus the following two bins are created:  $0 < R_{CL} < 0.75 \text{ °C h}^{-1}$  and  $0.75 \leq R_{CL} \leq 1.5 \text{ °C h}^{-1}$ . During nighttime there is no SW radiation, so clouds always have a warming effect on near-surface temperature.

**Data availability.** The SIRTA-ReOBS NetCDF file used in this study is available at <https://sirta.ipsl.fr/reobs.html> (last access: 10 April 2021) under <https://doi.org/10.14768/4F63BAD4-E6AF-4101-AD5A-61D4A34620DE> (SIRTA, 2021). The advection term and the temperature at the mixing layer depth were estimated using Copernicus Climate Change Service Information ERA5-Land and ERA5 [2009–2014], respectively.

**Author contributions.** OJRM carried out the data analysis and methodology and prepared all the figures. MC, SB, and JR contributed to the data analysis, advised on methods, and performed validation and interpretation of results. OJRM wrote the manuscript with contributions from MC and SB.

**Competing interests.** The contact author has declared that neither they nor their co-authors have any competing interests.

**Disclaimer.** Publisher's note: Copernicus Publications remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Acknowledgements.** The authors would like to acknowledge the Mesocentre ESPRI team at IPSL for providing storage resources and giving access to the SIRTA-Re-OBS dataset and also the SIRTA observatory for collecting and providing the geophysical variables and lidar data used in this study.

**Review statement.** This paper was edited by Johannes Quaas and reviewed by two anonymous referees.

## References

- Al-Hinti, I., Al-Muhtady, A., and Al-Kouz, W.: Measurement and modelling of the ground temperature profile in Zarqa, Jordan for geothermal heat pump applications, *Appl. Therm. Eng.*, 123, 131–137, <https://doi.org/10.1016/j.applthermaleng.2017.05.107>, 2017.
- Allan, R. P.: Combining satellite data and models to estimate cloud radiative effect at the surface and in the atmosphere, *Meteorol. Appl.*, 18, 324–333, <https://doi.org/10.1002/met.285>, 2011.
- Allan, R. P., Liu, C., Loeb, N. G., Palmer, M. D., Roberts, M., Smith, D., and Vidale, P.-L.: Changes in global net radiative imbalance 1985–2012, *Geophys. Res. Lett.*, 41, 5588–5597, <https://doi.org/10.1002/2014GL060962>, 2014.
- Archer, K. J. and Kimes, R. V.: Empirical characterization of random forest variable importance measures, *Comput. Stat. Data An.*, 52, 2249–2260, <https://doi.org/10.1016/j.csda.2007.08.015>, 2008.
- Arkin, P. A. and Meisner, B. N.: The Relationship between Large-Scale Convective Rainfall and Cold Cloud over the Western Hemisphere during 1982–84, *Mon. Weather Rev.*, 115, 51–74, [https://doi.org/10.1175/1520-0493\(1987\)115<0051:TRBLSC>2.0.CO;2](https://doi.org/10.1175/1520-0493(1987)115<0051:TRBLSC>2.0.CO;2), 1987.
- Arnold, N. S., Willis, I. C., Sharp, M. J., Richards, K. S., and Lawson, W. J.: A distributed surface energy-balance model for Haut Glacier d’Arolla, Valais, Switzerland, 42, 77–89, <https://doi.org/10.3189/S0022143000030549>, 1996.
- Bartolini, E., Claps, P., and D’Odorico, P.: Interannual variability of winter precipitation in the European Alps: relations with the North Atlantic Oscillation., *Hydrol. Earth Syst. Sci.*, 13, 17–25, <https://doi.org/10.5194/hess-13-17-2009>, 2009.
- Bastin, S., Chiriaco, M., and Drobinski, P.: Control of radiation and evaporation on temperature variability in a WRF regional climate simulation: comparison with colocated long term ground based observations near Paris, *Clim. Dynam.*, 51, 985–1003, <https://doi.org/10.1007/s00382-016-2974-1>, 2018.
- Bennartz, R., Shupe, M. D., Turner, D. D., Walden, V. P., Steffen, K., Cox, C. J., Kulie, M. S., Miller, N. B., and Pettersen, C.: July 2012 Greenland melt extent enhanced by low-level liquid clouds, *Nature*, 496, 83–86, <https://doi.org/10.1038/nature12002>, 2013.
