PONTIFICIA UNIVERSIDAD CATOLICA DE CHILE

ESCUELA DE INGENIERIA

DETERMINING ADVECTIVE FOG VARIABILITY AND STRATOCUMULUS CLOUD VERTICAL STRUCTURE COMBINING A THERMODYNAMIC CHARACTERIZATION OF THE MARINE BOUNDARY LAYER WITH A GROUND OPTICAL FOG OBSERVATION SYSTEM

ZEIDY L. HUAMÁN

Thesis submitted to the Office of Research and Graduate Studies in partial fulfillment of the requirements for the Degree of Master of Science in Engineering.

Advisor:
FRANCISCO SUÁREZ

Santiago de Chile, March 2022

PONTIFICIA UNIVERSIDAD CATOLICA DE CHILE
ESCUELA DE INGENIERIA

DETERMINING ADVECTIVE FOG VARIABILITY AND STRATOCUMULUS CLOUD VERTICAL STRUCTURE COMBINING A THERMODYNAMIC CHARACTERIZATION OF THE MARINE BOUNDARY LAYER WITH A GROUND OPTICAL FOG OBSERVATION SYSTEM

ZEIDY L. HUAMÁN SEVILLA

Members of the Committee:

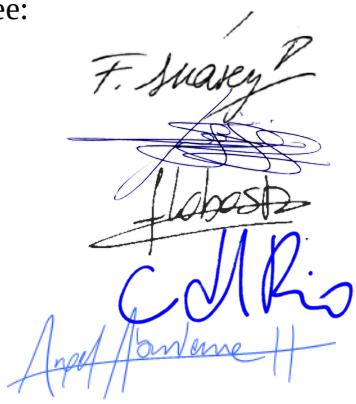
FRANCISCO SUÁREZ

JORGE GIRONÁS

FELIPE LOBOS

CAMILO DEL RÍO

ÁNGEL ABUSLEME



Thesis submitted to the Office of Research and Graduate Studies in partial fulfillment of the requirements for the Degree of Master of Science in Engineering.

Santiago de Chile, March 2022

ACKNOWLEDGMENTS

I would like to express my gratitude to my dear parents who are my continuous inspiration to move on and have instilled me to chasing my dreams with firmness and devotion. To my siblings who are my main motivation to give my best and never give up despite the obstacles.

I want to thank my advisor, Francisco Suárez, for trusting me and in my capabilities from the first moment. I am sure this would not have been possible without his careful supervision, his thoughtful encouragement and generous patience. To be his student, has been an extraordinary opportunity that allowed me to improve such as a person and professional.

Likewise, I am sincerely grateful with my best friends: Jhonny Valle, Daniela Pérez, Martín Morales, and Nicole Blin, for their unconditional support during this period. They encouraged me and got a smile out of me in my hardest days.

I also wish to thank my committee members who were more than generous with their expertise and precious time. A special thanks to Camilo del Río and Felipe Lobos, for their valuable mentoring, their kind endless help and unwavering enthusiasm for atmospheric sciences that kept me constantly engaged with my research.

I want to acknowledge the Programa Nacional de Becas y Crédito Educativo (PRONABEC-Perú), for the funding provided through the "Presidente de la República" scholarship to carry out my postgraduate studies. To the Agencia Nacional de Investigación y Desarrollo (ANID) for funding grants FONDECYT/11200789 and FONDECYT/1210221, to the Centro UC Desierto de Atacama and the Research Group for Earth Observation (Heidelberg University of Education) for providing me the

meteorological data for this research. Finally, I thank to the Centro de Desarrollo Urbano Sustentable (CEDEUS – ANID/FONDAP/15110020) and the Centro de Excelencia en Geotermia de los Andes (CEGA – ANID/FONDAP/15200001) for their support.

TABLE OF CONTENTS

LIST OF ABBREVIATIONS

CB: Cloud base

CT: Cloud top

DMC: Dirección Meteorológica de Chile (Meteorological Direction of Chile)

GOES: Geostationary Operational Environmental Satellite

GOFOS: Ground-based Optical Fog Observation System

LCL: Lifting Condensation Level

LRM: Logistic Regression Model

m ASL.: Meters above Sea Level

MBL: Marine boundary layer

MLRM: Multiple Linear Regression Model

MODIS: Moderate Resolution Imaging Spectroradiometer

Sc: Stratocumulus

SEPA: Southeast Pacific anticyclone

SLRM: Simple Linear Regression Model

θ: Potential temperature

q: Specific humidity

LIST OF FIGURES

LIST OF TABLES

ABSTRACT

The stratocumulus (Sc) cloud, responsible of the formation of advective fog at the coast of the Atacama Desert, is a valuable untapped freshwater source with potential to face the water scarcity in this region. However, little is known about features such as seasonal fog variability and cloud vertical structure that are essential to assess fog harvesting water potential. To investigate these features, a thermodynamic characterization of potential temperature (θ) and specific humidity (q) in the marine boundary layer (MBL), and a Ground Optical Fog Observation System (GOFOS) that directly measures local fog spatiotemporal variability, were combined. The vertical θ and q gradients are employed to find seasonal thresholds that define the MBL regimes related to fog formation (well-mixed MBL) and dissipation (stratified MBL); the lifting condensation level (LCL) and artificial neural networks (ANNs) are used to estimate the Sc cloud base (CB) elevation; and multiple linear regression models (MLRM) and ANNs are utilized to find the elevation of the Sc cloud top (CT). Even when a seasonal variability of the θ and q gradients was observed, the thresholds that defines the MBL regimes related to fog formation are nearly constant along the year. Fog detection using the θ vertical gradients agreed 95% of the time during winter and spring, whereas fog detection using the q gradients agreed in ~69% in the same seasons. A logistic regression analysis demonstrated that fog events can be classified with a good accuracy (0.82) using only θ as predictive variable. The CB estimated with ANNs show a better agreement (r^2=0.67) with GOFOS observations throughout the year than the LCL method; with errors smaller than 4%. Finally, both MLRM and ANNs show a good agreement with GOFOS observations to estimate CT (r^2=0.84 and r^2=0.92, respectively; with errors smaller than 4%). Our results reveal that the essential features of fog formation and its seasonal variability can be approximated through standard meteorological observations.

Keywords: advective fog, marine boundary layer, Atacama Desert, cloud base, cloud top.

RESUMEN

La nube estratocúmulo (Sc), responsable de la formación de niebla advectiva en la costa del Desierto de Atacama, es una valiosa fuente de agua dulce sin explotar con potencial para enfrentar la escasez de agua en esta región. Sin embargo, se sabe poco acerca de características como la variabilidad estacional de la niebla y la estructura vertical de la nube que son esenciales para evaluar el potencial hídrico de recolección de niebla. Para investigar estas características, se combinó una caracterización termodinámica de la temperatura potencial (θ) y la humedad específica (q) en la capa límite marina (MBL) con un sistema de observación óptico de niebla (GOFOS) que mide directamente la variabilidad espaciotemporal de la niebla. Los gradientes verticales θ y q se emplean para encontrar umbrales estacionales que definen los regímenes de MBL relacionados con la formación (MBL bien mezclada) y disipación de niebla (MBL estratificada); la elevación del nivel de condensación (LCL) y redes neuronales artificiales (ANN) se utilizaron para estimar la elevación de la base de la nube Sc (CB). Además, se utilizaron modelos de regresión lineal múltiple (MLRM) y ANNs para encontrar la elevación del techo de la nube Sc (CT). A pesar de que se observó una variabilidad estacional de los gradientes verticales de θ y q, los umbrales que definen los regímenes de MBL relacionados con la formación de niebla son casi constantes a lo largo del año. La detección de niebla usando los gradientes verticales de θ coincidió el 95% del tiempo durante el invierno y la primavera, mientras que la detección de niebla usando los gradientes q coincidió en ~69% en las mismas estaciones. Un análisis de regresión logística demostró que los eventos de niebla se pueden clasificar con una buena precisión (0,82), utilizando solamente θ como variable predictiva. Los CB's estimados con ANNs muestran una mejor concordancia (r^2 = 0.67) con las observaciones de GOFOS a lo largo del año, en comparación con el método LCL; con errores menores al 4%. Finalmente, tanto MLRM como ANN muestran una buena concordancia con las observaciones de GOFOS para estimar CT (r^2 = 0.84 y r^2 = 0.92, respectivamente; con errores menores al 4%). Los resultados de este estudio revelan que las características esenciales de la formación de niebla y su variabilidad estacional se pueden aproximar a través de observaciones meteorológicas estándar.

Palabras clave: niebla advectiva, capa límite marina, desierto de Atacama, base de la nube, techo de la nube.

I. INTRODUCTION

1.1. Background

Fog is a meteorological phenomenon that originates from the contact of saturated air masses with the Earth's surface (Fessehaye et al., 2014). Based on geographic characteristics, this phenomenon can be classified in three types: radiation, orographic and advective fog (Cereceda et al., 2002). Radiation fog occurs overnight when the air is cooled by contact with a cold surface, causing water vapor condensation in the air (Straub et al., 2012; Domen et al., 2014). Orographic fog happens when the humid air parcel rises due to the presence of the topography and adiabatically condenses (Lobos-Roco et al.,2018). Advective fog, which is the focus of this research, arises over the ocean where moist air cools as it passes over cooler waters, forming low-altitude stratocumulus (Sc) clouds that are then transported toward the coast (Domen et al., 2014; Fessehaye et al., 2014).

Fog has been investigated as a valuable freshwater resource with potential to face the water scarcity (Cereceda et al., 2008a, b; Klemm et al., 2012), and the sole water supply of fragile ecosystems in arid and semi-arid regions, such as those located at the coast of the Atacama Desert (Cereceda et al., 2002, 2008a; Osses et al., 2005; Pinto et al., 2006; Koch et al., 2019). Here, advective fog is frequent (Cereceda et al., 2002; del Río et al., 2021a). Its formation is related with the Sc cloud that occurs under statically stable lower-tropospheric conditions (Wood, 2012), and it is influenced by the interaction of two physical processes (Duynkerke, 1995; Lobos-Roco et al., 2018): subsidence and longwave radiative cooling (Figure 1a). As a result of the strong subsidence produced by the Hadley cell (Chang, 1995) and enhanced by the backward diurnal circulation between the Pacific Ocean and the Andes Altiplano (Rutllant et al., 2003; Lobos-Roco et al., 2021), a warm/dry air above the Sc cloud top (CT) is developed. When this warm/dry air interacts with the cold/wet and well-mixed marine boundary layer (MBL),

it generates a thermal inversion layer that together with the longwave radiative cooling promotes formation of semi-permanent Sc low clouds (Duynkerke et al., 1995; Bretherton and Park, 2009). The longwave radiative cooling generates a convective instability due to the heat exchange by entrainment of warm/dry air coming from the free troposphere that mixes into the MBL (See Figure 1a), raising the cloud base (CB) and thinning the cloud (Wood, 2012). The convective instability can be removed by the turbulence generated by cool downdrafts coming from the Sc CT and by warm compensating updrafts coming from the Sc CB that are generated by surface warming (Stevens et al., 2003). Finally, predominant S-SW winds transport the Sc cloud inland where it intercepts the coastal mountain range (del Río et al., 2021a,b).

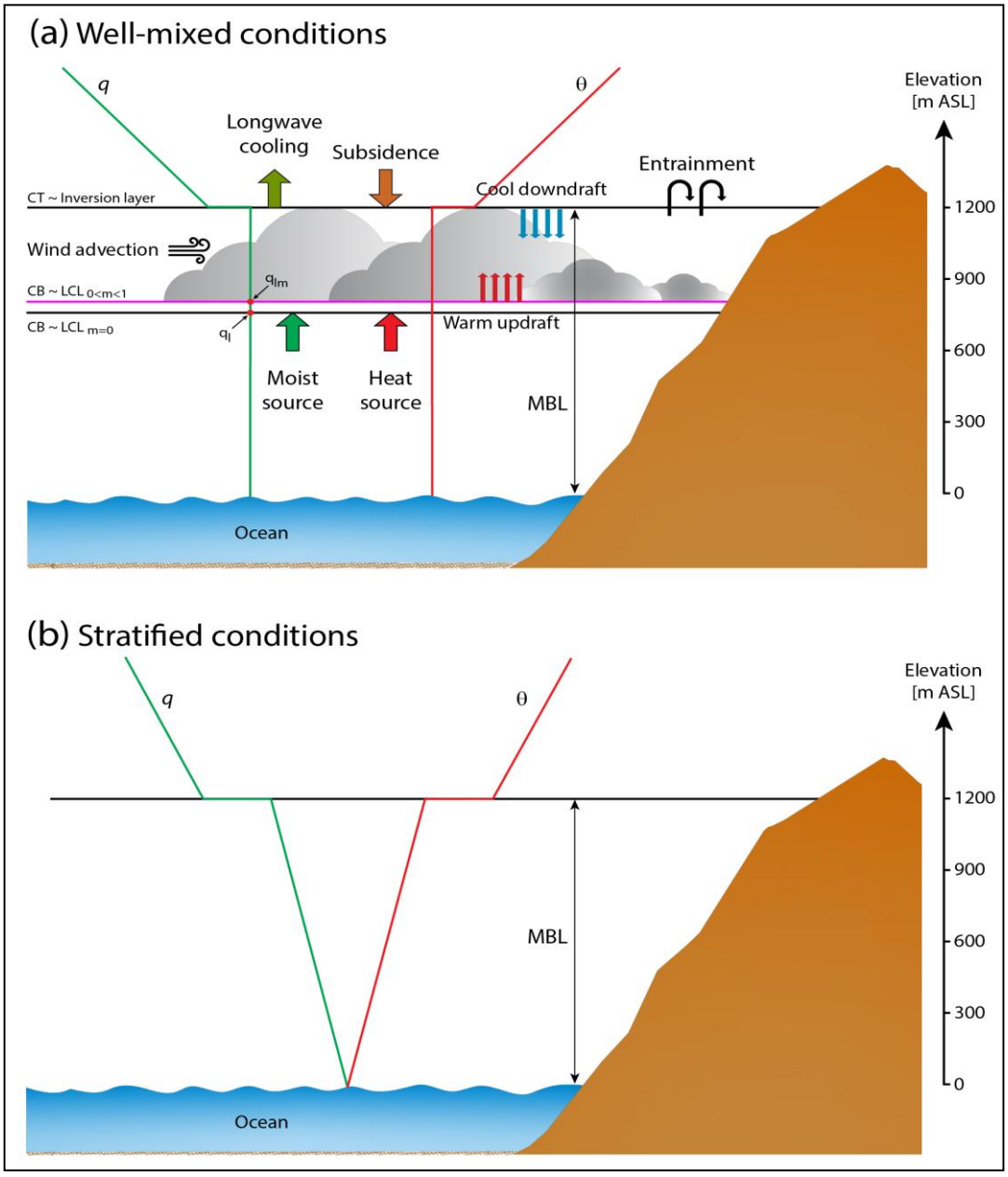


Figure 1. (a) Physical processes for the development of the Sc cloud that penetrates inland where the relief intercepts it and fog is formed (Duynkerke et al., 1995; Lobos Roco et al., 2018). The vertical profile of θ and q under well-mixed conditions within the marine boundary layer (MBL) is presented, as well as the identification of Sc cloud base (CB) and cloud top (CT), lifting condensation level (LCL), the adiabatic mixing ratio q_l, and the mixing ratio considering mixing with environment q_{lm}. (b) Profiles of θ and q under stratified conditions.

