ATMOSPHERIC WATER SUPPLY TO THE ATACAMA DESERT FROM NEWLY DEVELOPED SATELLITE REMOTE SENSING TECHNIQUES AND REANALYSIS

INAUGURAL–DISSERTATION

ZUR

ERLANGUNG DES DOKTORGRADES

DER MATHEMATISCH-NATURWISSENSCHAFTLICHEN FAKULTÄT

DER UNIVERSITÄT ZU KÖLN

VORGELEGT VON

CHRISTOPH BÖHM

AUS HALLE (SAALE)

KÖLN, 2020

BERICHTERSTATTER:
Prof. Dr. Susanne Crewell
Prof. Dr. Stephanie Fiedler
Prof. Dr. Jörg Bendix

TAG DER MÜNDLICHEN PRÜFUNG:
22.01.2021

Diese Arbeit wurde von der Mathematisch-Naturwissenschaftlichen Fakultät der Universität zu Köln im Jahr 2021 als Dissertation angenommenen.

ABSTRACT

Many facets of atmospheric water supply to the Atacama Desert are poorly understood. However, in-depth knowledge regarding water availability, moisture sources and the underlying mechanisms is required to investigate biological and geological processes and to identify potential mutual relationships.

This thesis provides a comprehensive meteorological perspective on the atmospheric water supply to the Atacama Desert within the context of the recent climate. Spatial and temporal variability of moisture as well as their controlling mechanisms depend on the type of water supply, i. e. clouds, water vapor, fog or precipitation.

To investigate the influence of the persistent stratocumulus cloud deck above the southeast Pacific on the desert region, a new cloud base height retrieval method is introduced. It allows to estimate the vertical position of these clouds, which can help to identify regions within the coastal desert that are potentially influenced by these clouds. A first application of this new method revealed a strong relation between stratocumulus properties and the isotopic composition of coastal *Tillandsia* populations.

The proximity of the Atacama Desert to main acting zones of the El Niño-Southern Oscillation (ENSO) phenomenon and of the Pacific Decadal Oscillation (PDO) together with results from previous studies suggest that modes of climate variability have strong influence on the moisture supply to this region. As oscillating extreme phases of these climate modes have recurring periods on the order of a few years to decades, a long data record is needed to study their impact. Therefore, spatio-temporal variability of integrated water vapor (IWV) provided by a century-spanning reanalysis data set is studied in relation to ENSO and PDO. It is shown that the reanalysis represents IWV in a suitable manner to study its long-term variability. On a decadal time scale, the PDO revealed a stronger coupling to IWV compared to ENSO.

According to a seasonal analysis, identified relationships between ENSO and IWV are in line with findings reported for precipitation in the northeastern Atacama. This suggests that IWV has the potential to serve as a proxy for precipitation. The ENSO signal is opposite for summer and winter season. The negative phase (La Niña) favors wetter summers and drier winters, whereas the positive phase (El Niño) is associated with drier summers and wetter winters. Besides, it is shown that enhanced IWV under La Niña conditions is not constrained to the northeastern part of the Atacama Desert but can reach even offshore regions near the west coast. This effect can be typically observed in the

summer season. Thus, the moisture can be supplied to the Atacama Desert from easterly or westerly sources depending on season and ENSO phase with regionally varying impacts.

Water vapor is a key variable which controls fog formation. While a few studies demonstrate the impact of fog on the coastal desert based on in-situ measurements as well as spatially and temporally limited satellited-based observations, this thesis introduces a novel satellite-based fog detection method which allows a region-wide assessment. An application of the algorithm for a 3-year period shows the spatial distribution of fog frequencies across the Atacama Desert. Aside from the coastal maximum, high fog frequencies are also revealed for isolated locations farther inland, which often coincide with salt flats within the central valley. The mechanisms driving fog formation within these inland regions remain unclear. The novel fog detection method creates the opportunity to further investigate this issue in future research.

Aside from westerly moisture sources associated with the Pacific Ocean and episodic easterly inflow from the continental interior, a third scenario is identified in this thesis. By investigating the role of atmospheric rivers for the Atacama Desert, it is revealed that moisture can be transported from the Amazon Basin across the Andes and the southeast Pacific towards the Atacama Desert. Furthermore, fractional precipitation rates of more than 50 % for various regions within the Atacama Desert demonstrate the importance of atmospheric rivers for this hyperarid environment.

ZUSAMMENFASSUNG

Viele Facetten des atmosphärischen Wassereintrags in die Atacama-Wüste sind kaum erforscht. Kenntnisse über die Wasserverfügbarkeit sowie die Quellen der Feuchtigkeit und die zugrunde liegenden Mechanismen sind jedoch erforderlich, um biologische und geologische Prozesse besser zu verstehen und mögliche wechselseitige Beziehungen zu identifizieren.

Diese Arbeit bietet eine umfassende meteorologische Perspektive auf den atmosphärischen Wassereintrag in die Atacama-Wüste im Kontext des gegenwärtigen Klimas. Die räumliche und zeitliche Variabilität des Wasserangebots und deren Kontrollmechanismen hängen davon ab, ob das Wasserangebot in Form von Wolken, Wasserdampf, Nebel oder Niederschlag betrachtet wird.

Um den Einfluss der persistenten Stratocumulus-Wolkendecke über dem Südostpazifik auf die Wüstenregion zu untersuchen, wird eine neue Methode zur Bestimmung der Wolkenbasishöhe entwickelt. So kann die vertikale Lage der Wolken vollständig bestimmt werden. Dies erlaubt es, Regionen entlang des Küstengebirges zu identifizieren, die möglicherweise von diesen Wolken beeinflusst werden. Eine erste Anwendung dieser neuen Methode ergab einen deutlichen Zusammenhang zwischen den Eigenschaften der Stratocumulus-Bewölkung und der Zusammensetzung von Stickstoffisotopen der Tillandsienpopulationen entlang der Küste.

Die Nähe der Atacama-Wüste zu den Hauptwirkungszonen der "El Niño-Southern Oscillation" (ENSO) und der Pazifischen Dekaden-Oszillation (PDO) sowie Ergebnisse früherer Studien legen den starken Einfluss dieser großskaligen Klimamoden auf den Wassereintrag in diese Region nahe. Da die sich abwechselnden Extremphasen dieser Klimamoden wiederkehrende Perioden in der Größenordnung von einigen Jahren bis Jahrzehnten aufweisen, ist ein langer Datensatz erforderlich, um ihren Einfluss zu untersuchen. In dieser Arbeit wird der integrierte Wasserdampf (IWV) aus einem Reanalyse-Datensatz, der mehr als ein Jahrhundert umspannt, verwendet, um den Einfluss von ENSO und PDO zu untersuchen. Zunächst wird gezeigt, dass die Reanalyse den IWV ausreichend genau repräsentiert, um seine Langzeitvariabilität zu untersuchen. Auf einer dekadischen Zeitskala zeigte die PDO eine stärkere Kopplung an IWV als ENSO.

Festgestellte saisonale Beziehungen zwischen ENSO und IWV ähneln den Beziehungen, die bereits in früheren Studien für die nordöstliche Atacama-Wüste in Bezug auf Niederschlag gezeigt wurden. Dies deutet darauf hin, dass IWV das Potenzial hat, als Approximation für Niederschlag zu dienen. Das ENSO-Signal ist für die Sommer- und Wintersaison entgegengesetzt. Die negative Phase (La Nina) begüns-

tigt feuchtere Sommer und trockenere Winter, während die positive Phase (El Niño) zu trockeneren Sommern und feuchteren Wintern führt. Insbesondere für die Sommersaison wird gezeigt, dass erhöhter integrierter Wasserdampfgehalt unter La Niña Bedingungen nicht auf den nordöstlichen Teil der Atacama-Wüste beschränkt ist, sondern sogar den der Westküste vorgelagerten Pazifik erreichen kann. Zusammenfassend ergibt sich, dass Wassereintrag in die Atacama-Wüste je nach Jahreszeit und ENSO-Phase aus östlich oder westlich gelegenen Quellen stammt, wobei es regionale Unterschiede geben kann.