- Broeke, M. V. D., Reijmer, C., As, D. V., and Boot, W.: Daily cycle of the surface energy balance in Antarctica and the influence of clouds, *Int. J. Climatol.* 26, 1587–1605, <https://doi.org/10.1002/joc.1323>, 2006.
- Brownlee, J.: Bagging and Random Forest Ensemble Algorithms for Machine Learning, *Machine Learning Mastery*, available at: <https://machinelearningmastery.com/bagging-and-random-forest-ensemble-algorithms-for-machine-learning/> (last access: 18 October 2021), 2016.
- Chakroun, M., Bastin, S., Chiriaco, M., and Chepfer, H.: Characterization of vertical cloud variability over Europe using spatial lidar observations and regional simulation, *Clim. Dynam.*, 51, 813–835, <https://doi.org/10.1007/s00382-016-3037-3>, 2018.
- Chen, T., Rossow, W. B., and Zhang, Y.: Radiative Effects of Cloud-Type Variations, *J. Climate*, 13, 264–286, [https://doi.org/10.1175/1520-0442\(2000\)013<0264:REOCTV>2.0.CO;2](https://doi.org/10.1175/1520-0442(2000)013<0264:REOCTV>2.0.CO;2), 2000.
- Chepfer, H., Bony, S., Winker, D., Cesana, G., Dufresne, J. L., Minnis, P., Stubenrauch, C. J., and Zeng, S.: The GCM-Oriented CALIPSO Cloud Product (CALIPSO-GOCCP), 115, D00H16, <https://doi.org/10.1029/2009JD012251>, 2010.
- Cheruy, F., Campoy, A., Dupont, J.-C., Ducharne, A., Hourdin, F., Haefelin, M., Chiriaco, M., and Idelkadi, A.: Combined influence of atmospheric physics and soil hydrology on the simulated meteorology at the SIRTA atmospheric observatory, *Clim. Dynam.*, 40, 2251–2269, <https://doi.org/10.1007/s00382-012-1469-y>, 2013.
- Chervakov, M.: Variability of Earth’s radiation budget components during 2009–2015 from radiometer IKOR-M data, *Geophys. Res. Abstr.*, EGU2016-9611, EGU General Assembly 2016, Vienna, Austria, 2016.
- Chiriaco, M., Bastin, S., Yiou, P., Haefelin, M., Dupont, J.-C., and Stéfanon, M.: European heatwave in July 2006: Observations and modeling showing how local processes amplify conducive large-scale conditions, *Geophys. Res. Lett.*, 41, 5644–5652, <https://doi.org/10.1002/2014GL060205>, 2014.
- Chiriaco, M., Dupont, J.-C., Bastin, S., Badosa, J., Lopez, J., Haefelin, M., Chepfer, H., and Guzman, R.: ReOBS: a new approach to synthesize long-term multi-variable dataset and application to the SIRTA supersite, *Earth Syst. Sci. Data*, 10, 919–940, <https://doi.org/10.5194/essd-10-919-2018>, 2018.
- SIRTA: The SIRTA-ReOBS Dataset, SIRTA [data set], <https://doi.org/10.14768/4F63BAD4-E6AF-4101-AD5A-61D4A34620DE>, last access: 10 April 2021.
- Conangla, L., Cuxart, J., Jiménez, M. A., Martínez-Villagrasa, D., Miró, J. R., Tabarelli, D., and Zardi, D.: Cold-air pool evolution in a wide Pyrenean valley, *Int. J. Climatol.*, 38, 2852–2865, <https://doi.org/10.1002/joc.5467>, 2018.
- Copernicus Climate Change Service: ERA5-Land hourly data from 2001 to present, *Climate Data Store* [data set], <https://doi.org/10.24381/CDS.E2161BAC>, 2019.
- ERA5: Fifth generation of ECMWF atmospheric reanalyses of the global climate, *Climate Data Store* [data set], available at: <https://cds.climate.copernicus.eu/cdsapp#!/home>, last access: 3 February 2020.