1.2. Motivation

To successfully find and assess optimal sites for fog harvesting, a better understanding of the features of the fog and low cloud (FLC) dynamics is needed. Relevant features of advective fog dynamics are fog formation, maintenance, and dissipation (duration), and Sc cloud vertical structure (Cereceda et al., 2002; Larrain et al., 2002; Domen et al., 2014; Lobos- Roco et al., 2018; del Río et al., 2021a).

1.3. Research gap

Some studies have been carried out to understand fog's annual variability (Muñoz et al., 2016; del Río et al., 2018, 2021b), its temporal and spatial distribution (Farias et al., 2005; Cereceda et al., 2008b; Torregrosa et al., 2016; Andersen & Cermak, 2018; del Río et al., 2021b), and its relationship with macroclimatic phenomena (Schulz et al., 2011; Muñoz et al., 2016; del Río et al., 2018). However, these studies have limitations related to the low quality of observations due to its low spatial and temporal resolutions. For instance, Farías et al. (2005) studied the spatiotemporal behavior of the FLC through Geostationary Operational Environmental Satellite (GOES) images every 90 min at the Coastal of the Atacama Desert during 2001-2003. They analyzed the FLC spatial dynamics for one representative month in winter and one representative month in summer, determining the daily and monthly fog cycles and its frequency. Nonetheless, this methodology has two main limitations. The first one is related to the characterization of the FLC vertical structure because satellite images do not precisely discriminate low clouds from fog. This issue occurs because the cloud is measured from above and there is little information regarding the proximity of the cloud base to the ground surface. The second limitation is the study period because they only investigated two months of the year. Unfortunately, these months did not include spring, which is the season that exhibits more fog frequency (e.g., See García et al., 2021).

Bendix et al. (2005) proposed an algorithm based on Moderate Resolution Imaging Spectroradiometer (MODIS) daytime data for Germany to distinguish between fog and

low clouds with a better spatial resolution than that used by Farías et al. (2005). Although this method represents a step forward in fog detection from space and in the characterization of the FLC vertical structure, the results revealed relatively high percentage errors and false alarm ratios due to the time lag between satellite overpass and ground observations.

To determine the FLC vertical structure, it is possible to estimate FLC CB and CT through sophisticated ground systems and remote sensing. For instance, CB height can be detected by lidar ceilometers that receive the backscattered profile with a vertical spatial resolution of ~15 m (Hogan et al., 2005; Serpetzoglou, 2008). However, these systems do not offer the global coverage of satellite data (Koračin et al., 2014). Lidar ceilometers also use complex algorithms to estimate a representative CB from the returned cloud signal (Jarraud, 2008). Satellite remote sensing, such as GOES (Ellrod and Gultepe, 2007) and MODIS (Bendix et al., 2005), allows identifying CB heights through passive visible and infrared imagery with better night performance than ceilometers, but with limitations to distinguish between ordinary low stratus clouds and fog at the terrain surface (Koračin et al., 2014). Likewise, CT can be calculated through active and passive remote sensing or with ground systems, but the most traditional method is using brightness temperatures from one or more radiosonde channels to calculate CT temperature, and to infer the height from a temperature profile, usually derived from a numerical model (Jarraud, 2008) with deviations of ~50 m (Bendix et al., 2004). However, it is also necessary an algorithm that extracts the CT height of extremely low stratus clouds by superimposing the initial binary stratus mask over a digital elevation model.

Although previous methods represent good alternatives for measuring CB and CT, these have important limitations such as high economic costs, complex computational resources, and low spatiotemporal resolution (Jarraud, 2008; Schween et al., 2020). Our motto is to investigate relatively simple and less expensive methods with acceptable accuracy to estimate Sc CB and CT and to understand the FLC vertical dynamic. Lobos-Roco et al. (2018) studied fog formation and dissipation by characterizing the MBL

under presence or absence of FLC. They found two MBL regimes based on the atmospheric thermal stability that defines FLC formation (well-mixed MBL) and dissipation (stratified MBL), through thresholds of vertical gradients of potential temperature (θ) and specific humidity (q). Also, they estimated the main FLC characteristics, such as CB and adiabatic liquid water mixing ratio (q_l), using the lifting-air parcel method (Wetzel, 1990). This method assumes that an air parcel is lifted dry adiabatically through the MBL without mixing during the ascension. Although this study is an important advance on this topic, their results are limited to only nine fog events during 2015. Recently, del Río et al. (2021a) proposed a novel Ground Optical Fog Observation System (GOFOS) that takes advantage of the local topography at the Atacama Desert to directly measure local FLC spatiotemporal variability and vertical structure. As described below, this system adds more realistic ground features to identify FLC dynamics and to understand the physics of the Sc cloud formation.

This study proposes a novel and economic methodology that combines the MBL thermodynamic characterization (Lobos-Roco et al., 2018) with GOFOS (del Rio et al., 2021a), to investigate the seasonal variability of advective fog formation/dissipation, and to characterize the FLC vertical structure using standard meteorological stations. Moreover, it aims to answer the following questions:

1.4. Research questions

1.4.1. General question:

How does the thermodynamic characterization of the marine boundary layer define the advective fog variability and the Stratocumulus cloud vertical structure, at the coast of the Atacama Desert?

1.4.2. Specific questions:

- How precise are the classification of the marine boundary layer regimes, when considering its seasonal variability to fog detection, at the coast of the Atacama Desert?
- How precise is the lifting-air parcel method to estimate the FLC base, at the coast of the Atacama Desert?
- Is it possible to estimate the altitude of the FLC base and Top through standard meteorological observations?

1.5. Hypothesis

The thermodynamic characterization of the Marine Boundary Layer through seasonal thresholds of θ and q gradients allows to predict the advective fog variability and estimate the main features of the Sc Cloud vertical structure, at the coast of the Atacama Desert.

1.6. Objectives

1.6.1. General objective:

To determine the advective fog variability and the Stratocumulus cloud vertical structure combining a thermodynamic characterization of the marine boundary layer with a Ground Optical Fog Observation System (GOFOS).

1.6.2. Specific Objectives:

- To find and validate seasonal thresholds of the θ and q vertical gradients that defines the fog formation (well-mixed conditions) and dissipation (stratified conditions).

- To evaluate the accuracy of the lifting-air parcel method to estimate the FLC Base in the study site.
- To estimate the elevation of FLC Base and Top using different methods populated with standard meteorological observations.

This document is organized in five main chapters: Chapter 1 reviews the relevance to study fog as potential alternative of freshwater in the face to the water scarcity and introduces the research gap existing in the scientific knowledge and literature. Likewise, the research questions, the hypothesis, and the objectives of this study are presented. Then, Chapter 2 reviews the main concepts related to the Marine Boundary Layer, Sc cloud base and Cloud top. Chapter 3 describes the methodology employed in this research. It includes the description of the study area and the selected meteorological stations, the GOFOS instrumentation, as well as the explanation of the data processing. Chapter 4 discusses the results from a quantitative and critical perspective. This section is divided as follows: thresholds to classify the Marine Boundary Layer, the estimations of FLC base and FLC top. Finally, Chapter 5 presents the main conclusions and outlooks for future work.

II. THEORICAL FRAMEWORK

2.1. Marine boundary layer regimes

Two types of MBL regimes are associated to advective fog: the well-mixed and the stratified regimes. On one hand, the well-mixed regime (Figure 1a) is associated with cloudy conditions (Muñoz et al., 2011) and fog formation (Lobos Roco et al., 2018). This regime presents a vertical structure within the MBL where the temperature follows a near–dry adiabatic lapse rate below the cloud and the moisture is adiabatic within the cloud layer (del Río et al., 2021b). On the other hand, the stratified regime (Figure 1b) is associated with clear conditions and fog dissipation, since atmosphere stability does not allow that the marine air layers mix by themselves and condensate (Muñoz et al., 2011). Lobos Roco et al. (2018) established the following thresholds, through the analysis of nine fog events during 2015, to classify the well-mixed and the stratified regimes as a function of the vertical gradients of potential temperature ($\partial\theta/\partial z$) and specific humidity ($\partial q/\partial z$):

$$\frac{\partial \theta}{\partial z} \begin{cases} < 3.10\times10^{-3}\ \frac{K}{m} & Well-mixed \\ > 3.10\times10^{-3}\ \frac{K}{m} & Stratified \end{cases} \quad \text{Eq. 1}$$

$$\frac{\partial q}{\partial z} \begin{cases} < 1.60\times10^{-3}\ \frac{g}{kg\ m} & Well-mixed \\ > 1.60\times10^{-3}\ \frac{g}{kg\ m} & Stratified \end{cases} \quad \text{Eq. 2}$$

where z is the height. These thresholds allow to describe the thermodynamic vertical structure of the Sc cloud and the diurnal variability of the fog formation/dissipation based on the MBL stability using θ and q to keep track of the movements of air parcels in an adiabatic process (Wang, 2013; Lobos Roco et al., 2018).

2.2. Cloud base

The CB is defined as the height from which the Sc cloud is formed. It corresponds to a visible accumulation of condensed water vapor (Forsythe et al., 2000; Zuidema et al., 2009). The FLC base can be represented by the lifting condensation level (LCL; see Figure 1a, Lobos Roco et al., 2018), since the FLC corresponds to a boundary layer cloud. At this elevation an air parcel saturates if it is lifted adiabatically (Wetzel, 1990; Daidzic, 2019).

The CB governs the longwave radiative properties from clouds to surface, determines the cloud emission temperature and strongly affects the incident infrared radiation at the surface (Sharma et al., 2016; Viúdez-Mora, et al., 2015). Moreover, the CB radiative heating is an important process that influences fog dissipation (Rogers and Koračin, 1992). It can contribute to destabilize and decouple the sub-cloud and the cloud layers when a temperature discontinuity is developed at the CB. In this case, vertical transport of moisture from the surface to the cloud layer would be restricted and it will restrict cloud development due to evaporation and an increase in CB height (Stull, 1988).

2.3. Cloud Top

The CT refers to the highest part of the Sc cloud, determined by the MBL top, where the thermal inversion layer occurs (Figure 1a). This inversion layer prevents the exchange between MBL and the free troposphere (Böhm et al., 2020). At the CT, there are two important processes that generate the mixing required to maintain/dissipate the Sc cloud (Bretherton and Park, 2009). The first one is the radiative cooling, which creates cool-air thermals that sink, generating the mixing that maintains the Sc cloud. The second one corresponds to the entrainment process at the top of MBL, which introduces air from the free troposphere, generating instability when drying and heating the humid and cold air of the MBL. This phenomenon causes the cloud to be forced to condense at higher altitudes. Thus, the CB height rises and can cause cloud dissipation if it reaches the MBL top (Stull, 1988).

III. MATERIALS AND METHODS

3.1. Study site

The study site is located on the coast of the Atacama Desert, northern Chile between 19° - 21°50’ S and 71° - 69 ° W in the Tarapacá region (Figure 2a). The coast of the Atacama Desert is characterized by a steep mountain range that rises a few km from the shore until reaching 1,500 m ASL. This topographic feature generates the optimal conditions for fog formation since the topography intercepts the MBL inland advection. Moreover, the steep and close-to-the-ocean topography brings exceptional conditions for observing the vertical distribution of the marine Sc cloud, working as a permanent vertical profile.