Wasserdampf ist eine Schlüsselvariable für Nebelbildung. Während einige Studien den Einfluss von Nebel auf den Küstenbereich der Atacama-Wüste anhand von in-situ Messungen zum einen und räumlich und zeitlich begrenzten satellitengestützten Beobachtungen zum anderen belegen, wird in dieser Arbeit eine neue satellitenbasierte Nebeldetektionsmethode vorgestellt, die eine regionale Abschätzung ermöglicht. Eine Anwendung des Algorithmus über einen Zeitraum von 3 Jahren zeigt die räumliche Verteilung der Nebelauftrittshäufigkeiten in der Atacama-Wüste. Neben dem Küstenmaximum zeigen sich auch hohe Nebelfrequenzen für einzelne Regionen weiter im Landesinneren, die vielfach mit Salzpfannen in der zentralen Depression in Verbindung stehen. Die Mechanismen, die die Nebelbildung in diesen Binnenregionen antreiben, sind weiterhin unbekannt. Die neuartige Nebeldetektionsmethode bietet die Möglichkeit, dieses Problem in zukünftigen Forschungsarbeiten weiter zu untersuchen.

Neben den mit dem Pazifik verbundenen westlichen Quellen der Feuchtigkeit und dem episodischen östlichen Zufluss aus der kontinentalen Feuchtzone Südamerikas wird in dieser Arbeit ein weiterer Pfad identifiziert, der mit einem Wassereintrag verbunden ist. Durch die Untersuchung der Rolle atmosphärischer Flüsse (engl. atmospheric rivers) für die Atacama-Wüste wird deutlich, dass Feuchte vom Amazonasbecken über die Anden und den Südostpazifik in Richtung Atacama-Wüste transportiert werden kann. Darüber hinaus zeigen Niederschlagsanteile von überwiegend mehr als 50 % in der Atacama-Wüste die Bedeutung atmosphärischer Flüsse für diese hyperaride Umgebung.

CONTENTS

ACRONYMS

AR Atmospheric River

CR2 Center for Climate and Resilience Research

CRC Collaborative Research Center

ECMWF European Centre for Medium-Ranged Weather Forecast

ENSO El Niño Southern Oscillation

EOF Empirical Orthogonal Function

ERA-20C ECMWF's Twentieth Century Reanalysis

ERA5 ECMWF's 5th generation reanalysis

GOES Geostationary Operational Environmental Satellite

HadISST1 Hadley Centre's sea ice and sea surface temperature data set

HYSPLIT Hybrid Single Particle Lagrangian Integrated Trajectory Model

IGRA Integrated Global Radiosonde Archive

IVT Integrated Water Vapor Transport

IWV Integrated Water Vapor

LCL Lifting Condensation Level

MBL Maritime Boundary Layer

MIBase MISR Cloud Base Height

MISR Multi-angle Imaging SpectroRadiometer

MODIS Moderate Resolution Imaging Spectroradiometer

PDO Pacific Decadal Oscillation

RMSE Root Mean Square Error

SRTM Shuttle Radar Topographic Mission

SST Sea Surface Temperature

WRF Weather Research and Forecasting Model

UTC Coordinated Universal Time

VOCALS-REx Variability of the American Monsoon System (VAMOS) Ocean-Cloud-Atmosphere-Land Study – Regional Experiment

Part I

INTRODUCTION

1 MOTIVATION

Figure 1.1: Photograph of the coastal Atacama Desert taken in March 2017 picturing the Pacific ocean, the maritime boundary layer, and the free troposphere. The maritime boundary layer is visible as the optically denser part of the air, intersecting with the emerging coastal cliff, which hosts locally adapted plant communuties. The free troposhere aloft is much clearer compared to the boundary layer.

Water is a vital ingredient in the evolution of the Earth. By triggering biological and geological processes, the availability and amount of water plays a decisive role in shaping our planet. Mostly, these processes are considered independently of each other and analyzed on a stand-alone basis. However, the fact that they respond to a common trigger encourages the assumption of mutual dependencies and possible interactions (Collaborative Research Center (CRC) 1211; sfb1211.uni-koeln.de; Dunai et al., 2020).

Innumerable biological and geological processes occur simultaneously on various time scales. This makes it difficult to disentangle potential relationships. To reduce some of this complexity, it would be beneficial to find a location where these processes happen more isolated and at a slower pace. An ecosystem where the abundance of water as the common trigger is reduced would best approximate these

criteria. As the oldest desert and presumably driest place on Earth, the Atacama Desert appears to be a suitable environment. While intermittent periods of enhanced water availability allow life and surfaces to evolve, enduring dry (hyperarid) conditions conserve the resulting traces through geological eras. The degree of aridity and the timing of pluvial episodes are controlled by the progress of climate change and associated atmospheric processes. This implies the immediate need for a meteorological perspective on this matter.

For many regions, precipitation is the prevailing water source. In hyperarid regions, however, other factors such as clouds and water vapor dominate the water cycle. To describe the atmospheric water supply to the Atacama Desert, following key aspects have to be considered.

The maritime boundary layer is the lowest tropospheric layer that is in contact with the ocean surface. Near the Atacama coast, it typically extends to a height of about 1 km (e.g. Rahn and Garreaud, 2010; Muñoz et al., 2016).

One aspect pose clouds which form over the southeast Pacific near the coast of the Atacama Desert. These low stratiform clouds frequently cap the moist maritime boundary layer. When they are transported onshore, they intercept with the local orography. The towering coastal cliff and mountain range with typical heights between 500 m and 2000 m constitute a natural barrier hindering these clouds from penetrating inland (Fig. 1.1). Wherever these mountains immerse into the cloud, fog oases arise enabling unique ecosystems (Pinto et al., 2006; Cereceda et al., 2008b; Lobos Roco et al., 2018). In these habitats, specialized plant communities, such as *Tillandsia*, satisfy their water and nutrition demand almost exclusively from fog water supply (Rundel et al., 1997; Pinto et al., 2006; Westbeld et al., 2009; González et al., 2011). The locations of these coastal fog oases are closely linked to the cloud heights.

A second aspect is the availability of water vapor. Making up about 99.5 % of the total water in the atmosphere (Stevens and Bony, 2013), water vapor is the most important source for precipitation and a key variable for fog formation and dew. Besides these indirect ways of dispensing liquid water after condensation, water vapor itself poses a direct water source for soils in arid regions and stimulates microbial activity and diversity via adsorption (McHugh et al., 2015; Crits-Christoph et al., 2013). Furthermore, relative humidity, which is closely related to water vapor content, determines phase transitions between gypsum, anhydrite and their intermediate phases (Tang et al., 2019; Ritterbach and Becker, 2020). This indicates its potential role in soil formation, which requires further exploration. Relative humidity and the isotopic composition of atmospheric water vapor, which depends on its source and pathway, determine isotopic fractionation during evaporation. Therefore, they are essential variables for the development of a paleo-humidity proxy (Craig and Gordon, 1965; Surma et al., 2018; Gázquez et al., 2018).

A third aspect is the distribution of fog, which is closely related to coastal cloud heights and water vapor. When coastal clouds are present at a sufficient height, they can cross the coastal cliff and penetrate

through the coastal mountain range. This is frequently observed at certain corridors such as canyons or distinct locations where the coastal cliff is lower (Farías et al., 2005). While in most cases clouds eventually dissipate due to mixing with dry inland air, the moisture enhanced maritime air can reach places even farther away from the coast (Schween et al., 2020). During nocturnal cooling via thermal radiation, this moisture enhanced air might reach saturation. Therefore, the initiation of condensation processes resulting in radiation fog seems plausible. Even though favorable conditions for its formation were found (Cereceda et al., 2002; Westbeld et al., 2009), radiation fog has not been reported due to lacking observations. While various studies show the impact of fog on localized coastal loma vegetation (e.g. Rundel et al., 1997; Pinto et al., 2006; Cáceres et al., 2007; Latorre et al., 2011; González et al., 2011; Río et al., 2018; Lehnert et al., 2018b; Lobos Roco et al., 2018), the importance of fog for biological and geological processes has barely been explored on a region-wide scale. Instead, Cáceres et al. (2007) and Westbeld et al. (2009) emphasize the urgent need for further investigations on the dynamics involved in fog formation and dissipation and the role of radiation fog, respectively.

Lomas are isolated and diverse vegetation formations in the coastal fog zone of the Atacama Desert (e.g. Pinto et al., 2006).

An investigation of these aspects would not only improve our understanding of the contemporary atmospheric water supply to the Atacama Desert, but also provide a meteorological basis for various biological and geological analyses in this region. For example, it would make it possible to constrain thresholds for biological and geological processes which are observable under current climatic conditions.

The framework for this multidisciplinary approach is provided by the CRC 1211, a research program funded by the German Research Foundation. With the focus on interactions between biological, geological and atmospheric processes, the program allows to exploit arising synergies and to conduct cross-validations. In its turn, it should lead to an improvement of applied methods and enable more solid conclusions in all three disciplines.