- Dai, A., Trenberth, K. E., and Karl, T. R.: Effects of Clouds, Soil Moisture, Precipitation, and Water Vapor on Diurnal Temperature Range, 12, 23, 2451–2473, [https://doi.org/10.1175/1520-0442\(1999\)012<2451:EOCSMP>2.0.CO;2](https://doi.org/10.1175/1520-0442(1999)012<2451:EOCSMP>2.0.CO;2), 1999.
- Dewitte, S. and Clerbaux, N.: Measurement of the Earth Radiation Budget at the Top of the Atmosphere – A Review, *Remote Sens.*, 9, 1143, <https://doi.org/10.3390/rs9111143>, 2017.
- Dione, C., Lohou, F., Chiriaco, M., Lohon, M., Bastin, S., Baray, J.-L., Yiou, P., and Colomb, A.: The Influence of Synoptic Circulations and Local Processes on Temperature Anomalies at Three

- French Observatories, *J. Appl. Meteorol. Clim.*, 56, 141–158, <https://doi.org/10.1175/JAMC-D-16-0113.1>, 2017.
- Dufresne, J.-L., Gautier, C., Ricchiazzi, P., and Fouquart, Y.: Longwave Scattering Effects of Mineral Aerosols, *J. Atmos. Sci.*, 59, 1959–1966, [https://doi.org/10.1175/1520-0469\(2002\)059<1959:LSEOMA>2.0.CO;2](https://doi.org/10.1175/1520-0469(2002)059<1959:LSEOMA>2.0.CO;2), 2002.
- Efthymiadis, D., Goodess, C. M., and Jones, P. D.: Trends in Mediterranean gridded temperature extremes and large-scale circulation influences, *Nat. Hazards Earth Syst. Sci.*, 11, 2199–2214, <https://doi.org/10.5194/nhess-11-2199-2011>, 2011.
- Gaevskaya, G. N., Kondrati'ev, K. Y., and Yakushevskaya, K. E.: Radiative heat flux divergence and heat regime in the lowest layer of the atmosphere, *Arch. Met. Geoph. Biokl. B.*, 12, 95–108, <https://doi.org/10.1007/BF02317955>, 1962.
- Genuer, R., Poggi, J.-M., and Tuleau-Malot, C.: Variable selection using Random Forests, 31, 2225–2236, <https://doi.org/10.1016/j.patrec.2010.03.014>, 2012.
- Grömping, U.: Estimators of Relative Importance in Linear Regression Based on Variance Decomposition, *Am. Stat.*, 61, 139–147, <https://doi.org/10.1198/000313007X188252>, 2007.
- Haefelin, M., Barthès, L., Bock, O., Boitel, C., Bony, S., Bouniol, D., Chepfer, H., Chiriaco, M., Cuesta, J., Delanoë, J., Drobinski, P., Dufresne, J.-L., Flamant, C., Grall, M., Hodzic, A., Hourdin, F., Lapouge, F., Lemaître, Y., Mathieu, A., Morille, Y., Naud, C., Noël, V., O'Hirok, W., Pelon, J., Pietras, C., Protat, A., Romand, B., Scialom, G., and Vautard, R.: SIRTa, a ground-based atmospheric observatory for cloud and aerosol research, *Ann. Geophys.*, 23, 253–275, <https://doi.org/10.5194/angeo-23-253-2005>, 2005.
- Hakuba, M. Z., Folini, D., Sanchez-Lorenzo, A., and Wild, M.: Spatial representativeness of ground-based solar radiation measurements, *J. Geophys. Res.-Atmos.*, 118, 8585–8597, <https://doi.org/10.1002/jgrd.50673>, 2013.
- Hartmann, D. L.: Chapter 6 Radiative Effects of Clouds on Earth's Climate, in: *International Geophysics*, Vol. 54, Elsevier, 151–173, [https://doi.org/10.1016/S0074-6142\(08\)60215-6](https://doi.org/10.1016/S0074-6142(08)60215-6), 1993.
- Hartmann, D. L., Ockert-Bell, M. E., and Michelsen, M. L.: The Effect of Cloud Type on Earth's Energy Balance: Global Analysis, *J. Climate*, 5, 1281–1304, [https://doi.org/10.1175/1520-0442\(1992\)005<1281:TEOCTO>2.0.CO;2](https://doi.org/10.1175/1520-0442(1992)005<1281:TEOCTO>2.0.CO;2), 1992.