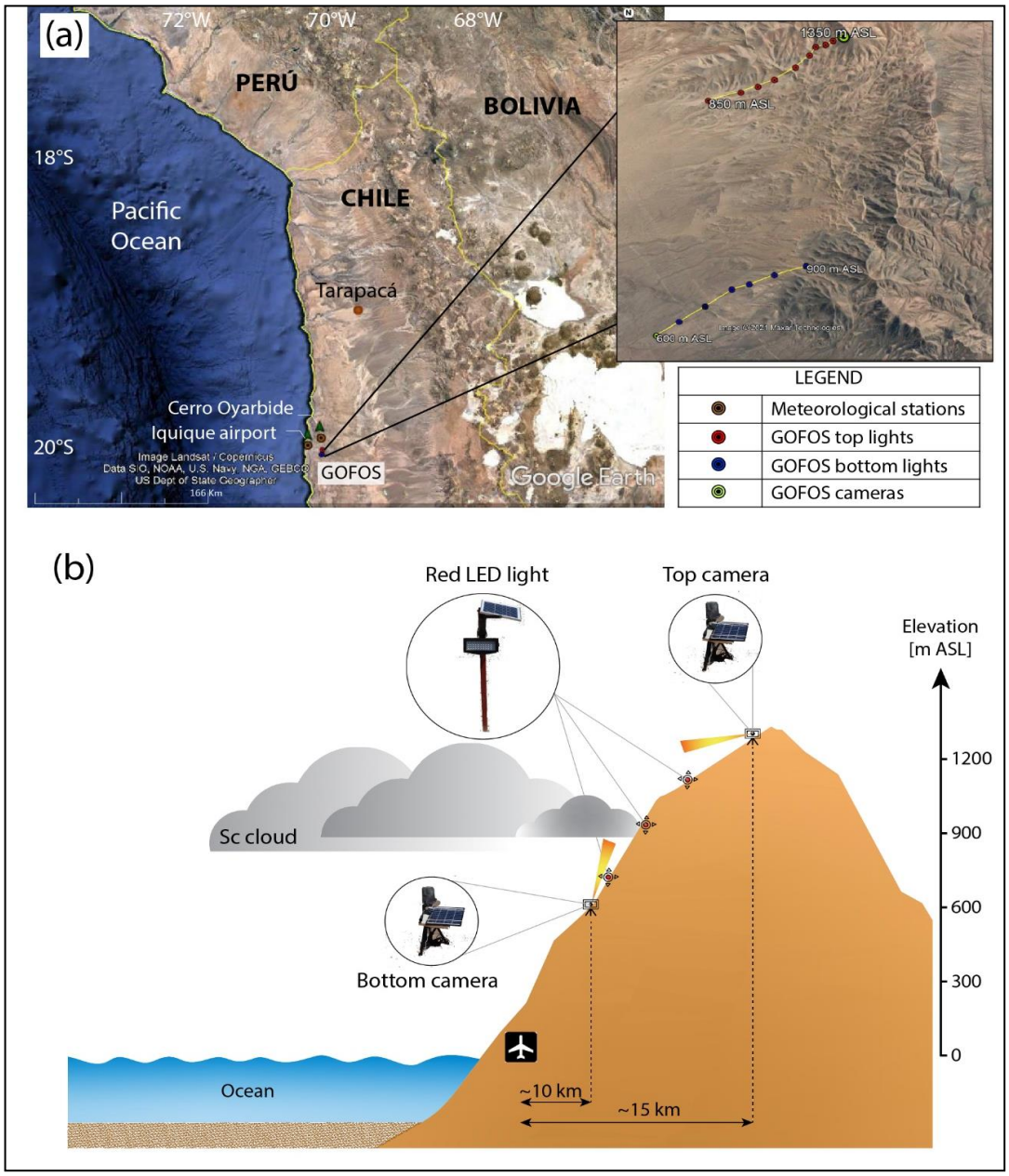


Figure 2. (a) Study site indicating how GOFOS is deployed in the coast of the Atacama Desert. The inset at the right-hand side shows the transects of GOFOS top and bottom. The GOFOS is installed near Cerro Oyarbide and the Iquique airport, where meteorological data are available. (b) Scheme of GOFOS with time lapse cameras, Red LED light and its solar panels. Approximate distances relative to the Iquique airport and elevations above sea level are also shown.

3.2. Meteorological Stations

Observations of relative humidity (*RH*), atmospheric pressure (*P*) and air temperature (*T*) from two nearby meteorological stations (Figure 2a) are available. These meteorological variables allowed to estimate θ and q using the atmospheric pressure of the lowest station as reference (Stull, 1988). The lowest station is at the Iquique airport (20.55° S, 70.17° W), less than 15 km from the GOFOS system, and belongs to the Meteorological Direction of Chile (DMC). It is located at 48 m ASL and has been recording hourly data since 1981. The highest meteorological station is located in *Cerro Oyarbide* (20.49°S, 70.06°W), less than 10 km from the GOFOS system. It is located at 1,211 m ASL and has been recording 10-min data since November 2016. It belongs to Centro UC Desierto de Atacama and Research Group for Earth Observation (RGEO), Department of Geography, Heidelberg University of Education, Heidelberg, Germany.

3.3. GOFOS

The GOFOS is a novel system proposed by del Río et al. (2021a). It was installed at the coast of the Atacama Desert (Figure 2a), at ~20 km from the shoreline to have in-situ measurements of the Sc cloud presence, CB, and CT. It is composed of two optical time-lapse cameras that capture images every 10 min, and a series of autonomous pole-lights that are in an altitudinal profile covering the entire range of regular fog presence (Figure 2b). The cameras were installed at 600 m ASL (20.63° S, 70.06° W) and at 1,350 m ASL (20.58° S, 70.03° W), on plastic tripods and were connected to 10-W solar panels. The lights consist of 44 LEDs in a red mode, installed on a 1.5 m galvanized iron pole and energized by a 6-W solar panel. The lights are spaced at a vertical distance of 50 m. As shown in the inset of Figure 2a, the bottom camera covers a first transect of six pole-lights installed between 600 and 900 m ASL and captured images from January to December 2017. The top camera covers a second transect of nine pole-lights located between 850 and 1,350 m ASL, in the E-W direction and captured images from August

2016 to December 2017. This analysis considered the period between January 19th and December 31st, 2017, in which both cameras were operative.

The GOFOS' dataset is a binary database obtained from images extracted from the recorded videos where (1) indicates fog presence, when the lights are not visible, and (0) indicates fog absence, when the lights are visible. The CB and CT heights were determined considering the location and height of pole-lights in the transect.

3.4. Data processing

3.4.1. Threshold estimation to classify the MBL regimes.

Following the methodology proposed by Lobos-Roco et al. (2018), first the hourly θ and q were calculated for each meteorological station using the pressure observed at 48 m ASL as a reference. From this calculation, the differences of θ and q over a vertical gradient from the reference level (48 m ASL, i.e., the elevation of the first meteorological station) to the highest level (1,211 m ASL, i.e., the elevation of the second meteorological station) were estimated. To do this, the following equations were used:

$$\theta = T(\frac{P_0}{P})^{\frac{R}{C_p}} \qquad \text{Eq.3}$$

where: θ is the Potential temperature, T (K) is absolute temperature, P_0 (Pa) is the reference pressure at 48 m ASL in this case, P (Pa) is the observed pressure at 1211 m ASL, and R/cp (0.286) is the Gas constant and specific heat capacity at a constant pressure.

$$q = rh_1 * q_s \qquad \text{Eq.4}$$

$$q_s = \frac{0.622 * e_s}{P/10} \qquad \text{Eq.5}$$

$$e_s = e_0 * e^{\frac{b*(t_k - t_1)}{t_k - t_2}} \qquad \text{Eq.6}$$

Where: q_s is the saturated mixing ratio, e_s is the saturation vapor pressure, e_0 (611 Pa) is the actual vapor pressure, rh_1= *rh*/100 is the relative humidity, *b* (17.2694) is a dimensionless parameter, t_k (^{o} C) is the air temperature at the meteorological station, t_1 (273.16 K) and t_2 (35.86 K) are temperature conversion factors.

Then, the hours with fog presence/absence were identified using the GOFOS' database, which was previously filtered. The filtering process consisted in choosing measurements where the inversion layer is located higher or equal to 1,211 m ASL to ensure that the meteorological measurements performed at this elevation are within the MBL. Also, the hours with incomplete GOFOS data were removed from the database. This filtering process resulted in a total of 5989 hours to analyze.

The hours used in the analysis are associated to 346 days, which were grouped by season to evaluate the temporal evolution of θ and q gradients throughout the year. The data consisted in 73 summer days (December to February), 91 autumn days (March to May), 92 winter days (June to August) and 90 spring days (September to November). Each season was divided in 150 steps to test different thresholds that allow to have a first classification of the MBL regimes (well-mixed or stratified). The results of this classification were compared with the GOFOS' dataset considering the following criterion: there is "agreement" if a) GOFOS indicates fog presence and the MBL regime is well-mixed, or if b) GOFOS indicates fog absence and the MBL regime is stratified. These agreements were expressed in percentage and the errors were calculated through their complement to sum up 100%. The best seasonal thresholds of daily $\partial\theta/\partial$z and $\partial q/\partial$z were defined as those that have the greater agreement and the smallest error for both regimes in each analyzed season.

To investigate the predictive power of daily $\partial\theta/\partial$z and $\partial q/\partial$z, two Logistic Regression Models (LRMs) were generated using the logit link function (Nielsen, 1998), and considering compliance with relevant assumptions, such as independence of errors, linearity in the logit for continuous variables, absence of multicollinearity, and lack of outliers and strongly influential values (Stoltzfus, 2011). The LRMs use the maximum

likelihood to find the highest probability that an event will occur based on predictors whenever the category to be estimated follows a binomial distribution (Hedeker, 2003). Thus, this regression model estimates the probability of a binary outcome category by transforming the linear regression equation of the independent variables in a logit scale (Stoltzfus, 2011). In this study, two LRMs were constructed: LRM1 (GOFOS-θ) and LRM2 (GOFOS- q); where fog presence and fog absence are the outcome categories, and θ and q hourly gradients are the independent variables used as single predictors in each model.

Before generating the LRMs, the data was randomly split in two sets for training/validation and for evaluation (Raykar and Saha, 2015; Korjus et al., 2016; Xu and Goodacre, 2018). The training and inner validation set comprised 4792 hours (80% of the data), whereas the evaluation set had 1197 hours (20% of the data). As the training dataset had 299 hours with fog events and 4493 days with fog absence (1:15 ratio), the data were balanced by using a weight of 15 in the fog events. This process avoids a biased classification in each LRM. Then, an internal validation was carried out applying the K-fold cross validation technique, which consists in randomly partition the dataset into K equal sized subsamples or folds where the single subsample is used to the validation data, and the remaining $k - 1$ subsamples are used as training data (Rodriguez et al., 2010). This process was repeated K = 5 times with each of the k subsamples, to allow a large fraction of data for the training process (See Figure 3). The selection of K = 5 equilibrates the tradeoff between bias and computational costs (Fushiki, 2011; Anguita et al., 2012).

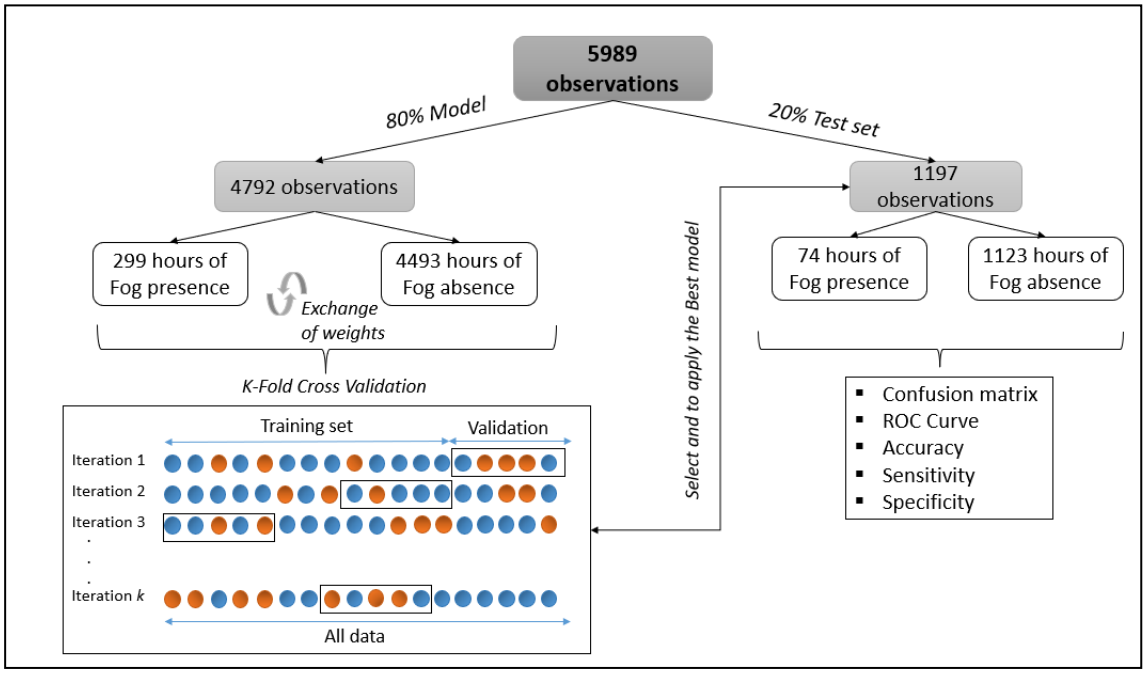


Figure 3. Scheme of the process to generate and validate the Logistic Regression models.

Model performance was evaluated using two global indicators: the confusion matrix and the receiver operating characteristic (ROC) curve (Giancristofaro and Salmaso, 2003). On one hand, the confusion matrix allows analyzing the proportion of a correct fog event classification performed by the model, with four possible outcomes: true positives (TP), which indicates that the model correctly forecasts a fog event; true negatives (TN) that are associated to correct model predictions of fog absence; false positives (FP), which occurs when the model forecasts a fog event that does not exist; and false negatives (FN) that happens when the model predict fog absence instead of an event. On the other hand, the ROC curve represents the trade-off between model's sensitivity (proportion of TP that are correctly identified, also known as true positive rate) and specificity (proportion of TN that are properly identified, also known as true negative rate). The area under the ROC curve ranges from 0 to 1. The larger the area under the ROC curve, the more accurate the model prediction is (Stoltzfus, 2011). These global indicators allow

obtaining model's accuracy, sensitivity, specificity, and precision (Giancristofaro and Salmaso, 2003).

3.4.2. Cloud Base estimation

The CB elevation was estimated only during days that showed well-mixed conditions or Sc cloud-fog formation (104 days) by applying two methods. The first one, was the lifting-air parcel method (Wetzel, 1990), which assumes that the LCL of a boundary layer cloud is similar to the CB height (Lobos-Roco et al., 2018), as shown in Figure 1a. This method uses thermodynamic variables, such as θ, q and the liquid water mixing ratio ($q_l \geq 0$) to estimate the elevation at which the air condensates (i.e., the LCL) that is where the saturated specific humidity (q_s) is equal to q.