The synergies from this approach are exemplified here from the meteorological point of view. The first and most immediate goal of the meteorological analysis is to gain more detailed insights into the underlying atmospheric processes. This involves identifying and quantifying relevant relationships between large-scale phenomenons, such as the internal climate mode El Niño Southern Oscillation (ENSO), synoptic features and regional water availability patterns in the Atacama Desert. Ultimately, such knowledge would contribute to the geoscience community in the following way: larger scale phenomenons could be linked to the regional water variability and thus be related to local biological or geological processes. For example, the isotopic composition of water bound in gypsum, which can serve as a medium for paleo-humidity reconstruction (Gázquez et al., 2018; Voigt, 2020), could be associated with a broader context. By using modern observa-

tions, functional relationships between variables quantifiable through climate archives on one side and varying ambient atmospheric conditions on the other side can be estimated and verified in the context of the recent climate. When analyzing desert archives, such calibrated relationships would allow a more comprehensive interpretation and support the reconstruction of climate history.

Summarizing the above, this thesis contributes to the overall purpose of the CRC by providing a comprehensive meteorological perspective on the atmospheric water supply to the Atacama Desert under recent climate conditions.

2

BACKGROUND AND CLIMATIC FEATURES OF THE ATACAMA DESERT

This chapter provides an overview on the state-of-the-art knowledge about atmospheric features associated with water supply to the Atacama Desert. After an introduction to the general climatic setting which results from the geographic location (Chapter 2.1), more details are provided on stratocumulus clouds and fog (Chapter 2.2) as well as the influence of large-scale drivers (Chapter 2.3) and synoptic drivers (Chapter 2.4).

2.1 CLIMATIC SETTING

Hyperarid is a term to describe regions for which precipitation is less than 5 % of the potential evapotranspiration (UNEP, 2011).

The Atacama Desert is located at the western coast of South America (Fig. 2.1). Hyperarid conditions persist between the southeast Pacific and the foothills of the high Andes from West to East and between 15°S and 30°S (UNEP, 2011; Houston and Hartley, 2003). The prevailing aridity is typical for subtropical locations at the subsiding branch of the Hadley cell circulation and can be found in other subtropical regions, e. g. the Namib Desert.

Due to the enduring large-scale subsidence, strong semi-persistent surface anticyclones emerge, forming the subtropical high pressure belt. This includes the southeast Pacific anticyclone, which has specific implications for the Atacama Desert. Resulting southerly winds at the eastern side of the high pressure system enact stress on the sea surface driving the Humboldt Current, which transports cold sea water from higher latitudes towards the tropics along the South American west coast (Montecino and Lange, 2009).

Furthermore, friction effects cause a counterclockwise rotation of the flow towards the west with increasing depth within the near surface ocean layer (Ekman spiral). This leads to near coastal upwelling of deep ocean water, which provides further cooling of the Sea Surface Temperature (SST) embedded in the Humboldt Current system. The strong temperature contrast between the cold SST and the subsiding warm air leads to a stable stratification. This hinders the development of precipitating clouds and results in a decoupling of the Maritime Boundary Layer (MBL) and the free troposphere aloft. Turbulent mixing within the boundary layer enforces an approximately adiabatic temperature profile and a nearly constant specific humidity profile. Under these conditions, the typically observed stratocumulus deck can form right below the base of the inversion (more details in Chapter 2.2).

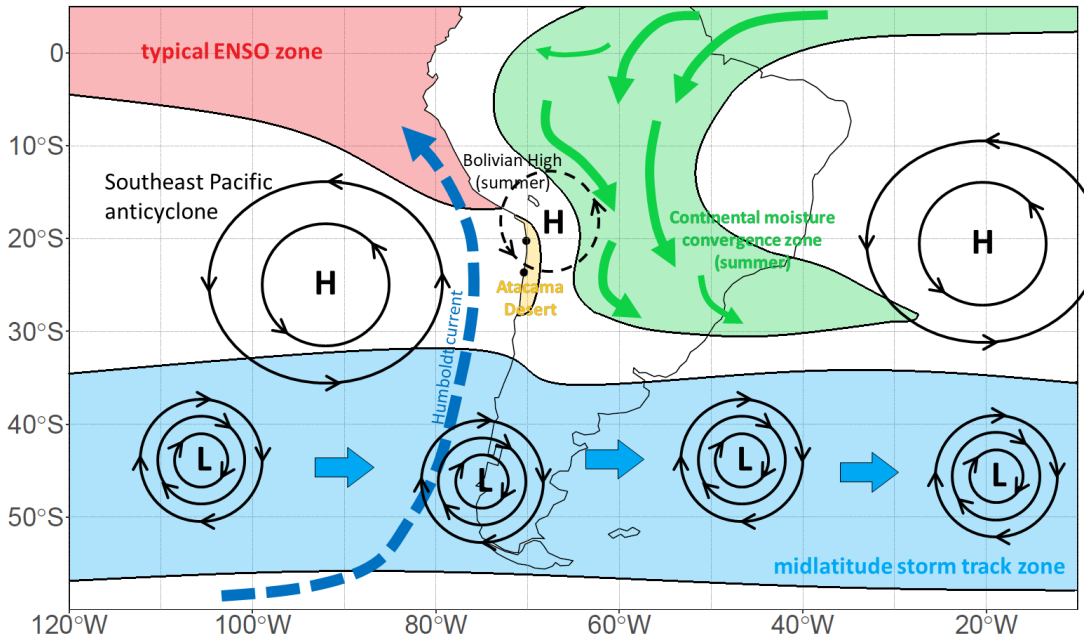

Figure 2.1: Scheme of key factors impacting the climate of the Atacama Desert (yellow shading): Midlatitude storm track zone (blue shading), ENSO zone (red shading), which denotes the region with typically largest SST variability, the continental moisture convergence zone (green shading) with Atlantic moisture transport (austral summer situation). Furthermore, the southeast Pacific anticyclone, the Bolivian high (summer seasonal upper tropospheric high pressure system) and the Humboldt Current (blue dashed arrow) are denoted.

In contrast to other subtropical west coast deserts, the aridity of the Atacama Desert is further enhanced due to the unique local topography. At the western margin, the coastal cliff poses an immediate barrier hindering the moist maritime layer from penetrating inland. Furthermore, diurnal heating of the land surfaces creates a strong thermal contrast between land and sea. These two factors result in the formation of two separate circulation cells (Rutllant et al., 2003). A lower cell is constrained to the maritime boundary layer with updrafts near the coastal cliff in the course of the diurnal heating. These updrafts result in near-surface westerly winds superimposed on the prevailing southerlies and an offshore return flow below the inversion layer. During night time, the circulation reverses and weakens (Rutllant et al., 2013). An upper cell is driven by strong daytime heating of the Andean slopes ("Andean pumping" Rutllant et al., 2013). This results in strong westerly winds near the surface of the desert and a return flow at higher altitudes. This flow pattern increases the subsidence above the coastal desert during the afternoon (Rutllant et al., 2013).

At the eastern margin, the high Andean mountain ranges are believed to create a rain shadow effect. According to in-situ measurements, precipitation decreases rapidly with decreasing heights at the

western slopes of the Andean cordillera which can be attributed to barrier effect of the Andes preventing moisture intrusion from the Amazon Basin (Houston and Hartley, 2003). Controversially, several modeling studies revealed that a lowering of the Andes does not have a significant effect on precipitation and moisture within the Atacama Desert (Ehlers and Poulsen, 2009; Sepulchre et al., 2011; Garreaud et al., 2010).

The interplay of the factors mentioned above result in annual precipitation rates below 2 mm in the hyperarid core (Houston, 2006). The precipitation pattern shows a North–South gradient due to winter storm tracks (Fig. 2.1) occasionally reaching the southern Atacama Desert. Additionally, a West–East gradient takes contour with higher precipitation rates at higher elevations at the western slopes of the Andes in connection to moist easterlies during the summer (Houston, 2006). These easterlies are caused by an upper tropospheric high pressure system called Bolivian high which is created by deep convection over the Amazon Basin during the austral summer season (Lenters and Cook, 1997). During episodic southward shifts of the Bolivian high, easterly flows transport large amounts of moisture causing heavy thunderstorms in the Altiplano (Garreaud et al., 2003). This mechanism is also responsible for the summer time precipitation maximum at the western slopes of the Andes in the northern Atacama. Seasonal peak precipitation rates decline rapidly with decreasing topography to the west (Houston and Hartley, 2003; Houston, 2006). Since precipitation rates are so low, other moisture sources such as fog (Chapter 2.2) and water vapor (Chapter 2.3) become dominant within the hyperarid core of the Atacama.