- Ionita, M., Lohmann, G., Rimbu, N., and Scholz, P.: Dominant modes of Diurnal Temperature Range variability over Europe and their relationships with large-scale atmospheric circulation and sea surface temperature anomaly patterns: VARIABILITY OF THE SEASONAL DTR, *J. Geophys. Res.*, 117, D15111, <https://doi.org/10.1029/2011JD016669>, 2012.
- Ionita, M., Boronean, C., and Chelcea, S.: Seasonal modes of dryness and wetness variability over Europe and their connections with large scale atmospheric circulation and global sea surface temperature, *Clim. Dynam.*, 45, 2803–2829, <https://doi.org/10.1007/s00382-015-2508-2>, 2015.
- James, G., Witten, D., Hastie, T., and Tibshirani, R.: *An Introduction to Statistical Learning*, Springer New York, New York, NY, <https://doi.org/10.1007/978-1-4614-7138-7>, 2013.
- Kato, S., Loeb, N. G., Rutan, D. A., Rose, F. G., Sun-Mack, S., Miller, W. F., and Chen, Y.: Uncertainty Estimate of Surface Irradiances Computed with MODIS-, CALIPSO-, and CloudSat-Derived Cloud and Aerosol Properties, *Surv. Geophys.*, 33, 395–412, <https://doi.org/10.1007/s10712-012-9179-x>, 2012.
- Kauppinen, J., Heinonen, J., and Malmi, P.: Influence of Relative Humidity and Clouds on the Global Mean Surface Temperature, *Energ. Environ.*, 25, 389–399, <https://doi.org/10.1260/0958-305X.25.2.389>, 2014.
- Kukla, G. and Karl, T. R.: Nighttime warming and the greenhouse effect, *Environ. Sci. Technol.*, 27, 1468–1474, <https://doi.org/10.1021/es00045a001>, 1993.
- Kushta, J., Kallos, G., Astitha, M., Solomos, S., Spyrou, C., Mitsakou, C., and Lelieveld, J.: Impact of natural aerosols on atmospheric radiation and consequent feedbacks with the meteorological and photochemical state of the atmosphere, *J. Geophys. Res.-Atmos.*, 119, 1463–1491, <https://doi.org/10.1002/2013JD020714>, 2014.
- Li, Z. and Leighton, H. G.: Global climatologies of solar radiation budgets at the surface and in the atmosphere from 5 years of ERBE data, *J. Geophys. Res.-Atmos.*, 98, 4919–4930, <https://doi.org/10.1029/93JD00003>, 1993.
- Llasat, M. C. and Puigcerver, M.: Cold air pools over Europe, *Meteorol. Atmos. Phys.*, 42, 171–177, <https://doi.org/10.1007/BF01314823>, 1990.
- Loh, W.-Y.: Regression Trees With Unbiased Variable Selection and Interaction Detection, *Stat. Sinica*, 12, 361–386, 2002.
- Malardel, S.: *Fondamentaux de Météorologie*, 2nd Edn., Cépaduès-éditions, Toulouse, 2009.
- Manish: Tree Based Algorithms: A Complete Tutorial from Scratch (in R & Python), Analytics Vidhya, available at: <https://www.analyticsvidhya.com/blog/2016/04/tree-based-algorithms-complete-tutorial-scratch-in-python/> (last access: 18 October 2021), 2016.
- Mariotti, A., Pan, Y., Zeng, N., and Alessandri, A.: Long-term climate change in the Mediterranean region in the midst of decadal variability, *Clim. Dynam.*, 44, 1437–1456, <https://doi.org/10.1007/s00382-015-2487-3>, 2015.
- McNider, R. T., Steeneveld, G. J., Holtzlag, A. A. M., Pielke Sr., R. A., Mackaro, S., Pour-Biazar, A., Walters, J., Nair, U., and Christy, J.: Response and sensitivity of the nocturnal boundary layer over land to added longwave radiative forcing, *J. Geophys. Res.-Atmos.*, 117, D14106, <https://doi.org/10.1029/2012JD017578>, 2012.
- Miao, S., Dou, J., Chen, F., Li, J., and Li, A.: Analysis of observations on the urban surface energy balance in Beijing, *Sci. China Earth Sci.*, 55, 1881–1890, <https://doi.org/10.1007/s11430-012-4411-6>, 2012.