To determine the LCL, the surface pressure at 48 m ASL was used as the reference in the calculations. Then, the changes of an air parcel lifting from there by using the values of P, θ and q at 48 m ASL, and of the station located at 1,211 m ASL, were estimated. The q_l was calculated following a simple criterion: $q_l = q_s - q$ (if $q_s \geq q$), which allows obtaining a first order estimate of q_l. This estimation was further improved by incorporating a mixing factor, m, which is the fraction of surface air contained in the parcel as it reaches the top of the MBL (Wetzel, 1990). It represents how the surrounding air mixes with the lifted-air parcel (Wetzel and Boone, 1995): m = 0 means that there is no air entrained and the parcel conserves its original θ from the surface to the top of the mixed layer; whereas m = 1 describes the condition when the air in the parcel is identical to that of its surroundings when it reaches the top of the MBL (Lobos-Roco et al., 2018). Hence, the parcel's θ and q at a height z for different mixing factors were calculated as (Lobos-Roco et al., 2018):

$$\theta^P(z) = \left(1 - m\frac{z}{z_{LCL}}\right)\theta^S + m\frac{z}{z_{LCL}}\theta^{MBL} \quad \text{Eq. 7}$$

$$q^P(z) = \left(1 - m\frac{z}{z_{LCL}}\right)q^S + m\frac{z}{z_{LCL}}q^{MBL} \quad \text{Eq. 8}$$

where $\theta^P(z)$ and $q^P(z)$ are the potential temperature and specific humidity of the air parcel that is at a height z, respectively; z_{LCL} is the height of the LCL; θ^{MBL} and q^{MBL} are the MBL's potential temperature and specific humidity (average between θ and q at 48 m ASL and 1,211 m ASL), respectively; and θ^S and q^S are the potential temperature and specific humidity at the surface (at 48 m ASL), respectively. The best mixing factor was determined by comparing the LCL results (CB calculated) with the CB observed by GOFOS (as described above). The error between predicted and observed values was quantified using the mean absolute error (MAE), and the root mean square error (RMSE).

The second method used artificial neural networks (ANNs) considering *RH*, *P* and *T* from the thermodynamic profile in the MBL as predictor variables. The ANNs were generated with the Neural Network toolbox of Matlab® using 80% (83 days) of the dataset for training/validation, whereas 20% (21 days) of the data was used for evaluation. To compare the results of CB through LCL and ANNs, the performance metrics of the entire data set are reported. The best ANN parameters were determined using the Levenberg-Marquardt backpropagation algorithm to minimize network error (Lourakis, 2005). The feedforward network structure included 10 hidden layers, which is the default value provided by Matlab®, as variations in the hidden layers did not significantly change the CB estimations. Finally, the MAE and the RMSE were calculated to compare the simulated values with the observations reported by GOFOS.

3.4.3. Cloud Top estimation

With the aim of predicting CT in a simple way using known parameters, linear regression models and ANNs were constructed to relate LCL, θ and q vertical gradients, percentage of daily hours when fog is present according to the MBL regimes classification, i.e., fog presence percentage (FPP), and meteorological variables (*RH*, *P*, *T*, wind speed) with CT.

For both methodologies, only the days under well-mixed conditions were analyzed (104 days), of which 80% (83 days) were used as training set/validation, and the remainder of the data was used for evaluation (21 days).

First, for the linear regression models, the input data were rescaled into a range between 0 and 1 with the aim of improving the interpretation of the predictor's influence on the response variable (Ali et al., 2014). Then, the correlation between the independent variables and the dependent variable (CT) was evaluated to choose the predictors with more incidence to estimate CT. With the selected predictors, different models were generated considering compliance with relevant assumptions of the linear regressions (Jacob, 2013), such as independence of errors, linearity for continuous variables, absence of multicollinearity, and homoscedasticity. These models were assessed using the Student's *t*-distribution with a significance level of 0.05. The r^2 and the Akaike information criterion (AIC) (Kabacoff, 2011) were used to determine the goodness of fit of these models. Finally, the best models were internally validated through the K-fold cross validation method (K = 5) and were tested with the 100% of the dataset, where the MAE and the RMSE between the simulated and observed (GOFOS) data were used as a regression diagnosis (Hastie et al., 2009; Jacob, 2013).

Second, CT was also estimated through ANNs considering *RH*, *P* and *T* from both meteorological stations as predictor variables and using the Levenberg-Marquardt backpropagation algorithm with 10 hidden layers to the feedforward neural network. To compare the results of ANNs with the linear regression models, the performance metrics of the entire data set are reported. Finally, the MAE and the RMSE were calculated to compare the simulated values with the observations reported by GOFOS.

IV. RESULTS

4.1. Seasonal thresholds of MBL regimes.

On one hand, the vertical gradients of θ (Figure 4a) and q (Figure 4b) measured during FLC presence are smaller than those under FLC absence. These results demonstrate that the small or large gradients can be related to the well-mixed (fog formation) or stratified (fog dissipation) regimes within the MBL, respectively (Lobos-Roco et al. 2018, García et al.,2021).

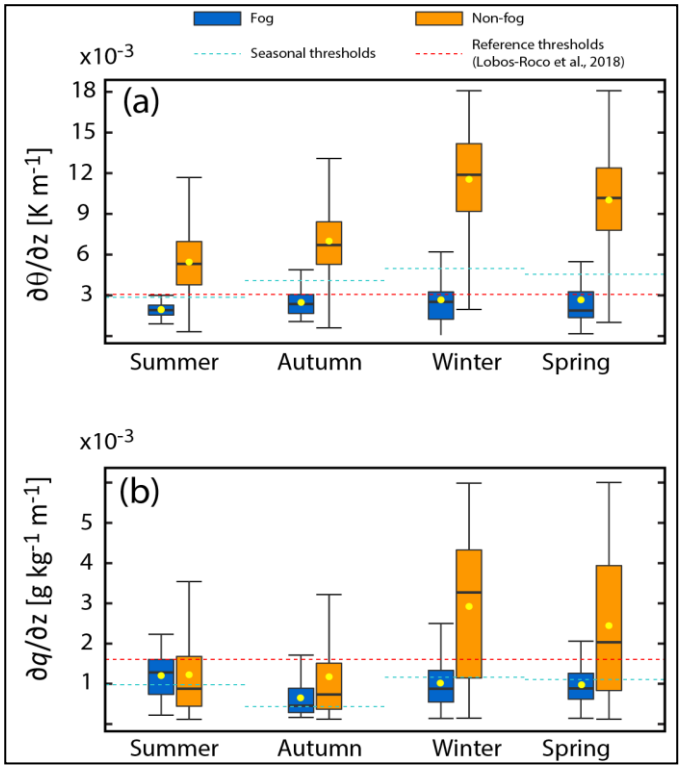


Figure 4. Descriptive statistics of (a) potential temperature vertical gradients ($\partial\theta/\partial z$) and (b) specific humidity vertical gradients ($\partial q/\partial z$) separated by fog and non-fog events. Each boxplot rectangle represents the interquartile range, the whiskers correspond to the minimum and maximum data values (excluding outliers), the horizontal line within the rectangle represents the median, and the yellow circles correspond to the average. The dashed-red line represents the reference thresholds proposed by Lobos- Roco et al. (2018) and the cyan line represents the seasonal thresholds found in this study.

On the other hand, under FLC conditions a small seasonal variability of θ and q gradients is observed (with differences of ~2.5 K and ~1 g/kg between the lowest and highest meteorological stations, respectively), whereas the opposite occurs under fogless conditions where the variability is larger and noticeable (~10 K and ~2 g/kg, in average respectively). These latter differences are larger during winter and spring compared to summer and autumn. This behavior is explained by the seasonal thermal variability of the free troposphere.

Figures 5 and 6 presents examples of the methodology utilized to estimate the thresholds of θ and q vertical gradients. These thresholds, which were calculated for the different seasons of the year, are also presented in Figure 4 and compared to those presented by Lobos-Roco et al. (2018). On one hand, the greatest $\partial\theta/\partial z$ thresholds occur in winter (4.99×10^{-3} K/m) and spring (4.58×10^{-3} K/m). These values exceed the Lobos-Roco et al. (2018) reference threshold (3.10×10^{-3} K/m) by ~61% and ~48%, respectively. On other hand, the lowest $\partial\theta/\partial z$ thresholds occur in summer (2.88×10^{-3} K/m) and autumn (4.13×10^{-3} K/m) being ~7% less and ~33% more than the reference threshold, respectively. Also, as shown in Figure 4b, the estimated $\partial q/\partial z$ thresholds in all the seasons, are smaller than the reference threshold (1.60×10^{-3} g/kg-m) proposed by Lobos-Roco et al. (2018). The $\partial q/\partial z$ thresholds for summer (9.62×10^{-4} g/kg-m) and autumn (4.19×10^{-4} g/kg-m) are ~40% and ~74% less than the reference threshold, respectively. While in winter (1.15×10^{-3} g/kg-m) and spring (1.09×10^{-3} g/kg-m), the thresholds are ~6% and ~32% less than the reference threshold, respectively. However, these differences are relevant only for summer and autumn where both fog events (well-mixed regime) and non-fog events (stratified regime) are below the reference threshold.

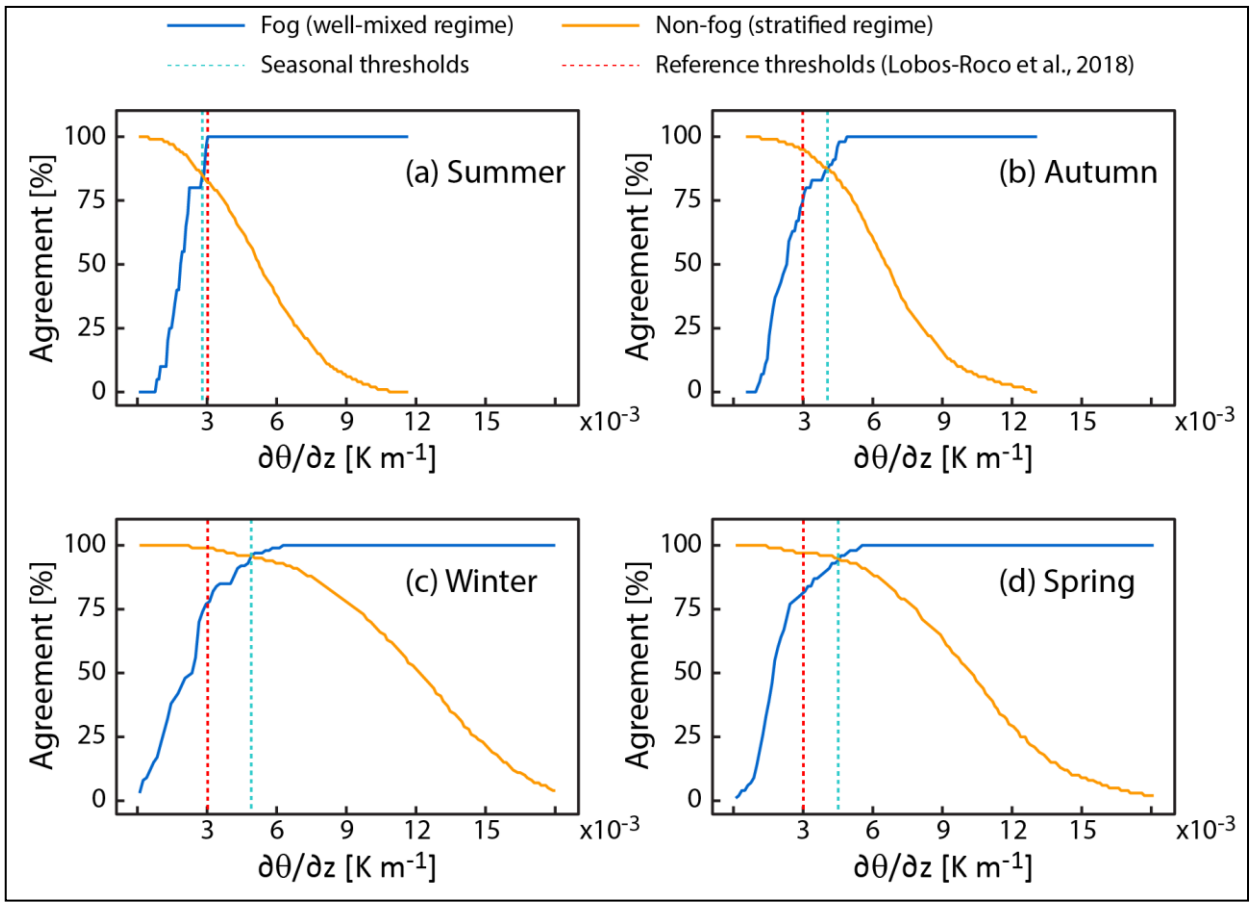


Figure 5. Percentage of agreement between the $\partial\theta/\partial z$ seasonal thresholds to classify the MBL regimes and GOFOS observations by season. The blue line represents the agreements between fog and the well-mixed MBL regime, whereas the orange line represents the agreement between non-fog and the stratified MBL regime. The dashed cyan line represents the seasonal threshold that maximizes the agreement of both fog and non-fog conditions, whereas the dashed red line represents the reference threshold proposed by Lobos-Roco et al. (2018).