2.2 STRATOCUMULUS AND FOG

The stratocumulus turning into fog at its interception with the coastal cliff and mountain range constitutes the life vein for *Tillandsia* and other plant communities (Rundel et al., 1997; Muñoz-Schick et al., 2001; Pinto et al., 2006; Westbeld et al., 2009; González et al., 2011). Recent discoveries also attribute activation of photosynthesis of soil organisms to fog water supply pointing out the importance of this "living skin" for carbon and nitrogen fixation as well as soil formation through bio-weathering (Lehnert et al., 2018a; Jung et al., 2020). Intuitively, the exact location of these biologically active zones depends on the top and base height and on the frequency of the maritime stratocumulus. Underlying concepts of stratocumulus formation and dissipation are introduced in Chapter 2.2.1. These concepts, which have been reviewed by Wood (2012), encompass the relation between atmospheric drivers and cloud properties, such as height and thickness. Thereafter, current findings regarding variability of cloud heights and cloud cover fraction

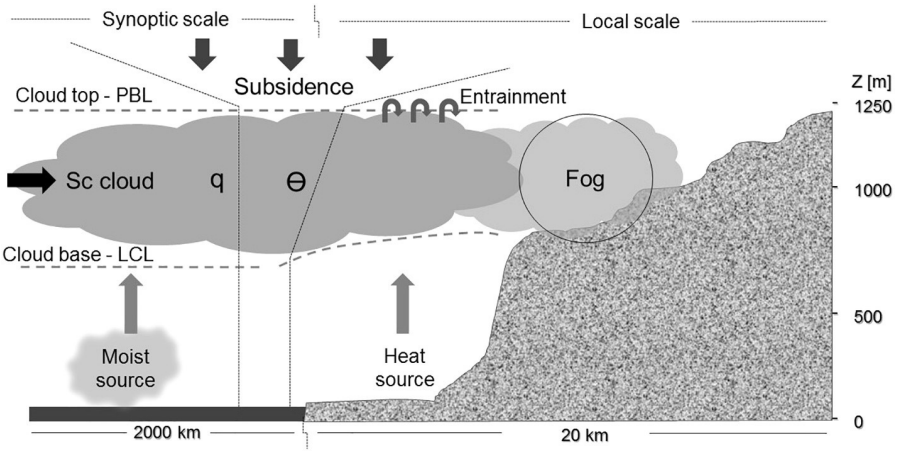

Figure 2.2: Physical processes of the marine stratocumulus in the coastal region. A well mixed MBL features a vertically constant specific humidity q and a constant potential temperature θ which increases slightly within the cloud layer towards the cloud top. Figure taken from Lobos Roco et al. (2018). ©2018 Elsevier B.V. Used with permission.

(Chapter 2.2.2) as well as fog water supply and fog detection (Chapter 2.2.3) are described.

2.2.1 *Stratocumulus – conceptual description*

Above the cold Humboldt Current and below large-scale subsidence of warm and dry air, the southeast Pacific region offers ideal conditions for stratocumulus formation. Consequently, it hosts the largest and most persistent subtropical stratocumulus deck on Earth (Klein and Hartmann, 1993). Conceptually, these clouds can form when turbulent mixing moistens the boundary layer and stratifies the temperature resulting in a neutral layer (Paluch and Lenschow, 1991). The initial turbulence is either generated by vertically sheared horizontal winds or by buoyancy resulting from the moisture and heat exchange at the ocean-atmosphere interface. The latter is provided when evaporation and temperature exchange enhance the virtual temperature of an air parcel allowing ascent (Paluch and Lenschow, 1991, and Fig. 2.2 for a schematic overview).

When the subsidence inversion is above the Lifting Condensation Level (LCL), cloud formation starts at the height at which the upper part of the mixed layer reaches saturation (Randall and Suarez, 1984). The cloud may grow up to the inversion base height. Once the cloud began to form, it emits infrared radiation at its top which leads to a cooling of the cloud layer (Paluch and Lenschow, 1991). The radiative cooling sharpens the inversion and generates a convective circulation which becomes the main driver maintaining the cloud (Lilly, 1968). This convection enhances the turbulence, which, in turn,

homogenizes the cloud layer and couples it to the moisture source (ocean-atmosphere interface). Several feedback mechanisms may alter the height of the MBL, the LCL, the thickness of the cloud and as a result its liquid water content (Albrecht et al., 1990; Bretherton and Wyant, 1997; Wood, 2012).

While the stratocumulus forms over ocean, it is modulated near coastal areas due to diurnally varying land-sea interactions (Sunuararajan and Tjernström, 2000). At night time, a stronger cooling of the land surface compared to the sea surface leads to predominant descent of air parcels at the ocean side of the coastal cliff. This results in a boundary layer local circulation cell with an onshore flow near the top of the boundary layer, advecting the stratocumulus inland (Rutllant et al., 2003).

Fog is formed where the cloud intercepts with the orography depending on the heights of the cloud margins. During daytime, the circulation of the maritime boundary layer reverses and intensifies in particular in the afternoon due to stronger heating of the slopes of the coastal cliff compared to the sea surface (Rutllant et al., 2003). Thus, the stratocumulus is pushed offshore. The heated land surface also lifts the LCL thinning the cloud from below (Rogers and Koraccn, 1992).

Additionally, the upper circulation cell driven by the "Andean pumping" (Rutllant et al., 2013) strengthens the coastal subsidence, which lowers the inversion base height (Rutllant et al., 2003). This leads to dissipation of the cloud by thinning from the top. During the afternoon, isolated convective cumulus clouds can form at the coastal cliff. They are related to the updraft of the daytime marine boundary layer circulation and form frequently at identical locations possibly related to the shape of the coastal shoreline and orography (Cereceda et al., 2002).

2.2.2 *Coastal cloud observations*

A recent study based on more than 40 years of observations at three coastal sites (Arica, Antofagasta, Iquique) analyzes the diurnal cycle of coastal clouds (Muñoz et al., 2016). While higher cloud cover fraction together with lower cloud base heights prevail during the night, lower cloud cover fraction together with higher cloud base heights predominate during the day. The observed cloud cycle is consistent with the previous conceptual considerations.

These observations also reveal a distinct seasonal cycle with the highest cloud cover fraction and lowest cloud base heights during winter and spring (May–November) and lower cloud fractions and higher cloud base heights for summer and fall (December–April) (Muñoz et al., 2016). This is consistent with the seasonality of the SST and the subsidence strength according to the position of the southeast

Pacific anticyclone, which shifts southward during summer leaving a weaker subsidence (Garreaud et al., 2009).

Furthermore, negative trends in cloud base heights with different onset times for the three locations were identified between 1995 and 2010 (Muñoz et al., 2016). For example, the annual mean cloud base height at Antofagasta decreased about 200 m between 1985 and 2005. For Arica, the overall trend is also negative but weaker because a positive trend is detected for the core summer season (January and February). Varying seasonal trends are also reported regarding cloud cover fraction with an increasing signal for winter and spring (June–November) and decreasing signal for summer and fall (December–May).

The negative cloud base height trends are concomitant with a negative trend of inversion base height determined from radiosonde data for the period between 1960–2009 (Schulz et al., 2012). However, the decrease of the inversion base height is of lesser magnitude compared to the decrease of the cloud base height, which implies that the cloud thickness increased within the analyzed period (Muñoz et al., 2016). A thickening cloud layer is consistent with an increasing inversion strength determined from reanalysis data for the southeast Pacific region (period 1984–2009, Seethala et al., 2015). The stronger inversion hinders cloud dissipation by decreasing the entrainment rate (Wood, 2012).

Different cloud cover fraction trends with opposite signs for spring and fall season compared to the report by Muñoz et al. (2016) are revealed if cloud cover fraction averages over a total of seven available coastal stations between 20°S and 30°S are considered (Eastman and Warren, 2013). While Muñoz et al. (2016) argue the difference could be caused by differences in the data processing or slightly varying time periods, it may also indicate that the spatial variability of the coastal stratocumulus is not sufficiently represented by these few observations. This is further supported by different onset times of the negative cloud base height trend (Muñoz et al., 2016).

However, a region-wide cloud height assessment is currently lacking in the literature. A region-wide coastal cloud height climatology would allow to constrain the role of clouds for biological and geological processes.