- Miller, N. B., Shupe, M. D., Cox, C. J., Noone, D., Persson, P. O. G., and Steffen, K.: Surface energy budget responses to radiative forcing at Summit, Greenland, *The Cryosphere*, 11, 497–516, <https://doi.org/10.5194/tc-11-497-2017>, 2017.
- Pal, S. and Haefelin, M.: Forcing mechanisms governing diurnal, seasonal, and interannual variability in the boundary layer depths: Five years of continuous lidar observations over a suburban site near Paris, 120, 11936–11956, <https://doi.org/10.1002/2015JD023268>, 2015.
- Parding, K., Olseth, J. A., Dagestad, K. F., and Liepert, B. G.: Decadal variability of clouds, solar radiation and temperature at a high-latitude coastal site in Norway, *Tellus B*, 66, 25897, <https://doi.org/10.3402/tellusb.v66.25897>, 2014.
- Pinardi, N. and Masetti, E.: Variability of the large scale general circulation of the Mediterranean Sea from observations

- and modelling: a review, *Palaeogeogr. Palaeoclimatol.*, 158, 153–173, [https://doi.org/10.1016/S0031-0182\(00\)00048-1](https://doi.org/10.1016/S0031-0182(00)00048-1), 2000.
- Popiel, C. O. and Wojtkowiak, J.: Temperature distributions of ground in the urban region of Poznan City, *Exp. Therm. Fluid Sci.*, 51, 135–148, <https://doi.org/10.1016/j.expthermflusci.2013.07.009>, 2013.
- Raschke, E., Ohmura, A., Rossow, W. B., Carlson, B. E., Zhang, Y.-C., Stubenrauch, C., Kottke, M., and Wild, M.: Cloud effects on the radiation budget based on ISCCP data (1991 to 1995), *Int. J. Climatol.*, 25, 1103–1125, <https://doi.org/10.1002/joc.1157>, 2005.
- Rebetez, M., Dupont, O., and Giroud, M.: An analysis of the July 2006 heatwave extent in Europe compared to the record year of 2003, *Theor. Appl. Climatol.*, 95, 1–7, <https://doi.org/10.1007/s00704-007-0370-9>, 2009.
- Satheesh, S. K. and Krishna Moorthy, K.: Radiative effects of natural aerosols: A review, *Atmos. Environ.*, 39, 2089–2110, <https://doi.org/10.1016/j.atmosenv.2004.12.029>, 2005.
- Shi, X., McNider, R. T., Singh, M. P., England, D. E., Friedman, M. J., Lapenta, W. M., and Norris, W. B.: On the Behavior of the Stable Boundary Layer and the Role of Initial Conditions, *Pure Appl. Geophys.*, 162, 1811–1829, <https://doi.org/10.1007/s00024-005-2694-7>, 2005.
- Stapf, J., Ehrlich, A., Jäkel, E., Lüpkes, C., and Wendisch, M.: Re-assessment of shortwave surface cloud radiative forcing in the Arctic: consideration of surface-albedo–cloud interactions, *Atmos. Chem. Phys.*, 20, 9895–9914, <https://doi.org/10.5194/acp-20-9895-2020>, 2020.
- Strobl, C., Boulesteix, A.-L., Zeileis, A., and Hothorn, T.: Bias in random forest variable importance measures: Illustrations, sources and a solution, *BMC Bioinformatics*, 8, 25, <https://doi.org/10.1186/1471-2105-8-25>, 2007.
- Stull, R. B.: An Introduction to Boundary Layer Meteorology, Springer Netherlands, <https://doi.org/10.1007/978-94-009-3027-8>, 1988.
- Trigo, R. M., Osborn, T. J., and Corte-Real, J. M.: The North Atlantic Oscillation influence on Europe: climate impacts and associated physical mechanisms, *Clim. Res.*, 20, 9–17, 2002.
- Vautard, R. and Yiou, P.: Control of recent European surface climate change by atmospheric flow, *Geophys. Res. Lett.*, 36, L22702, <https://doi.org/10.1029/2009GL040480>, 2009.
- Wallace, J. and Hobbs, P.: Atmospheric Science – An Introduction Survey, 2nd Edn., Academic Press, Burlington, MA, 504 pp., 2006.