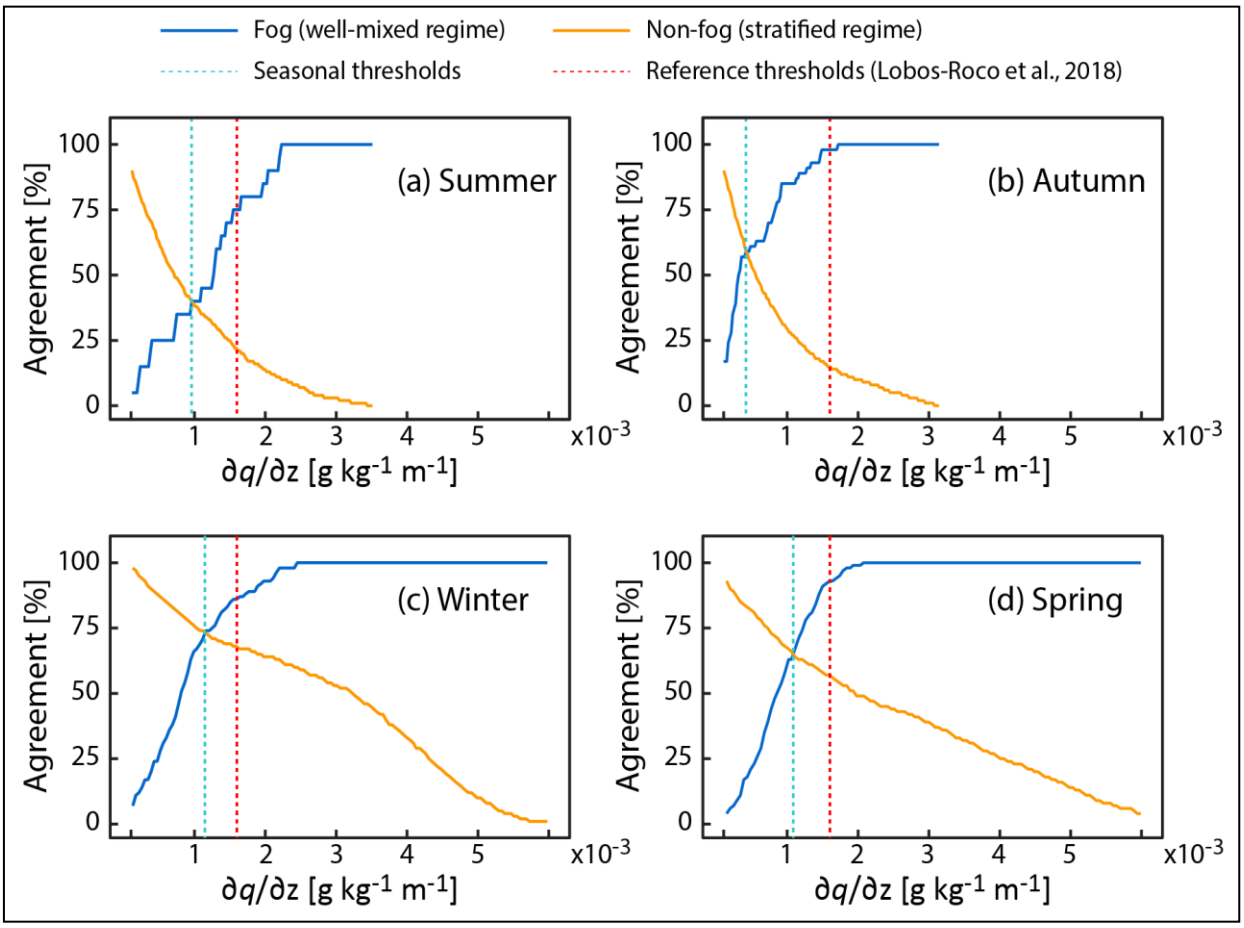


Figure 6. Percentage of agreement between the $\partial q/\partial z$ seasonal thresholds to classify the MBL regimes and GOFOS observations by season. The blue line represents the agreements between fog and the well-mixed MBL regime, whereas the orange line represents the agreement between non-fog and the stratified MBL regime. The dashed cyan line represents the seasonal threshold that maximizes the agreement of both fog and non-fog conditions, whereas the dashed red line represents the reference threshold proposed by Lobos-Roco et al. (2018).

Fog detection using the proposed $\partial\theta/\partial z$ seasonal thresholds, as shown in Figure 7a, agrees in ~95% of the time for winter and spring, and in ~86% of the time for summer and autumn. When using the $\partial\theta/\partial z$ reference threshold, fog detection agreement reaches ~79% for winter and spring, and ~87% for summer and autumn. Similarly, fog absence detection using either seasonal or reference $\partial\theta/\partial z$ thresholds result in good agreements

(~97% during winter/spring and ~87% for summer/autumn) when compared to the GOFOS observations (Figure 7b). Figures 7c and d also shows that the agreement percentage of fog detection when using the proposed $\partial q/\partial z$ seasonal thresholds is ~69% in winter and spring, and ~49% in summer and autumn. Note also that even when the $\partial q/\partial z$ threshold proposed by Lobos-Roco et al. (2018) in some cases achieves ~98% of agreement for the well-mixed regime, it has much lower agreements for the stratified regime compared to the proposed seasonal thresholds (see Figure 6).

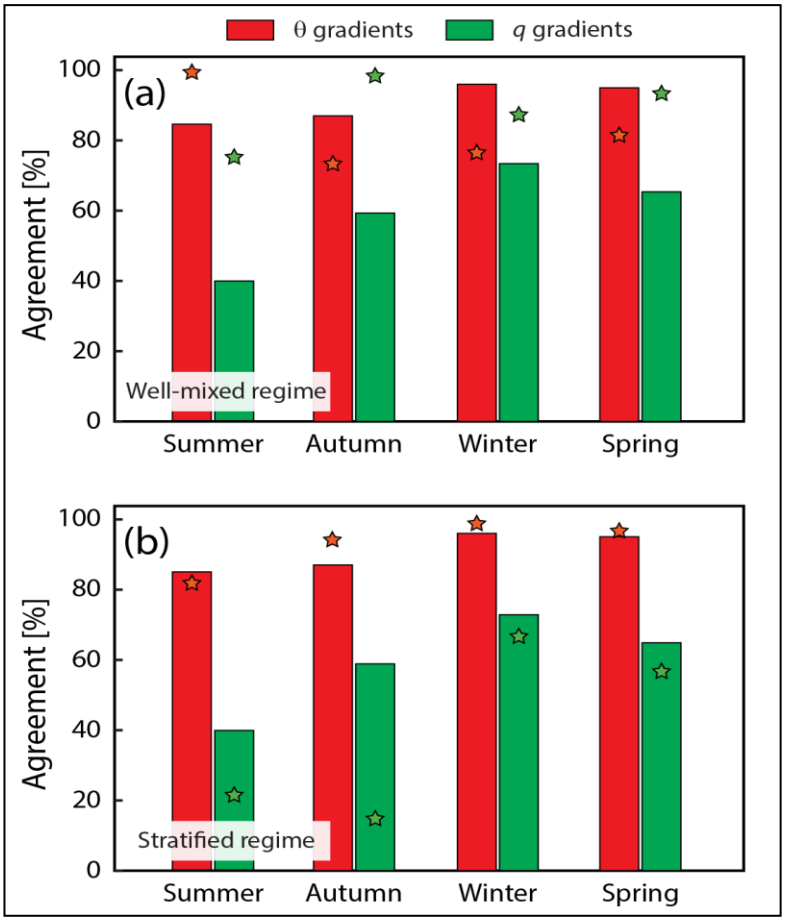


Figure 7. Percentage of agreement between θ and q vertical gradient thresholds to classify the MBL regimes and GOFOS observations for the different seasons. (a) Well-mixed regime; and (b) stratified regime. The orange and light green stars correspond to the percentage of agreement achieved with the reference thresholds proposed by Lobos-Roco et al. (2018).

As presented in Table 1, the LRMs generated between GOFOS-θ (LMR1) and GOFOS-q (LMR2) supports the previous findings. LRM1 shows a greater prediction power and accuracy than LRM2. For instance, LRM1 has a better quality and fit to the data because its AIC (398) is less than that of LRM2 (739). On one hand, the LMR1 prediction has 66 true positives (TPs), 206 false positives (FPs), 8 false negative (FN), and 917 true negatives (TNs). Hence, this logistic model correctly classifies 983 out of 1197 datapoints, reaching an accuracy of 0.82. On the other hand, LMR2 has 60 TPs, 589 FPs, 14 FNs, and 534 TNs. In this case, the logistic model correctly classifies 594 datapoints and has an accuracy of 0.50.

Table 1. Summary of the Logistic Regression Models (LRMs) results. LMR1 refers to the GOFOS-θ LRM, whereas LRM2 corresponds to the GOFOS-q LRM. For the Akaike information criterion (AIC), five independent variables were used (K = 5) and L is the maximum value of the likelihood function obtained for each model.

INDICATOR	FORMULA	LRM 1	LRM 2
Akaike information criterion (AIC)	$2K - 2\ln(L)$	398	739
True positives (TPs)	-	66	60
True negatives (TNs)	-	917	534
False Positives (FPs)	-	206	589
False Negatives (FNs)	-	8	14
Accuracy (Acc)	(TP+TN)/(TP+TN+FP+FN)	0.82	0.50
False Positive Rate (FPR)	FPR= FP/(FP+TN)	0.18	0.52
False Negative Rate (FNR)	FNR= FN/ (FN+TP)	0.11	0.14
Sensitivity or true positive rate (TPR)	TPR = 1-FNR	0.89	0.86
Specificity or true negative rate (TNR)	TNR= 1-FPR	0.82	0.48

The ROC curve for each model is shown in Figure 8. On one hand, there is a very good relationship between sensitivity and specificity in LRM1, because the area under the ROC curve indicates that there is a probability of ~91% that the model will distinguish between TPs and FPs. On the other hand, LRM2 has a probability of 65% to distinguish between FNs and TNs, but it has a poor prediction of TPs and FPs.

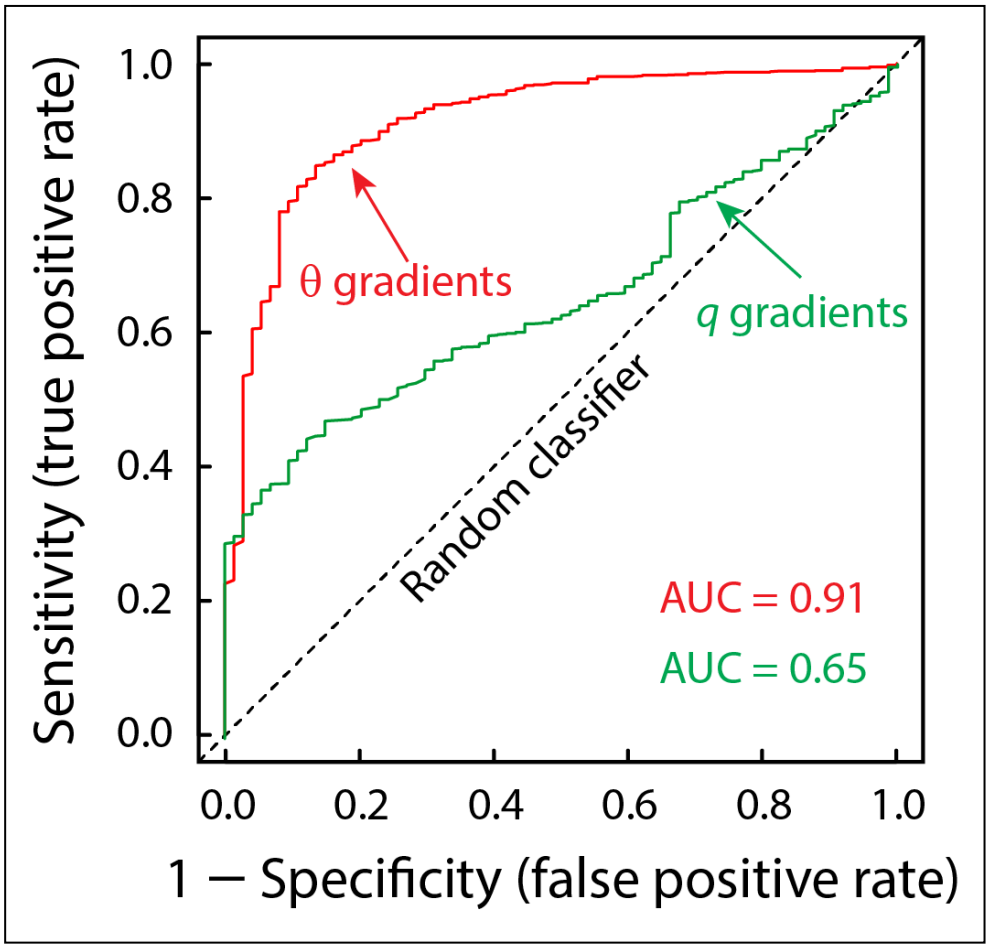


Figure 8. Receiver operating characteristic (ROC) curve of q (red line) and q (green line) vertical gradients. AUC indicates the area under the curve of each ROC curve.

4.2. Cloud base

The best agreement between CB observed by GOFOS and CB estimated by the LCL method occurs when the mixing factor is m = 1.0 (see Figure 9).

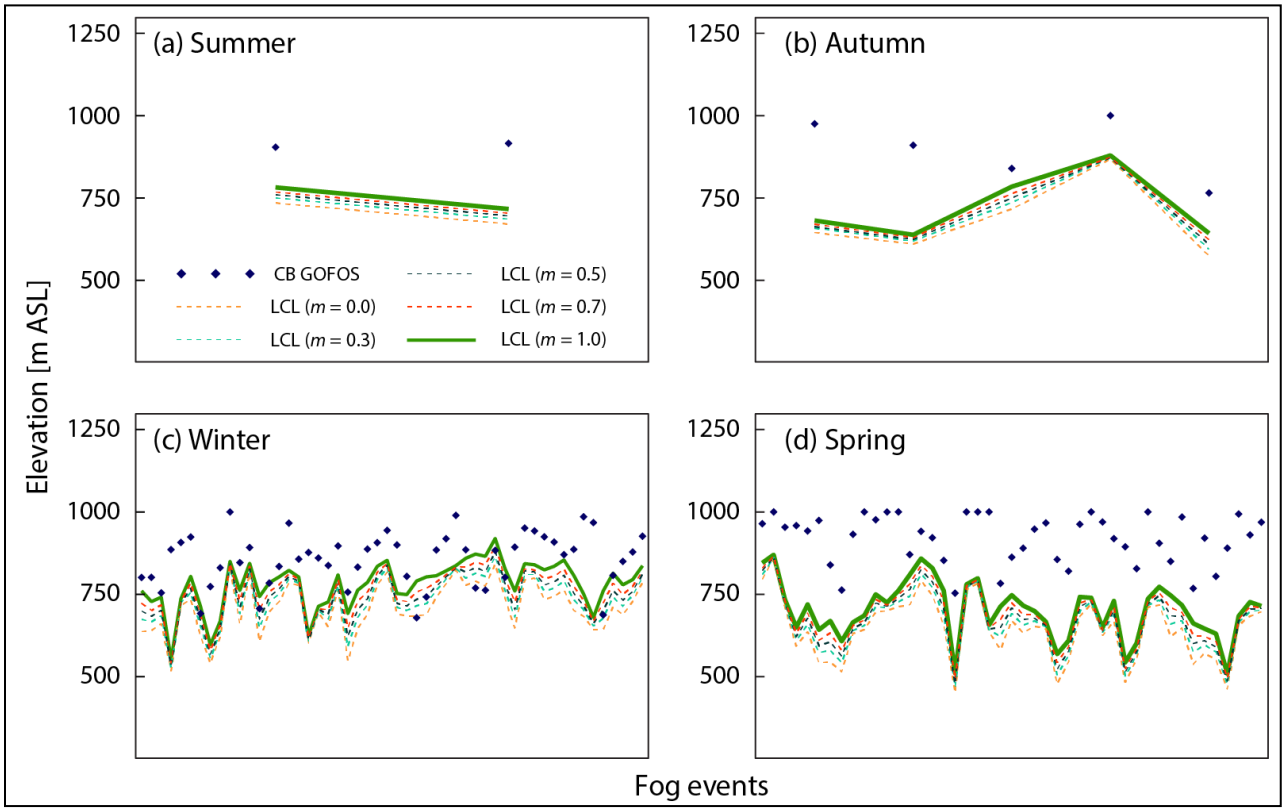


Figure 9. Seasonal variability of the Sc cloud base (CB) observed by GOFOS (CB GOFOS) and that estimated with the lifting condensation level (LCL) using different mixing factors (*m*).