A data set of cloud top and base heights with high spatial resolution could be used for model and reanalysis evaluation. Ultimately, the use of reanalysis could be beneficial to extent the limited temporal coverage of satellite-based measurements. To achieve the desired spatial coverage, satellite remote sensing would be beneficial. However, to date, no operational satellite-based retrieval method of cloud base height is available.

2.2.3 *Fog observations*

Fog and low cloud cover over the coastal part of the Atacama Desert are strongly correlated to maritime stratocumulus cover (Cereceda et al., 2008b). The research carried out to study fog in this region has been mainly based on (i) in-situ fog collection (e. g. Cereceda et al., 2002; Larraín et al., 2002; Cáceres et al., 2007; Cereceda et al., 2008b; Lobos Roco et al., 2018; Río et al., 2018) mostly using standard fog collectors (Schemenauer and Cereceda, 1994a) and (ii) satellite remote sensing of low clouds (e. g. Farías et al., 2005; Osses et al., 2005; Cereceda et al., 2008b; Lehnert et al., 2018b; Río et al., 2018).

A major study site for in-situ fog collection is Alto Patache, a research station operated by the Centro del Desierto de Atacama UC. It is located on top of the coastal cliff at a height of 850 m above sea level and directly faces the Pacific ocean. For this location, the seasonal cycle of fog water yields has been derived from a 17-year long monthly time series (Río et al., 2018). Whereas highest fog water yields are observed for winter (July, August, September), lowest average fog water amounts are collected during summer seasons (December-April). These findings are consistent with the seasonal cycle of cloud cover fractions observed at other coastal stations below the coastal cliff (Muñoz et al., 2016) and with the seasonality of the general driving factors for stratocumulus (c. f. Chapter 2.2.1).

Furthermore, a vertical gradient has been observed along an inclining transect along the coastal cliff near Alto Patache (Cereceda et al., 2008b). Annual fog water yields decrease rapidly from more than $7\,\mathrm{l\,m^{-2}}$ at the top (850 m) to almost zero at 650 m. This matches with mean annual cloud base heights of around 800 m at Iquique (north of Alto Patache) and 650 m at Antofagasta (south of Alto Patache) (Muñoz et al., 2016) considering that the liquid water content increases between cloud base and cloud top height (Bretherton, 1997).

Farther inland, amounts of fog water collected at individual sites decrease (Cereceda et al., 2002; Cáceres et al., 2007; Cereceda et al., 2008b). For a transect between the coastal region and the Yungay area, declining abundance of hypolithic cyanobacteria and reduced fog water yields have been observed consistently (Warren-Rhodes et al., 2006; Cáceres et al., 2007).

Moreover, few fog events have been observed even beyond the coastal mountain range 45 km inland from Alto Patache within the Pampa de Tamarugal (central depression) during a 10-month field campaign (Cereceda et al., 2002). Occasionally, no concomitant fog was observed at the coastal station, which might be an indication of other processes being at play besides advection. These could be, for example, dissipating fog leaving a humid air mass to condense locally or diurnal evaporation of ground water which condenses during nocturnal

cooling (Cereceda et al., 2002). However, not enough observations are available to investigate such hypotheses.

Satellite remote sensing of low clouds has been applied to investigate spatial patterns to derive seasonal and diurnal variations (e. g. Farías et al., 2001; Farías et al., 2005; Cereceda et al., 2008b; Lehnert et al., 2018b). To distinguish between low and high clouds, spectral thresholds were applied. However, it is difficult to infer from satellite data whether such low clouds intercept with the ground. Nevertheless, these observations could confirm the seasonal cycle derived from the in-situ measurements with most persistent cloud cover over the coastal region during winter (Farías et al., 2005; Cereceda et al., 2008b; Lehnert et al., 2018b). The diurnal cycle of continental low cloud coverage has been determined using observations from the Geostationary Operational Environmental Satellite (GOES) for two individual months (August 2001, January 2002). A diurnal minimum appears between 10 a. m. and 3 p. m. local time.

Furthermore, satellite-based observations confirmed the occasional occurrence of low clouds in the central valley (Farías et al., 2005; Cereceda et al., 2008b). These occurrences were linked to cloud advection through corridors which were identified from satellite images and grouped into coastal corridors and interior corridors (Farías et al., 2005).

Satellite-based observations are advantageous regarding spatial coverage compared to in-situ measurements. However, satellite-based studies have been carried out only for short periods and limited regions within the Atacama Desert. In addition, fog and low clouds have not been distinguished so that it is not clear whether detected cloud features have the potential to supply water to the soil and plants and trigger biological and geological processes.

2.3 LARGE-SCALE DRIVERS

In this thesis, the term large-scale drivers refers to phenomena acting on time scales between multiple months and decades and spatial scales larger than 1000 km. Large-scale drivers are distinguished from synoptic drivers (Chapter 2.4) which refer to atmospheric phenomena of spatial and temporal dimensions up to a few 100 km and a few days, respectively. For example, dominant patterns of climate variability, such as the ENSO phenomenon and the Pacific Decadal Oscillation (PDO), act on the large scale both temporally and spatially with periods between 2 to 7 years (ENSO; Timmermann et al., 2018) and longer (PDO; Newman et al., 2016) and with primary acting zones spanning the tropical (ENSO) and the northern Pacific (PDO) ocean basin. These two climate modes have been associated with impacts on weather patterns even in remote regions through global teleconnections. While other climate indices exist, impacts on weather patterns across the Atacama

Desert and South America have mostly been related to ENSO and PDO in previous studies (Chapter 2.3.2). Oscillations of these two climate modes result from coupled ocean-atmosphere feedback mechanisms (Chapter 2.3.1).

2.3.1 *Pacific climate modes*

The ENSO phenomenon is the strongest year-to-year climate fluctuation of the global climate system (Timmermann et al., 2018). It is characterized by periodically recurring phases of warm (El Niño) and cold (La Niña) SSTs in the central or eastern tropical Pacific (Fig. 2.3) with recurring periods of extreme phases between three and seven years which have global implications, for instance, on precipitation and surface temperature (e. g. Davey et al., 2014). The state-of-the-art understanding of the guiding processes driving ENSO is reviewed by Timmermann et al. (2018). During the neutral phase, prevailing easterly trade winds steer surface water from the eastern to the western tropical Pacific. While on the eastern ocean rim Ekman transport leads to upwelling of cold water, convergence in the western Pacific results in downwelling of warm surface water. This process maintains a tilt of the thermocline along the equatorial Pacific with greater depth in the west and shoal depth in the east resulting in a charged western Pacific warm pool with great heat content.

Anomalous westerly wind events, which are favored by enhanced SST (Vecchi and Harrison, 2000), result from weather noise. They can trigger an oceanic downwelling Kelvin wave to propagate eastward. It transports warm water to the east expanding the western Pacific warm pool and suppresses the upwelling of cold water. This results in a deepening of the thermocline in the central and possibly the eastern Pacific, which gives rise to the development of an El Niño event.

Once a positive SST anomaly is provided, the interplay of positive feedback mechanism enhances this initial warming. A weaker SST gradient leads to weaker trade winds which, in turn, weaken the west-east SST gradient even more (Bjerknes feedback). Furthermore, the weaker trade winds weaken the advection of cold water from the eastern Pacific (zonal advective feedback) and reduce the upwelling of colder deep ocean water (Ekman feedback). Due to the deepened thermocline such upwelling gives rise to anomalously warm water in the eastern Pacific (thermocline feedback).

Negative feedback mechanisms dampen the intensifying positive SST anomaly. Via enhanced radiative and turbulent heat fluxes associated with tropical convection and cloud formation, heat is transferred from ocean to atmosphere (thermal damping). Horizontal and vertical transports of colder surrounding water into the region of enhanced SST anomaly counteract the positive feedback mechanisms (dynamical damping Lübbecke and McPhaden, 2013).