- Walters, J. T., McNider, R. T., Shi, X., Norris, W. B., and Christy, J. R.: Positive surface temperature feedback in the stable nocturnal boundary layer, *Geophys. Res. Lett.*, 34, L12709, <https://doi.org/10.1029/2007GL029505>, 2007.
- Wang, K. and Dickinson, R. E.: Contribution of solar radiation to decadal temperature variability over land, *P. Natl. Acad. Sci. USA*, 110, 14877–14882, <https://doi.org/10.1073/pnas.1311433110>, 2013.
- Wang, Y., Xie, S.-P., Wang, B., and Xu, H.: Large-Scale Atmospheric Forcing by Southeast Pacific Boundary Layer Clouds: A Regional Model Study, *J. Climate*, 18, 934–951, <https://doi.org/10.1175/JCLI3302.1>, 2005.
- Wendisch, M., Macke, A., Ehrlich, A., Lüpkes, C., Mech, M., Chechin, D., Dethloff, K., Velasco, C. B., Bozem, H., Brückner, M., Clemen, H.-C., Crewell, S., Donth, T., Dupuy, R., Ebell, K., Egerer, U., Engelmann, R., Engler, C., Eppers, O., Gehrmann, M., Gong, X., Gottschalk, M., Gourbeyre, C., Griesche, H., Hartmann, J., Hartmann, M., Heinold, B., Herber, A., Herrmann, H., Heygster, G., Hoor, P., Jafariserajehlou, S., Jäkel, E., Järvinen, E., Jourdan, O., Kästner, U., Kecorius, S., Knudsen, E. M., Köllner, F., Kretzschmar, J., Lelli, L., Leroy, D., Maturilli, M., Mei, L., Mertes, S., Mioche, G., Neuber, R., Nicolaus, M., Nomokonova, T., Notholt, J., Palm, M., Pinxteren, M. van, Quaas, J., Richter, P., Ruiz-Donoso, E., Schäfer, M., Schmieder, K., Schnaiter, M., Schneider, J., Schwarzenböck, A., Seifert, P., Shupe, M. D., Siebert, H., Spreen, G., Stapf, J., Stratmann, F., Vogl, T., Welti, A., Wex, H., Wiedensohler, A., Zanatta, M., and Zeppenfeld, S.: The Arctic Cloud Puzzle: Using ALOUD/PASCAL Multiplatform Observations to Unravel the Role of Clouds and Aerosol Particles in Arctic Amplification, *B. Am. Meteorol. Soc.*, 100, 841–871, <https://doi.org/10.1175/BAMS-D-18-0072.1>, 2019.
- Wild, M.: Towards Global Estimates of the Surface Energy Budget, 3, 87–97, <https://doi.org/10.1007/s40641-017-0058-x>, 2017.
- Wild, M., Folini, D., Hakuba, M. Z., Schär, C., Seneviratne, S. I., Kato, S., Rutan, D., Ammann, C., Wood, E. F., and König-Langlo, G.: The energy balance over land and oceans: an assessment based on direct observations and CMIP5 climate models, *Clim. Dynam.*, 44, 3393–3429, <https://doi.org/10.1007/s00382-014-2430-z>, 2015.
- Willson, R. C., Gulkis, S., Janssen, M., Hudson, H. S., and Chapman, G. A.: Observations of Solar Irradiance Variability, *Science*, 211, 700–702, <https://doi.org/10.1126/science.211.4483.700>, 1981.
- Wulf, H., Mulder, T., Schaepman, M. E., Keller, A., Jörg, P. C., and Schaepman, M. E.: Remote sensing of soils, Research Report, University of Zurich, <https://doi.org/10.5167/uzh-109992>, 2015.
- Xoplaki, E., González-Rouco, J., Gyalistras, D., Luterbacher, J., Rickli, R., and Wanner, H.: Interannual summer air temperature variability over Greece and its connection to the large-scale atmospheric circulation and Mediterranean SSTs 1950–1999, *Clim. Dynam.*, 20, 537–554, <https://doi.org/10.1007/s00382-002-0291-3>, 2003.
- Xoplaki, E., González-Rouco, J. F., Luterbacher, J., and Wanner, H.: Wet season Mediterranean precipitation variability: influence of large-scale dynamics and trends, *Clim. Dynam.*, 23, 63–78, <https://doi.org/10.1007/s00382-004-0422-0>, 2004.