The CB seasonal variability reported by GOFOS, LCL and ANNs during the year under investigation is presented in Figures 10 and 11a. The lowest average GOFOS-estimated CB elevation is observed in autumn and winter (898 and 857 m ASL, respectively), while the highest average GOFOS-estimated CB altitudes occur in spring and summer (919 and 910 m ASL, respectively). In general, the LCLs presented in Figures 10 and 11a show lower altitudes in autumn and spring (725 and 702 m ASL, respectively), and higher altitudes in summer and winter (749 and 781 m ASL). When comparing the GOFOS-estimated CB with the LCL, an average annual difference of 154 ± 95 m is obtained, and the smallest error occurs in winter with a difference of 98 ± 69 m. These

differences represent less than 17% of the annual average of the GOFOS-estimated CB and are similar to those obtained by Lobos-Roco et al. (2018), which obtained differences of ~110 m when they applied the LCL method to estimate CB. In both cases the lifting-air parcel method underestimates the CB values, especially in summer and spring where errors of 173 ± 104 m and 217 ± 84 m, respectively, were found.

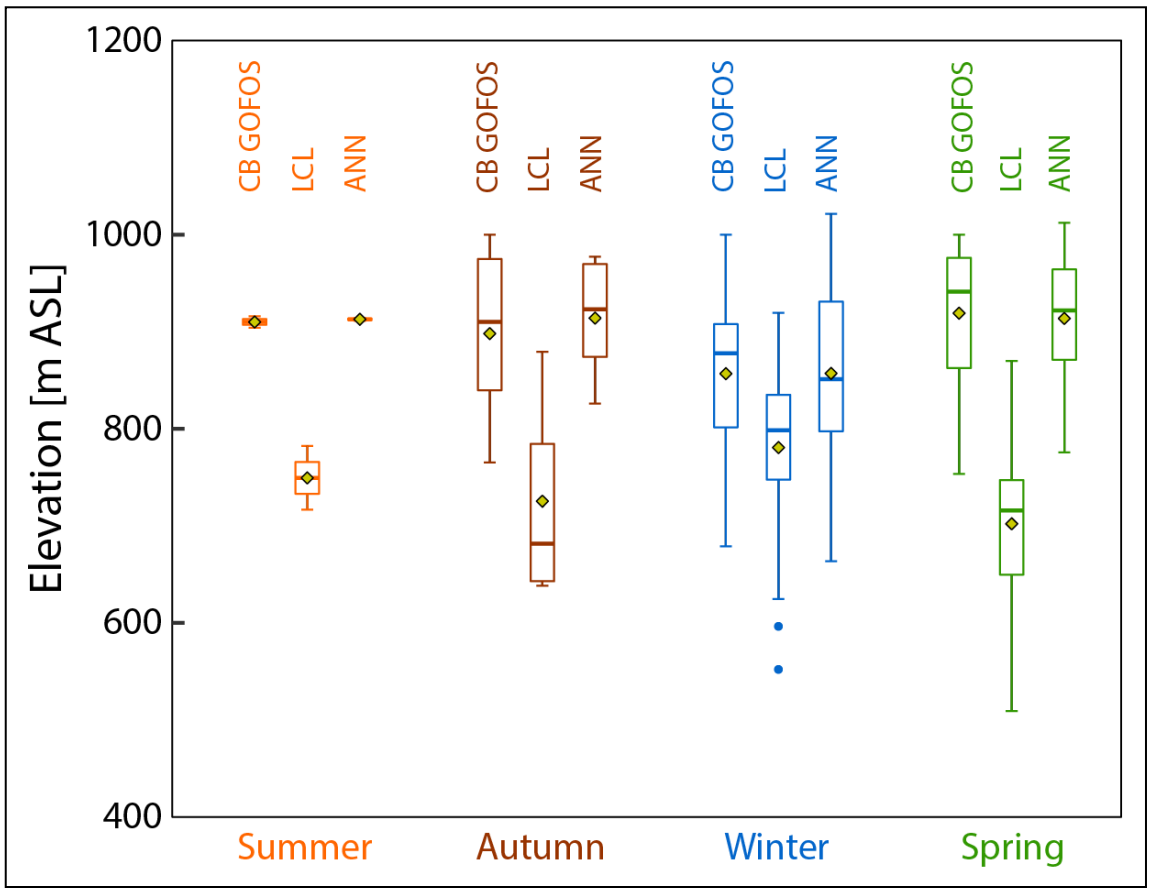


Figure 10. Seasonal boxplots of GOFOS-estimated cloud base (CB GOFOS), lifting condensation level (LCL), which is the approximation of the CB calculated by the lifting-air parcel method with a mixing factor of m = 1.0, and CB estimated using artificial neural networks (ANNs).

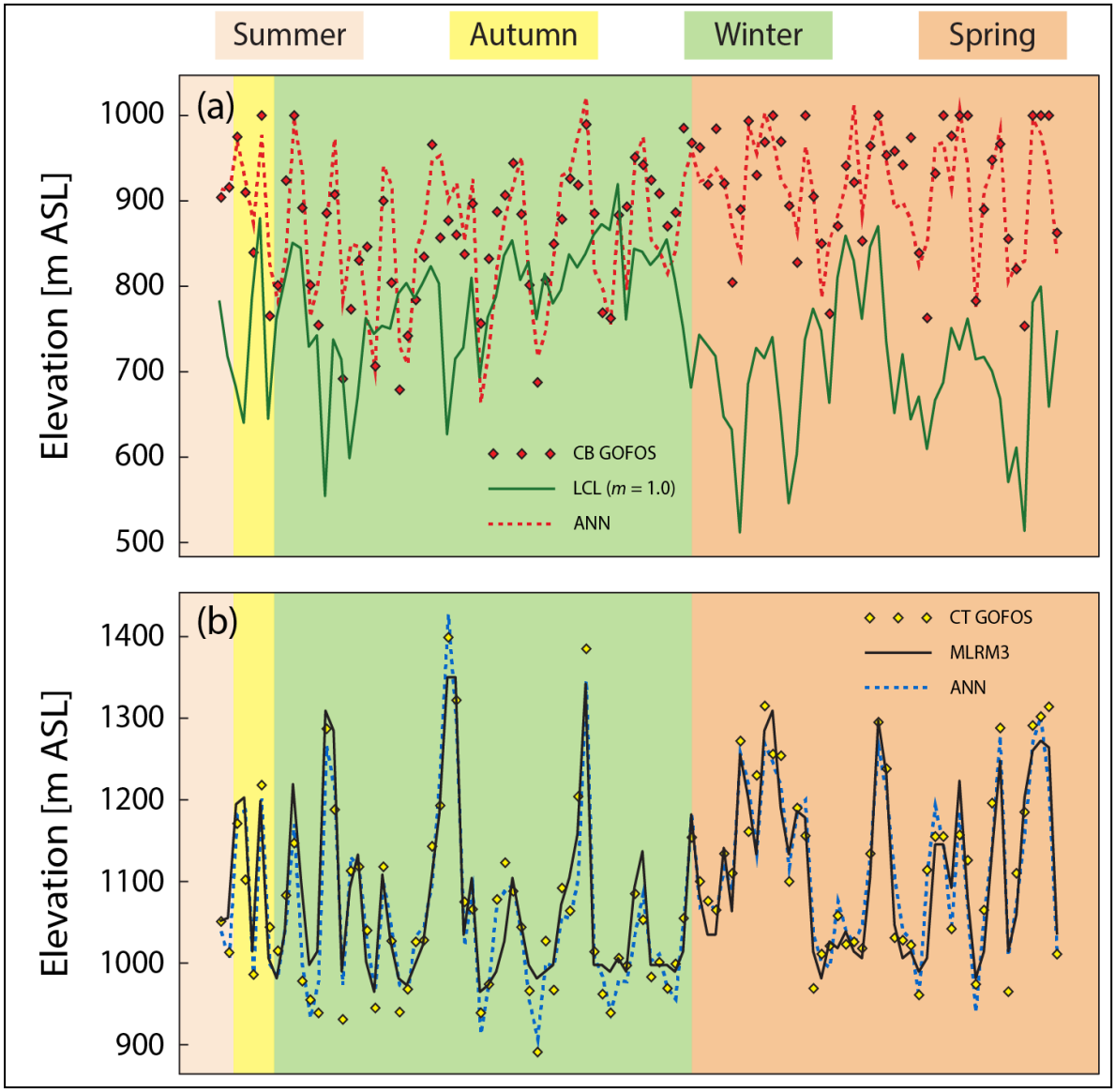


Figure 11. Seasonal cloud base (CB) and cloud top (CT) elevations: (a) CB observed by GOFOS (CB GOFOS) and predicted using the lifting condensation level (LCL) and artificial neural networks (ANNs). (b) CT observed by GOFOS (CT GOFOS) and predicted using the multiple linear regression model (MLRM) and ANNs.

Artificial Neural Networks show better estimations of GOFOS-estimated CB than LCL throughout the year (Figures 10 and 11a) with an average annual difference of 39 ± 29 m, which represent less than 4% of the observed annual mean CB. Moreover, Figure 10 shows that the probability distribution of CB estimated by ANNs resembles that of the GOFOS-estimated CB. The best performance of the ANNs occurs in summer and autumn, with differences of less than 3% compared to the observed CB elevations in these seasons. The lowest errors occurred in summer with a difference of 7 ± 4 m, whereas the highest errors occurred in winter with a difference of 45 ± 30 m.

4.3. Cloud Top

Table 2 shows the best regression models that were found to predict CT. CT showed a good correlation ($r > 0.7$) with LCL, FPP, θ and q gradients and a poor correlation with wind speed ($r = 0.4$). The best simple linear regression model (SLRM) only used one predictor, i.e., $\partial\theta/\partial z$, to estimate CT, whereas the best multiple linear regression models (MLRMs) used FPP, LCL and θ gradients as predictors. In MLRM1 and MLRM3, FPP and $\partial\theta/\partial z$ have the greatest influence on the response variable, while the effect of LCL on CT is small. Something similar happens in MLRM2 where among the two predictors (FPP and LCL), only FPP has a significant effect on CT estimation. According to the AIC and the R^2, MLRM3 represents better the evolution of CT. The mean value ± standard deviation of CT estimated with MLRM3 is of 1,090 ± 105 m ASL, which is an excellent estimation of the observed CT heights estimated with GOFOS, i.e., CT = 1,090 ± 113 m ASL ($r^2 = 0.84$; AIC = 87; $Pr(>|t| = 2 \times 10^{-16}$). As shown in Figure 11b, the MLRM3 and observed CT elevations follow the same trend, with MAE and RMSE of 35 and 43 m, respectively (less than 4% of the observed annual mean CT). The best performance of the MLRM 3 occurs in winter, spring, and summer with differences of 3% of the CT elevations observed in these seasons. The lowest errors occurred in summer with MAE and RMSE of 21 and 29 m, respectively, whereas the highest errors occurred in autumn, with MAE and RMSE of 42 and 51 m, respectively.

Table 2. Simple linear regression model (SLRM) and the multiple linear regression models (MLRMs) used to estimate cloud top (CT). AIC is the Akaike information criterion, LCL is the estimated cloud base, $\partial\theta/\partial z$ is the potential temperature gradient, and FPP is the percentage of daily hours when fog is present, estimated using the MBL regimes classification.

MODEL	STANDARDIZED EQUATION	SAMPLE STATISTICS	AIC	R^2
SLRM	$Y = 8.3 * 10^{-1}a - 2.9 * 10^{-16}$ Where: Y= $1/CT^2$; a= $log(\partial\theta/\partial z)$ CT is in m; $(\partial\theta/\partial z)$ is in K/m.	Mean (Y) = $8.7 * 10^{-7}$ Sd (Y) = $1.7 * 10^{-7}$ Mean (a) = -5.1 Sd (a)= 0.6	142	0.70
MLRM1	$CT = -2.2 * 10^{-2}LCL + 7 * 10^{-1}FPP - 2.5 * 10^{-1}(\partial\theta/\partial z) + 7.8 * 10^{-16}$	Mean (CT)= 1,090.5 m Sd (CT) = 114.6 m Mean (LCL) = 745.2 m Sd (LCL)=90.8 m Mean (FPP) = $2.3 * 10^{-1}$ Sd (FPP)= $2.8 * 10^{-1}$ Mean $(\partial\theta/\partial z)$ = $6.9 * 10^{-3}$K/m Sd $(\partial\theta/\partial z)$= $3.4 * 10^{-3}$ K/m	88	0.84
MLRM2	$CT = -1.2 * 10^{-2}LCL + 9.1 * 10^{-1}FPP + 8 * 10^{-16}$	Mean (CT)= 1,090.5 m Sd (CT) = 114.6 m Mean (LCL) = 745.2 m Sd (LCL)=90.8 m Mean (FPP) = $2.3 * 10^{-1}$ Sd (FPP)= $2.8 * 10^{-1}$	95	0.83
MLRM3	$CT = 7.1 * 10^{-1}FPP - 2.4 * 10^{-1}(\partial\theta/\partial z) + 7.9 * 10^{-16}$	Mean (CT)= 1,090.5 m Sd (CT) = 114.6 m Mean (FPP) = $2.3 * 10^{-1}$ Sd (FPP)= $2.8 * 10^{-1}$ Mean $(\partial\theta/\partial z)$ = $6.9 * 10^{-3}$K/m Sd $(\partial\theta/\partial z)$= $3.4 * 10^{-3}$ K/m	87	0.84

The ANNs used six parameters to predict CT (*RH*, *P* and *T* at 48 and 1,211 m ASL, respectively). As depicted in Figure 11b, CT estimations using ANNs were slightly better than those performed with MLRM 3, with r^2 = 0.92 and a CT of 1,086 ± 107 m ASL (a difference of less than ~0.5% with the GOFOS-estimated CT values). ANNs allow predictions of CT elevations with small errors (MAE = 25 m and RMSE = 33 m,

which represent less than 3% the observed annual mean CT). As opposed to MLRM CT estimations, ANNs yields excellent results during summer, winter, and spring, with differences smaller than 2% with respect to the observed CT, whereas in autumn this difference is smaller than 3%.