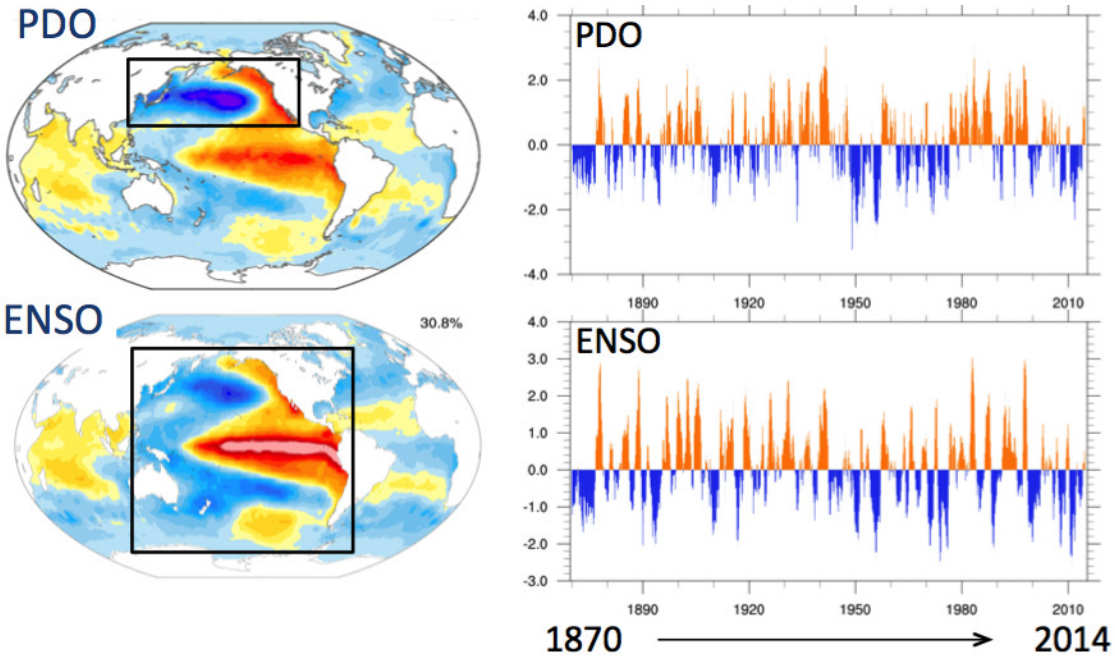

Figure 2.3: SST anomaly patterns for the PDO (top left) and ENSO (bottom left) derived by regressing monthly SST anomalies onto the principle component time series (PDO at top right, ENSO at bottom right) of the leading empirical orthogonal function for the area indicated by the black rectangles. Analysis is based on the Hadley Centre's sea ice and sea surface temperature data set (HadISST1) (Rayner et al., 2003) for the period between 1870–2014. This figure is taken from `https://climatedataguide.ucar.edu/climate-data/pacific-decadal-oscillation-pdo-definition-and-indices` (Deser et al., 2016) and has originally been adapted from Deser et al. (2010).

Anomalous westerlies associated with tropical deep convection are strongest around the equator and decrease towards the poles (Clarke, 1994). This results in a wind stress curl anomaly which induces poleward transport of upper level ocean water ultimately discharging the heat from the equatorial region (Clarke et al., 2007). Therefore, after full development of the El Niño state, the dampening feedbacks become more dominant. Reinstated easterly trade winds feature an upwelling oceanic Kelvin wave which propagates westward. Thereby, the ocean state transitions into the cold La Niña phase. This decrease of the eastern and central Pacific SSTs is enhanced by the aforementioned positive feedback mechanisms (Bjerknes, zonal advective, Ekman and thermocline feedback) as a response to enhanced trade winds and stronger west-east SST gradient.

A seasonal phase locking of the ENSO cycle is indicated by observed variability of eastern and central tropical Pacific SST. The highest variability is detected for austral late spring and early summer season, while the lowest variability is observed for austral fall (Wengel et al.,

2018; Timmermann et al., 2018). The reason for this phase lock is seasonally varying stability of the coupled ocean atmosphere system. The positive feedback mechanisms are greatest between September and December when eastern Pacific cold tongue is normally most pronounced, whereas the negative feedback mechanisms are greatest between January and April (Wengel et al., 2018). Therefore, El Niño development is typically initiated in austral fall, intensifying through winter and spring and reaching maximum SST anomalies in the central or eastern tropical Pacific in early summer. In the following year, the transition to La Niña begins in austral fall and peaks in the following summer (Timmermann et al., 2018). La Niña conditions are typically of lesser magnitudes compared to El Niño and can persist up to several years. The seasonal phase lock leads to an austral fall predictability barrier of ENSO development (Levine and McPhaden, 2015).

Individual El Niño events can vary considerably regarding their spatial and temporal evolution (Ashok et al., 2007; Kug et al., 2009; Capotondi et al., 2015). A common distinction between eastern Pacific El Niño events and central Pacific El Niño events, sometimes referred to as El Niño modoki (Ashok et al., 2007), is made. Superposition of these different flavors, which are also associated with respective patterns from the two leading Empirical Orthogonal Functions (EOFs) of tropical Pacific SST anomalies, results in a continuous spectrum of ENSO diversity (Timmermann et al., 2018). Through teleconnections, these ENSO flavors can have different implications for remote regions (Ashok et al., 2007; Tedeschi et al., 2013). However, due to short historical records, it is often difficult to distinguish different modes of this diversity to identify varying implications on remote regions (Timmermann et al., 2018).

The intensity and occurrence frequency of central Pacific El Niño events has been increasing since the 1990s (Lee and McPhaden, 2010). This increase has been attributed to a decadal oscillation of ENSO flavor variability related to low frequency weather noise (Newman et al., 2011), which is consistent with model simulations without external forcing (Wittenberg et al., 2014). However, the question whether tropical Pacific decadal variability might be attributable to fundamental decadal scale processes has not been fully resolved yet (Newman et al., 2016; Timmermann et al., 2018) partly due to missing long-term data records (Timmermann et al., 2018).

Low frequency ENSO variability is believed to be one of multiple drivers for decadal variability in the North Pacific (Newman et al., 2016). For the northern Pacific, the time series of the SST pattern which explains the most temporal variability of this basin is termed the PDO (Fig. 2.3). It is defined as the variation of the principle component of the leading EOF determined from SST anomalies (Mantua et al., 1997; Zhang et al., 1997). Different phases of this climate mode are associated with precipitation and temperature anomalies across various

regions on the globe (Mantua and Hare, 2002). Furthermore, the PDO modulates the way ENSO impacts various regions (Wang et al., 2014).

While the pattern resembles the ENSO pattern, the PDO acts on longer time scales, such that since the beginning of the satellite era (1980s) only one phase shift occurred. A prevailing warm phase with enhanced SSTs at the eastern margins of the Pacific ended in the late 1990s. It was followed by the onset of a cold phase which lasted at least until 2015 and may recently be shifting back to the warm phase (Newman et al., 2016). While the ENSO pattern reveals a marked amplitude maximum for the tropical Pacific, the PDO is characterized by a broader pattern extending into higher latitudes (Fig. 2.3).

The mechanisms which are currently considered to be driving forces behind the PDO are revisited by Newman et al. (2016). In this work, the PDO is considered as a synthesis of multiple processes comprising stochastic atmospheric forcing, teleconnections from the tropics and midlatitude ocean dynamics.

The stochastic forcing is introduced by variation of the Aleutian Low. A stronger Aleutian Low enhances advection of colder and drier air on its western side and drives stronger westerlies over the central North Pacific and stronger southerlies along the west coast of North America. This leads to stronger heat fluxes between ocean and atmosphere at the central North Pacific and enhanced Ekman transport feedback of the upper ocean layer (Alexander and Scott, 2008), which, in turn, results in colder SSTs at the central North Pacific and warmer SSTs at the eastern North Pacific. Therefore, an intensified Aleutian Low potentially yields a positive phase of the PDO while the mechanism is reversed for a weaker Aleutian Low. Lagged correlations reveal that the variability of the Aleutian Low leads the PDO implying its driving force (Newman et al., 2016).

Teleconnections between the tropics and the North Pacific are threefold (Newman et al., 2016). Via a so called atmospheric bridge (Alexander et al., 2002; Liu and Alexander, 2007), a positive or negative ENSO phase is linked to a stronger or weaker Aleutian Low and, thus, leads to a positive or negative impact on the PDO, respectively (Alexander et al., 2002; Alexander and Scott, 2008). Furthermore, eastward propagating equatorial Kelvin waves associated with the onset of El Niño are redirected polewards at the eastern Pacific boundary. On this poleward course, they create sea level and SST anomalies and excite oceanic Rossby waves (Johnson and O'Brien, 1990; Clarke, 1994). That way, such Kelvin waves have a direct impact on the ocean state of the North Pacific. Finally, decadal scale ENSO variability, i. e. low pass filtered SST anomalies, results in a primary EOF which resembles the PDO pattern (Zhang et al., 1997). Therefore, ENSO is considered to be linked to the PDO and to lead it according to lagged correlations (Newman et al., 2016). Observations reveal that decadal scale variations in the tropical

oceans are coherent with variations in the North Pacific, indicating a coupling between these regions (Deser et al., 2004).