V. DISCUSSION

The $\partial\theta/\partial z$ and $\partial q/\partial z$ have a seasonal variability that can be related to air temperature and relative humidity. These meteorological variables change throughout the day and the seasons, influencing the available energy for both evaporation and dry-air entrainment into the MBL (Raupach, 2000). For instance, in summer, it is common to have the highest air temperatures and the lowest relative humidity (Houston, 2006). These characteristics do not promote condensation, decreasing $\partial q/\partial z$ and increasing $\partial\theta/\partial z$, because when the atmosphere warms the excess of available energy is released as sensible heat (van Heerwaarden et al., 2009). The vertical gradients variability can also be associated with the intensity of air subsidence that changes throughout the year, and influences fog formation and dissipation: in warmer seasons there are fewer fog events because air subsidence is weak and higher temperatures diminish the thermal inversion layer. Hence, air entrainment is stronger and dissipates fog (Stevens et al., 2003; Wood, 2012). Despite the variability of $\partial\theta/\partial z$ and $\partial q/\partial z$, and of the seasonal thresholds obtained in this work, the constant thresholds found by Lobos-Roco et al. (2018) and improved by del Río et al., (2021a), are robust limits of the MBL regimes, which can be used to detect fog formation/dissipation. Nonetheless, the methodology proposed in this work has a better performance to classify both MBL regimes (well-mixed and stratified) during all the seasons because the proposed thresholds were selected accounting for the best agreement and the smallest error for both regimes in each analyzed period.

We found that $\partial\theta/\partial z$ is a better predictor of fog formation than $\partial q/\partial z$. This result was expected because $\partial\theta/\partial z$ is less sensible than $\partial q/\partial z$ to physical processes, such as evaporation, which modulate the moist in the MBL (van Heerwaarden et al., 2010). The sensitivity of q also occurs because the contribution of heat entrainment to the total temperature tendency is smaller than the contribution of dry-air entrainment to the tendency of specific humidity (van Heerwaarden et al., 2009). Consequently, q variations have a strong impact on the MBL evolution in the presence of low

temperatures, whereas the impact of q variations is less pronounced in the presence of high temperatures. As a result, the MBL evolution is mostly determined by temperature variations (van Heerwaarden et al., 2009; van Heerwaarden et al., 2010; Wood, 2012).

Fog detection using the MBL regimes agrees well with GOFOS observations and thus, the proposed methodology for determining vertical Sc cloud variability is reliable. However, GOFOS is a system that presents some uncertainties, such as its spatial resolution (that depends on the number of red LED lights) and installation range, which significantly influence the observations through topography and slope (del Río et al., 2021a). The GOFOS' spatial resolution can be improved and adapted by incorporating more sensors that allows to measure meteorological variables at different altitudes within the MBL, to achieve more representative observations of the vertical profile of this layer.

Regarding the CB estimations, as expected, more fog events appeared when radiation decreases the surface heat fluxes and increases the moisture sources, allowing condensation at lower levels (Lobos-Roco et al., 2018). In our study, these conditions occurred between winter and spring. The best estimations of the CB elevation, when using the LCL as a surrogate for CB, were achieved with a mixing factor of m = 1.0, which means that the air in the parcel is identical to that of its surroundings when it reaches the top of the MBL (Lobos-Roco et al., 2018). The underestimation of CB using the lifting-air parcel method can be explained by the location of the meteorological stations (48 m ASL and 1,211 m ASL), as they may be not representative of the entire thermodynamic profile in the MBL used in the LCL calculation. More accurate LCL values can be obtained installing a meteorological station nearer to the coastline at an elevation in between 48 and 1,211 m ASL. For instance, García et al. (2021) used a meteorological station at 750 m ASL that was nearer to the coastline. In that meteorological station, they found that fog events were more common than those observed at 1,211 m ASL, because as relative humidity values are greater, the marine conditions are more uniform, and the exchange between marine and continental air is lower than their meteorological stations located at higher altitudes. Moreover, García et

al. (2021) used a station at 750 m ASL in the middle of our transect. They found that fog presence and fog water yield in this location is greater than that registered at 1,211 m ASL. So, there is a negative relationship between fog presence to both the distance to the coast and elevation of the meteorological stations. Unfortunately, the meteorological station located at 750 m ASL used by García et al. (2021) only had two months of data in our investigation period and thus, it was not considered in the analysis.

ANNs allow estimating CB using *RH*, *P* and *T* directly from two meteorological stations located within the MBL with good accuracy. This method offers a low-cost alternative with simple parameters to estimate Sc CB elevations. However, ANNs do not consider the physical processes that determine the CB height, which are explained better by the LCL. Unfortunately, it is still unknown how the surface fluxes and the topography affect the LCL estimation in the coast of the Atacama Desert. Thus, this research area must be further explored in the future.

Even when CT heights correlated well with LCL, FPP, and $\partial\theta/\partial z$ and $\partial q/\partial z$, it showed a poor correlation with wind speed (data not shown). This poor correlation may occur because wind speed within the MBL has a stronger influence on the sub-cloud layer than the cloud layer. In the sub-cloud layer, the surface turbulent fluxes increase with higher wind speeds, whereas in the cloud layer is common that radiative cooling drives entrainment generation at the CT (Man-Kui Wai, 1987; Rieck et al., 2012). The observed seasonal changes in CT elevations can be associated to the location of the Southeast Pacific Anticyclone (SEPA) and the air mass subsidence that influence fog presence (Myers and Norris, 2013; del Río et al., 2021a). This air mass causes the thermal inversion layer to occur at lower altitudes and, at the same time, induces cooler air temperatures than usual. Consequently, lower CT elevations are observed in winter due to changes in the SEPA orientation (towards north). This phenomenon increases subsidence and limits the MBL growth.

The MLRM 3 estimated well the Sc CT during the year in a simple way only using two predictors (FPP and $\partial\theta/\partial z$), whereas ANNs needed six traditional meteorological parameters. However, ANNs offer the possibility to estimate CT using *RH*, *P* and *T* that

are variables that can be obtained directly from the meteorological stations in the thermodynamic profile in the MBL; whereas the MLRM 3 requires additional processing to determine FFP and $\partial\theta/\partial z$. Hence, we conclude that ANN models are a low-cost alternative to estimate Sc CT with good accuracy. As a future perspective, additional independent variables should be investigated to improve the Sc CT prediction. We suggest exploring the vertical wind speed profile at a higher spatial resolution than that used in this work, especially near CT, where these variables can enhance entrainment.

VI. CONCLUSIONS AND OUTLOOKS

Fog variability and the thermodynamic Sc cloud vertical structure in the coastal Atacama Desert were characterized using novel ground observations and a thermodynamic characterization of the MBL.

On one hand, the $\partial\theta/\partial z$ seasonal thresholds found in this study, or the reference threshold proposed by Lobos-Roco et al. (2018) can be successfully used to determine fog formation/dissipation. On the other hand, even when the $\partial q/\partial z$ reference threshold proposed by Lobos-Roco et al. (2018) has a good performance to detect fog presence/absence, it can lead to errors in the identification of the stratified regimes, especially in summer and autumn. For these seasons, the $\partial q/\partial z$ seasonal thresholds proposed in this work perform better to identify the MBL regimes.

Fog presence prediction using LRMs generated by GOFOS data and vertical gradients of θ and q was successful, with fog event detection up to 82%. A comparison between these models revealed that when using $\partial\theta/\partial z$, there is a probability of ~91% to distinguish between fog presence and absence, compared to 65% when using $\partial q/\partial z$. Hence, fog detection using the $\partial\theta/\partial z$ threshold is a better predictor than that based on $\partial q/\partial z$.

The lifting-air parcel method underestimates CB elevations with a systematic mean error of 154 m during the year. To obtain more realistic values of CB with this method, a meteorological station installed between 48 and 1,211 m ASL should be used to improve the estimations of θ and q throughout the MBL. Moreover, ANN models are a good-accuracy alternative to estimate Sc CB through simple meteorological parameters, although they do not have a physical basis.

Finally, MLRM 3 and ANNs were used to predict CT. Both models represented well the temporal evolution of the GOFOS-based CT estimates during the year, but we propose the application of ANNs since it is a low-cost alternative that only needs to meteorological observations (*RH*, *P* and *T*) from the thermodynamic profile in the MBL.

Our findings reveal that the essential features of fog formation analyzed using one year of data can be approximated through standard meteorological observations, which contributes to the prediction of this phenomenon. It is suggested exploring the vertical wind speed profile at a higher spatial resolution than that used in this work, especially near CT, where these variables can enhance entrainment.

REFERENCES

Ali, P. J. M., Faraj, R. H. (2014). Data normalization and standardization: a technical report. Mach Learn Tech Rep, 1(1), 1-6. https://doi.org/10.13140/RG.2.2.28948.04489

Andersen, H., Cermak, J. (2018). First fully diurnal fog and low cloud satellite detection reveals life cycle in the Namib. Atmospheric Measurement Techniques, 11(10), 5461-5470. https://doi.org/10.5194/amt-11-5461-2018

Anguita, D., Ghelardoni, L., Ghio, A., Oneto, L., Ridella, S. (2012). The 'K'in K-fold cross validation. In 20th European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning (ESANN) (pp. 441-446). i6doc. com publ.

Bendix, J., Thies, B., Cermak, J. (2004). Fog Detection with Terra-MODIS and MSG-SEVIRI. In Proceedings of the 2003 Meteorological Satellite Conference (pp. 427-435).
Bendix, J., Thies, B., Cermak, J., Nauß, T. (2005). Ground fog detection from space based on MODIS daytime data—A feasibility study. Weather and Forecasting, 20(6), 989-1005. https://doi.org/10.1175/WAF886.1

Böhm, C., Reyers, M., Schween, J. H.,Crewell, S. (2020). Water vapor variability in the Atacama Desert during the 20th century. Global and Planetary Change, 190, 103192. https://doi.org/10.1016/j.gloplacha.2020.103192

Bretherton, C. S., Park, S. (2009). A new moist turbulence parameterization in the Community Atmosphere Model. Journal of Climate, 22(12), 3422-3448. https://doi.org/10.1175/2008JCLI2556.1

Cereceda, P., Osses, P., Larrain, H., Farıas, M., Lagos, M., Pinto, R., Schemenauer, R. S. (2002). Advective, orographic and radiation fog in the Tarapacá region, Chile. Atmospheric Research, 64(1-4), 261-271. https://doi.org/10.1016/S0169-8095(02)00097-2
Cereceda, P., Larrain, H., Osses, P., Farías, M., Egaña, I. (2008a). The spatial and temporal variability of fog and its relation to fog oases in the Atacama Desert, Chile. Atmospheric Research, 87(3-4), 312-323. https://doi.org/10.1016/j.atmosres.2007.11.012

Cereceda, P., Larrain, H., Osses, P., Farías, M., Egaña, I. (2008b). The climate of the coast and fog zone in the Tarapacá Region, Atacama Desert, Chile. Atmospheric Research, 87(3-4), 301-311. https://doi.org/10.1016/j.atmosres.2007.11.011

Chang, E. K. (1995). The influence of Hadley circulation intensity changes on extratropical climate in an idealized model. Journal of Atmospheric Sciences, 52(11), 2006-2024. https://doi.org/10.1175/1520-0469(1995)052<2006:TIOHCI>2.0.CO;2

Daidzic, N. E. (2019). A new model for lifting condensation levels estimation. International Journal of Aviation, Aeronautics, and Aerospace, 6(5), 1. https://doi.org/10.15394/ijaaa.2019.1341

del Río, C., Garcia, J.-L., Osses, P., Zanetta, N., Lambert, F., Rivera, D., Siegmund, A., Wolf, N., Cereceda, P., Larraín H., Lobos, F. (2018). ENSO influence on coastal fog-water yield in the Atacama Desert, Chile. Aerosol and Air Quality Research, 18 (1), 127–144. https://doi.org/10.4209/aaqr.2017.01.0022

del Río, C., Lobos, F., Siegmund, A., Tejos, C., Osses, P., Huaman, Z., Meneses, JP., García, J. L. (2021a). GOFOS, ground optical fog observation system for monitoring the vertical stratocumulus-fog cloud distribution in the coast of the Atacama Desert, Chile. Journal of Hydrology, 597, 126190. https://doi.org/10.1016/j.jhydrol.2021.126190

del Río C, Lobos-Roco F, Siegmund A, Latorre C, Koch MA, Garcia JL,Osses P, Lambert F, Alfaro F, Siegmund A (2021b) Spatial distribution and interannual variability of coastal fog and low clouds cover in the hyperarid Atacama Desert and implications for past and present Tillandsia landbeckii ecosystems. Plant Systematics and Evolution, 307(5), 1-23. https://doi.org/10.1007/s00606-021-01782-z

Domen, J. K., Stringfellow, W. T., Camarillo, M. K., Gulati, S. (2014). Fog water as an alternative and sustainable water resource. Clean Technologies and Environmental Policy, 16(2), 235-249. https://doi.org/10.1007/s10098-013-0645-z

Duynkerke, P. G., Zhang, H. Q., Jonker, P. J. (1995). Microphysical and turbulent structure of nocturnal stratocumulus as observed during ASTEX. Journal of the atmospheric sciences, 52(16), 2763-2777. https://doi.org/10.1175/1520-0469(1995)052<2763:MATSON>2.0.CO;2

Ellrod, G. P., Gultepe, I. (2007). Inferring low cloud base heights at night for aviation using satellite infrared and surface temperature data. In Fog and Boundary Layer Clouds: Fog Visibility and Forecasting (pp. 1193-1205). Birkhäuser Basel. https://doi.org/10.1007/978-3-7643-8419-7_6

Farías, M., Cereceda, P., Osses, P., Núñez, R. (2005). Comportamiento espacio-temporal de la nube estratocúmulo, productora de niebla en la costa del desierto de Atacama (21 lat. S., 70 long. W.), durante un mes de invierno y otro de verano. Investigaciones geográficas, (56), 43-61.

Fessehaye, M., Abdul-Wahab, S. A., Savage, M. J., Kohler, T., Gherezghiher, T., Hurni, H. (2014). Fog-water collection for community use. Renewable and Sustainable Energy Reviews, 29, 52-62. https://doi.org/10.1016/j.rser.2013.08.063

Forsythe, J. M., Vonder Haar, T. H., Reinke, D. L. (2000). Cloud-base height estimates using a combination of meteorological satellite imagery and surface reports. Journal of Applied Meteorology, 39(12), 2336-2347. https://doi.org/10.1175/1520-0450(2000)039<2336:CBHEUA>2.0.CO;2

Fushiki, T. (2011). Estimation of prediction error by using K-fold cross-validation. Statistics and Computing, 21(2), 137-146. https://doi.org/10.1007/s11222-009-9153-8

García, J. L., Lobos-Roco, F., Schween, J. H., del Río, C., Osses, P., Vives, R., Pezoa,M., Siegmund, A., Latorre, C., Koch, M., Loehnert, U. (2021). Climate and coastal low-cloud dynamic in the hyperarid Atacama fog Desert and the geographic distribution of Tillandsia landbeckii (Bromeliaceae) dune ecosystems. Plant Systematics and Evolution, 307(5), 1-22. https://doi.org/10.1007/s00606-021-01775-y

Giancristofaro, R. A., Salmaso, L. (2003). Model performance analysis and model validation in logistic regression. Statistica, 63(2), 375-396. https://doi.org/10.6092/issn.1973-2201/358

Gultepe, I., Tardif, R., Michaelides, S., Cermak, J., Bott, A., Bendix, J., Müller, M. D., Pagowski, M., Hansen, B., Ellrod, G., Jacobs, W., Toth, G., Cober, S. G. (2007). Fog research: A review of past achievements and future perspectives. Pure and applied geophysics, 164(6), 1121-1159. https://doi.org/10.1007/s00024-007-0211-x

Hastie, T., Tibshirani, R., Friedman, J. (2009). Unsupervised learning. In The elements of statistical learning (pp. 485-585). Springer, New York, NY. https://doi.org/10.1007/978-0-387-84858-7_14

Hedeker, D. (2003). A mixed-effects multinomial logistic regression model. Statistics in medicine, 22(9), 1433-1446. https://doi.org/10.1002/sim.1522

Hogan, R.J., Gaussiat, N., Illingworth, A.J., 2005. Stratocumulus liquid water content from dual-wavelength radar. Journal of Atmospheric and Oceanic Technology, 22(8), 1207-1218. https://doi.org/10.1175/JTECH1768.1

Houston, J. (2006). Evaporation in the Atacama Desert: An empirical study of spatio-temporal variations and their causes. Journal of Hydrology, 330(3-4), 402-412. https://doi.org/10.1016/j.jhydrol.2006.03.036

Jacob, P. (2013). Probability and Statistics for Engineers and Scientists (9th Edition). https://doi.org/10.1080/09332480.2013.845456

Jarraud, M. (2008). Guide to meteorological instruments and methods of observation (WMO-No. 8). World Meteorological Organization: Geneva, Switzerland, 29.
Kabacoff, R. I. (2011). R in Action: Data Analysis and Graphics with R. Shelter Island: Manning.

Klemm, O., Schemenauer, R.S., Lummerich, A., Cereceda, P., Marzol, V., Corell, D., van Heerden, J., Reinhard, D., Gherezghiher, T., Olivier, J., Osses, P., Sarsour, J.,Frost, E., Estrela, M.J., Valiente, J.A., Fessehaye, G.M. (2012). Fog as a fresh-water resource: Overview and perspectives. Ambio, 41(3), 221-234. https://doi.org/10.1007/s13280-012-0247-8

Koch, M.A., Kleinpeter, D., Auer, E., Siegmund, A., del Río, C., Osses, P., García, J.L., Marzol, M.V., Zizka, G., Kiefer, C. (2019). Living at the dry limits. Ecological genetics of Tillandsia lomas in the Chilean Atacama Desert. Plant systematics and Evolution, 305(10), 1041-1053. https://doi.org/10.1007/s00606-019-01623-0

Koračin, D., Dorman, C. E., Lewis, J. M., Hudson, J. G., Wilcox, E. M., Torregrosa, A. (2014). Marine fog: A review. Atmospheric Research, 143, 142-175. https://doi.org/10.1016/j.atmosres.2013.12.012

Korjus, K., Hebart, M. N., Vicente, R. (2016). An efficient data partitioning to improve classification performance while keeping parameters interpretable. PloS one, 11(8), e0161788. https://doi.org/10.1371/journal.pone.0161788

Larraín, H., Velásquez, F., Cereceda, P., Espejo, R., Pinto, R., Osses P., Schemenauer R.S. (2002). Fog measurements at the site "Falda Verde" north of Chañaral compared with other fog stations of Chile. Atmospheric Research, 64(1-4), 273-284. https://doi.org/10.1016/S0169-8095(02)00098-4

Lobos-Roco F, Villa-Guerau de Arellano J, Pedruzo-Bagazgoitia X (2018) Characterizing the influence of the marine stratocumulus cloud on the land fog at the Atacama Desert. Atmospheric Research, 214, 109-120. https://doi.org/10.1016/j.atmosres.2018.07.009

Lobos-Roco, F., Hartogensis, O., Vilà-Guerau de Arellano, J., De La Fuente, A., Muñoz, R., Rutllant, J., Suárez, F. (2021). Local evaporation controlled by regional atmospheric circulation in the Altiplano of the Atacama Desert. Atmospheric Chemistry and Physics, 21(11), 9125-9150. https://doi.org/10.5194/acp-21-9125-2021, 2021.

Lourakis, M. I. (2005). A brief description of the Levenberg-Marquardt algorithm implemented by levmar. Foundation of Research and Technology, 4(1), 1-6.

Muñoz, R. C., Zamora, R. A., Rutllant, J. A. (2011). The coastal boundary layer at the eastern margin of the southeast Pacific (23.4° S, 70.4° W): Cloudiness-conditioned climatology. Journal of climate, 24(4), 1013-1033. https://doi.org/10.1175/2010JCLI3714.1

Muñoz, R. C., Quintana, J., Falvey, M. J., Rutllant, J. A., Garreaud, R. (2016). Coastal clouds at the eastern margin of the southeast Pacific: Climatology and trends. Journal of Climate, 29(12), 4525-4542. https://doi.org/10.1175/JCLI-D-15-0757.1

Myers, T. A., Norris, J. R. (2013). Observational evidence that enhanced subsidence reduces subtropical marine boundary layer cloudiness. Journal of Climate, 26(19), 7507-7524. https://doi.org/10.1175/JCLI-D-12-00736.1

Nielsen, H. S. (1998). Discrimination and detailed decomposition in a logit model. Economics Letters, 61(1), 115-120. https://doi.org/10.1016/S0165-1765(98)00155-4

Osses McIntyre, P. E., Salvador, M. F., Cárdenas, R. N., Troncoso, P. C., Barros, H. L. (2005). Coastal fog, satellite imagery, and drinking water: student fieldwork in the Atacama Desert. Geocarto International, 20(1), 69-74. https://doi.org/10.1080/10106040508542338

Pinto, R., Barria, I.,Marquet, P. A. (2006). Geographical distribution of Tillandsia lomas in the Atacama Desert, northern Chile. Journal of arid environments, 65(4), 543-552. https://doi.org/10.1016/j.jaridenv.2005.08.015

Raupach, M. R. (2000). Equilibrium evaporation and the convective boundary layer. Boundary-Layer Meteorology, 96(1), 107-142. https://doi.org/10.1023/A:1002675729075

Raykar, V. C., Saha, A. (2015). Data Split Strategies for Evolving Predictive Models. In Joint European Conference on Machine Learning and Knowledge Discovery in Databases (pp. 3-19). Springer, Cham. https://doi.org/10.1007/978-3-319-23528-8_1

Rieck, M., Nuijens, L., Stevens, B. (2012). Marine boundary layer cloud feedbacks in a constant relative humidity atmosphere. Journal of the Atmospheric Sciences, 69(8), 2538-2550. https://doi.org/10.1175/JAS-D-11-0203.1

Rogers, D. P., Koračin, D. (1992). Radiative transfer and turbulence in the cloud-topped marine atmospheric boundary layer. Journal of Atmospheric Sciences, 49(16), 1473-1486. https://doi.org/10.1175/1520-0469(1992)049<1473:RTATIT>2.0.CO;2

Rutllant JA, Fuenzalida H, Aceituno P (2003). Climate dynamics along the arid northern coast of Chile: The 1997–1998 Dinámica del Clima de la Región de Antofagasta (DICLIMA) experiment. Journal of Geophysical Research: Atmospheres, 108(D17). https://doi.org/10.1029/2002JD003357

Schween JH, Hofmeister D, Loehnert U (2020). Filling the observational gap in the Atacama Desert with a new network of climate stations. Global Planetary Change 184,103034. https://doi.org/10.1016/j.gloplacha.2019.103034

Schulz, N., Boisier, J.P., Aceituno, P. (2011). Climate change along the arid coast of northern Chile. International Journal of Climatology, 32(12), 1803-1814. https://doi.org/10.1002/joc.2395

Serpetzoglou, E., Albrecht, B. A., Kollias, P., Fairall, C. W. (2008). Boundary layer, cloud, and drizzle variability in the southeast Pacific stratocumulus regime. Journal of climate, 21(23), 6191-6214. https://doi.org/10.1175/2008JCLI2186.1

Sharma, S., Vaishnav, R., Shukla, M. V., Kumar, P., Kumar, P., Thapliyal, P. K., ... Acharya, Y. B. (2016). Evaluation of cloud base height measurements from Ceilometer CL31 and MODIS satellite over Ahmedabad, India. Atmospheric Measurement Techniques, 9(2), 711-719. https://doi.org/10.5194/amt-9-711-2016

Stevens, B., Lenschow, D. H., Vali, G., Gerber, H., Bandy, A., Blomquist, B., ... Van Zanten, M. C. (2003). Dynamics and chemistry of marine stratocumulus—DYCOMS-II. Bulletin of the American Meteorological Society, 84(5), 579-594. https://doi.org/10.1175/BAMS-84-5-579

Stoltzfus, J. C. (2011). Logistic regression: a brief primer. Academic Emergency Medicine, 18(10), 1099-1104. https://doi.org/10.1111/j.1553-2712.2011.01185.x

Straub, D. J., Hutchings, J. W., Herckes, P. (2012). Measurements of fog composition at a rural site. Atmospheric environment, 47, 195-205. https://doi.org/10.1016/j.atmosenv.2011.11.014

Stull, R. (1988). An Introduction to Boundary Layer Meteorology. Kluwer Academic Publishers, The Netherlands. 670 pp.

Torregrosa, A., Combs, C., Peters, J. (2016). GOES-derived fog and low cloud indices for coastal north and central California ecological analyses. Earth and Space Science, 3(2), 46-67. https://doi.org/10.1002/2015EA000119

Van Heerwaarden, C. C., Vilà-Guerau de Arellano, J., Moene, A. F., Holtslag, A. A. (2009). Interactions between dry-air entrainment, surface evaporation and convective

boundary-layer development. Quarterly Journal of the Royal Meteorological Society: A journal of the atmospheric sciences, applied meteorology and physical oceanography, 135(642), 1277-1291. https://doi.org/10.1002/qj.431

Van Heerwaarden, C. C., Vilà-Guerau de Arellano, J., Gounou, A., Guichard, F., Couvreux, F. (2010). Understanding the daily cycle of evapotranspiration: A method to quantify the influence of forcings and feedbacks. Journal of Hydrometeorology, 11(6), 1405-1422. https://doi.org/10.1175/2010JHM1272.1

Viúdez-Mora, A., Costa-Surós, M., Calbó, J., González, J. A. (2015). Modeling atmospheric longwave radiation at the surface during overcast skies: The role of cloud base height. Journal of Geophysical Research: Atmospheres, 120(1), 199-214. https://doi.org/10.1002/2014JD022310

Man-Kui Wai, M. (1987). A numerical study of the marine stratocumulus cloud layer. Boundary-layer meteorology, 40(3), 241-267. https://doi.org/10.1007/BF00117450

Wang, P. K. (2013). Physics and dynamics of clouds and precipitation. Cambridge University Press.

Wetzel, P. J. (1990). A simple parcel method for prediction of cumulus onset and area-averaged cloud amount over heterogeneous land surfaces. Journal of Applied Meteorology and Climatology, 29(6), 516-523. https://doi.org/10.1175/1520-0450(1990)029<0516:ASPMFP>2.0.CO;2

Wetzel, P. J., Boone, A. (1995). A Parameterization for Land–Atmosphere-Cloud Exchange (PLACE): Documentation and testing of a detailed process model of the partly cloudy boundary layer over heterogeneous land. Journal of Climate, 8(7), 1810-1837. https://doi.org/10.1175/1520-0442(1995)008<1810:APFLCE>2.0.CO;2

Wood, R. (2012). Stratocumulus clouds. Monthly Weather Review, 140(8), 2373-2423. https://doi.org/10.1175/MWR-D-11-00121.1

Xu, Y., Goodacre, R. (2018). On splitting training and validation set: a comparative study of cross-validation, bootstrap and systematic sampling for estimating the generalization performance of supervised learning. Journal of Analysis and Testing, 2(3), 249-262. https://doi.org/10.1007/s41664-018-0068-2

Zuidema, P., Painemal, D., De Szoeke, S., Fairall, C. (2009). Stratocumulus cloud-top height estimates and their climatic implications. Journal of Climate, 22(17), 4652-4666. https://doi.org/10.1175/2009JCLI2708.1