Different Ocean dynamics affect the nature of the PDO in terms of frequency and seasonal characteristics. For instance, a deeper mixed layer in boreal winter results in downward mixing of potential SST anomalies to depths which are decoupled from the surface during the summer season. A shallower mixed layer during summer embodies a lower heat inertia so that it responds faster to atmospheric processes resulting in greater SST variability. However, when the mixed layer depth increases during the subsequent winter, the conserved SST anomalies reemerge to the surface. This so called reemergence mechanism (Alexander and Deser, 1995) leads to higher autocorrelation of boreal winter and spring season with following winter and spring seasons compared to summer and fall (Newman et al., 2016).

Another form of ocean dynamics which affect the PDO is triggered by changes in the strength of the Aleutian Low. The resulting anomalous wind stress curl excites westward propagating Rossby waves, which, in turn, leads to a dynamic adjustment of the upper ocean gyre circulation (Qiu and Chen, 2005; Taguchi et al., 2007). These adjustments manifest in shifts of the SST anomaly pattern in the subarctic frontal zone of the western North Pacific. These SST shifts cause changes of heat and moisture exchange between ocean and atmosphere in this region (Tanimoto et al., 2003; Taguchi et al., 2012). The mediating Rossby waves take about 3–10 years to cross the Pacific basin (Sasaki and Schneider, 2011; Sasaki et al., 2013) feeding the interannual characteristic of the PDO (Newman et al., 2016). The way in which the resulting SST changes feed back onto the PDO remains poorly understood (Newman et al., 2016).

2.3.2 *Regional impacts*

There are several studies focussing on or including an analysis of the interannual atmospheric variability in the Atacama Desert. Such studies typically utilize reanalysis data, radiosoundings, in-situ precipitation measurements and gridded precipitation data sets derived from gauge stations. Continuous data from precipitation gauges are limited to the outer edge of the Atacama Desert, i. e. stations at the western coast or at the slopes of the Andes and at the Altiplano at the eastern side (Houston and Hartley, 2003; Houston, 2006; Schulz et al., 2012; Valdés-Pineda et al., 2016). These studies mostly reveal a link between ENSO and interannual variability of precipitation. This link is shaped differently depending on region and season.

At the Bolivian Altiplano in the northeast of the Atacama core desert region, the wet season, i. e. austral summer, coincides with the typical peak season of El Niño events (Chapter 2.3.1). Thus, summer precipitation correlates to the ENSO state: While El Niño is associated

with drier conditions, La Niña is associated with wetter conditions (e. g. Vuille, 1999; Garreaud and Aceituno, 2001; Garreaud et al., 2003; Houston, 2006; Garreaud et al., 2009; Marín and Barrett, 2017; Canedo-Rosso et al., 2019).

During austral summer, i. e. the wet season, precipitation across the Altiplano is related to moist easterly episodes which result from a southward shift of the Bolivian High (Vuille, 1999; Garreaud et al., 2003). While the seasonal mean zonal wind is not related to intensity of single precipitation events, it shows an almost linear relationship to the number of days with precipitation and, thus, to the seasonally accumulated precipitation (Garreaud and Aceituno, 2001).

The frequency of these moist easterlies is modulated by ENSO in the following way. El Niño related SST anomalies in the eastern Pacific lead to an expansion of the troposphere. These positive geopotential height anomalies in upper levels result in enhanced westerly winds so that less wet episodes can occur across the Altiplano (Vuille, 1999; Garreaud and Aceituno, 2001; Garreaud et al., 2003). On the other hand, La Niña is associated with negative SST anomalies which lead to stronger subsidence over the southeast Pacific. This allows more frequent wet easterly episodes (Vuille, 1999).

A similar summer seasonal relationship between zonal wind and moisture supply in terms of Integrated Water Vapor (IWV) has been observed farther south at the Chajnantor Plateau at a height of about 4800 m (Marín and Barrett, 2017). This region is characterized by extremely dry conditions suitable for very demanding astronomical observations (Giovanelli et al., 2001). They are carried out, for instance, at the Atacama Large Millimeter Array (ALMA Brown et al., 2004) and the Atacama Pathfinder Experiment (APEX Güsten et al., 2006).

Even though the fundamental link between ENSO phase and precipitation has been identified, not every El Niño or La Niña is associated with an anomalously dry or wet summer, respectively, resulting in low correlations. This can be attributed to the diversity of the ENSO phases resulting in different meridional positions of the zonal wind anomalies (Garreaud and Aceituno, 2001). Another reason might be the different temporal offsets of the ENSO peak and the Altiplanic wet season (Garreaud et al., 2003).

While ENSO influence is not significant for austral winter at the Altiplano (Vuille, 1999), the situation is opposite at the western coast of the Atacama Desert. El Niño conditions appear to enhance precipitation for both winter and summer (Houston, 2006; Vargas et al., 2006; Schulz et al., 2012). During austral winter, the enhanced precipitation is typically associated with a developing El Niño, featuring a blocking situation in midlatitudes. The latter pattern shifts the storm tracks farther north, resulting in enhanced precipitation of the coastal and southern Atacama Desert (Vargas et al., 2006). The relationship with La Niña is generally weaker and of opposite sign (Houston, 2006).

The recurring period of wet winter conditions at the coastal stations is about 12 years, whereas El Niño occur every 3.3 years on average (Houston, 2006). This indicates that the ENSO state might favor certain conditions but does not directly cause them. Therefore, specific synoptic conditions have to coincide for anomalous precipitation (Houston, 2006). Furthermore, the ENSO signal may be amplified or damped by the PDO (Andreoli and Kayano, 2005; Kayano and Andreoli, 2007; Valdés-Pineda et al., 2018).

Given its decadal time scale, the PDO received less attention in previous studies on impacts on the Atacama Desert, possibly due to the requirement of very long data records. Valdés-Pineda et al. (2018) retrieved a suitable precipitation record from the gauge station at Copiapó located at the southern Atacama coast (27°S). They reveal a strong phase agreement between decadal scale precipitation variability and PDO. Furthermore, for another station at Antofagasta (23.6°S), a long time series of annual precipitation (1904–2000) reveals that anomalously wet years cluster together during positive PDO phases (Vargas et al., 2006, their Fig. 2).

The previous studies show the importance of large-scale climate indices for the moisture supply to the Atacama Desert. Still, the studies are mostly limited to either the Altiplano at the northeast or to very few coastal stations. Attempts have been made to interpolate the composite precipitation signal using a kriging algorithm to fill the gaps in between (Houston, 2006). However, only three respective positive and negative ENSO phases were incorporated. While the mechanism between large-scale tropical variability and precipitation anomalies in the Bolivian Altiplano has been recognized, it remains unclear whether the same mechanism guides the ENSO response at other inland locations within the Atacama Desert. For instance, it is not known how far west moisture can be transported by recurring easterly episodes during the summer above the Altiplano. Composite vertical profiles of specific humidity based on radiosoundings at the coastal city Antofagasta reveal significantly higher values at higher tropospheric levels for wet easterly episodes over the Altiplano. Nevertheless, the analogue composite for La Niña conditions does not reveal a significant moistening (Vuille, 1999).

For the PDO, a positive interference with the ENSO signal has been indicated by studies which investigated South America on a continental scale (Andreoli and Kayano, 2005; Kayano and Andreoli, 2007; Garreaud et al., 2009). However, only very few results are provided with specific focus on the Atacama Desert. Using a wavelet analysis of precipitation measurements at the Altiplano for a 68-year period (1948–2016), no clear relationship to the PDO could be identified (Canedo-Rosso et al., 2019). Longer station records were considered for very few coastal stations, revealing potential implications of this climate mode for the Atacama (Vargas et al., 2006; Schulz et al., 2012; Valdés-Pineda

et al., 2018). This indicates that century-long data records might be required to obtain significant results.

2.4 SYNOPTIC DRIVER

The climate of the Atacama Desert is characterized by stable conditions with prevailing subsidence of warm dry air (Chapter 2.1). Possible disturbances are troughs and cut-off lows which emanate northward from higher latitudes across the Pacific and coastal region. Furthermore, thunderstorm activity and precipitation episodes occur across the Altiplano during austral summer which may reach over the western cordillera of the Andes into the Atacama Desert. These synoptic-scale events are associated with a southward displacement of the Bolivian High. Another disturbance could be posed by Atmospheric Rivers (ARs). Their role has not yet been investigated for this region.

Cut-off lows are closed cold low pressure systems which emanate from a trough within the basic westerly flow at midlatitudes. The cut-off process is typically associated with Rossby wave breaking and a split jet structure (Peters and Waugh, 2003; Ndarana and Waugh, 2010; Reyers and Shao, 2019). Within a region slightly south of the Atacama Desert between 30°S and 40°S, cut-off lows can occur all year round with an average of seven events per annum with a seasonal preference of fall, winter and spring (Pizarro and Montecinos, 1999). These cut-off lows typically produce rainfall at the associated inland region. Exceptions are very weak cut-off lows which are warmer and characterized by a weaker pressure gradient (Pizarro and Montecinos, 1999).

Further north in direct connection to the Atacama Desert, extended winter time (April–September) cut-off lows have been attributed to enhanced inland moisture and a modulated wind regime (Reyers and Shao, 2019). More details on these regionally occurring cut-off lows including their formation mechanisms and differences to the last glacial maximum are provided by Reyers and Shao (2019).

The March 2015 Atacama flood (Bozkurt et al., 2016) was in fact accompanied by an atmospheric river according to the AR catalogue by Guan and Waliser (2015).

The importance of cut-off lows is also displayed by an episodes of extreme precipitation which occurred in March 2015. For this case, heavy rain was linked to positive SST anomalies in the eastern Pacific and the presence of a cut-off low anomalously north near the Atacama coast (Bozkurt et al., 2016).

Cut-off lows and troughs have been associated with enhanced diurnal northwesterly winds and calm nocturnal conditions for a measurement site within the central Atacama region (Jacques-Coper et al., 2015). In contrast, undisturbed conditions favor easterly winds at night (Jacques-Coper et al., 2015). Enhanced IWV at the Altiplano has also been related to northwesterly midtropospheric flow resulting from troughing in particular for austral winter (Marín and Barrett, 2017).

During the austral summer, precipitation occurs episodically across the Altiplano. These wet episodes are related to a southward shift of the Bolivian High and have a typical duration of a few days (Garreaud et al., 2003) and, thus, can be considered synoptic scale events. Momentum of these upper level easterlies is mixed downward on the eastern side of the Andes so that the upslope flow of a very moist boundary layer air of the interior continent is intensified. This moisture advection fuels the Altiplano with additional latent heat, allowing deep convection and intense precipitation (Garreaud et al., 2003) The storm activity is mostly limited to the Altiplano and declines rapidly at the western slopes of the Andean mountain range.

First termed "tropospheric rivers" (Newell et al., 1992), filamentary structures of enhanced water vapor in the troposphere have been identified and associated with major moisture and heat transport from the tropics towards higher latitudes. The term AR was introduced later on by Zhu and Newell (1998), who pioneered identification algorithms for these objects. Nowadays, ARs are defined as a "long, narrow, and transient corridor of strong horizontal water vapor transport that is typically associated with a low-level jet stream ahead of the cold front of an extratropical cyclone" (Ralph et al., 2018). They constitute an emerging topic to the hydrological and meteorological science and engineering community (Wilson et al., 2020). Upon landfall at the eastern boundaries of the oceans, ARs can trigger extreme precipitation events associated with floods, strong winds as well as essential water supply (e. g. Newell et al., 1992; Zhu and Newell, 1998; Neiman et al., 2008; Ralph and Dettinger, 2011; Ramos et al., 2016; Waliser and Guan, 2017; Paltan et al., 2017; Blamey et al., 2018; Nash et al., 2018; Viale et al., 2018; Rauber et al., 2020).

From a global distribution of AR landfall frequency, it is evident that these systems are mainly impacting coastal regions of the midlatitudes with decreasing frequencies towards the equator (Guan and Waliser, 2015). Specifically for the South American continent, ARs occur mainly between 30°S and 60°S with peak frequencies between 45°S and 50°S (Viale et al., 2018). However, indication is given that a few landfalls happen even as far north as the coast of the Atacama Desert (Guan and Waliser, 2015, cf. their Fig. 8a). Even though their frequency is comparably low in this region, they may be related to a significant fraction of the overall very rare precipitation. Still, the role of ARs for the Atacama Desert has not been assessed yet.

3 OVERVIEW OF THE STUDIES

The goal of this thesis is to provide a comprehensive view on the moisture supply to the Atacama Desert. In this chapter, the overarching hypotheses are raised together with general goals. Thereafter, the corresponding studies carried out to achieve these goals are introduced and the respective objectives are specified in more detail.

3.1 AIMS AND HYPOTHESES

From the review of previous studies regarding the climatic setting of the Atacama Desert and related atmospheric processes, it is clear that different sources of moisture are at play for different parts of this region. At the northeastern edge of the Atacama, episodes with easterly winds are associated with the prevailing summer precipitation over the Altiplano (e. g. Garreaud and Aceituno, 2001; Garreaud et al., 2009). The rain producing moisture originates in the interior of the South American continent, the Amazon Basin (Garreaud et al., 2003). On the other end of the Atacama Desert, coastal vegetation and fog ecosystems rely on the advection of the Pacific stratocumulus (e. g. Cereceda et al., 2002; Farías et al., 2005; Pinto et al., 2006; Cereceda et al., 2008b; Lobos Roco et al., 2018; Lehnert et al., 2018b; Río et al., 2018). For the central Atacama between Altiplano and coast, the main moisture source is still unclear.

Furthermore, various studies reveal that large-scale patterns, such as ENSO or the PDO, exert different effects depending on region and season. While a negative relationship is reported for the Altiplano, where decreased summer precipitation is associated with the positive ENSO phase (El Niño) (e. g. Vuille, 1999; Garreaud and Aceituno, 2001; Garreaud et al., 2003; Canedo-Rosso et al., 2019), the opposite is observed at the southwestern coast. There, increased precipitation amounts are related to the positive ENSO phase (e. g. Vuille, 1999; Houston, 2006; Vargas et al., 2006; Schulz et al., 2012). The ENSO signal appears to be amplified by the PDO whenever both climate indices are in the same phase (Garreaud et al., 2009; Andreoli and Kayano, 2005; Kayano and Andreoli, 2007). Moreover, the positive PDO phase is related to overall wetter decades at the coast (Vargas et al., 2006; Schulz et al., 2012; Valdés-Pineda et al., 2018). To sum up, the seasonal influences by these large-scale patterns appear to be opposite between the southwest and northeast of the Atacama Desert. The border between these two poles remains undetermined.

Large-scale climate modes do not directly cause water supply, e. g. through precipitation, but might favor related synoptic conditions (Houston, 2006). An example of such synoptic features are cut-off lows which originate from higher latitudes over the Pacific and are associated with enhanced moisture supply in the Atacama Desert (Bozkurt et al., 2016; Pizarro and Montecinos, 1999; Reyers and Shao, 2019). They weaken the inversion strength and may lift the inversion base height, which allows the moist maritime boundary layer to penetrate farther inland. However, cut-off lows are typically located in height levels above the boundary layer. Therefore, they steer air masses from the northwest through the free troposphere towards the Atacama Desert. These circumstances suggest that the origin of the related moisture might be the remote Pacific. However, this assumption has not been verified yet by the scientific community.

Another synoptic phenomenon which might be relevant for the Atacama Desert could be ARs even tough their role has not been addressed properly for this region. According to Guan and Waliser (2015), a few landfalls are detected at the coast of northern Chile. Yet, AR-related precipitation fraction as well as associated mechanisms remain unknown.

In order to close existing gaps regarding moisture sources and the role of large-scale and synoptic-scale driven water supply in the Atacama Desert, this thesis raises two overarching hypotheses:

Overarching hypotheses of this work.

1. The main source of moisture supplied to the Atacama Desert is the Pacific ocean.

2. The variability of moisture availability in this region is driven by synoptic-scale rather than large-scale patterns.

These hypotheses provide the main guidance for the arrangement of the analyses presented in Part II. They comprise three main studies and one additional analysis. The way they are related to one another and to the moisture cycle in the Atacama Desert is depicted in Figure 3.1.

One of major forms of moisture supply to fog-dependent ecosystems in the Atacama are stratocumulus clouds, which provide moisture from the Pacific ocean by intercepting with the coastal orography upon inland penetration. In order to identify the locations influenced by stratocumulus, cloud top and base heights should be determined. One of the most suitable ways to determine these height on a region-wide scale is to use satellite observations. Whereas satellite-based products are readily available for cloud top height, this is not the case for cloud base height. Therefore, this thesis aims at deriving a suitable satellite-based cloud base height retrieval method, which is introduced in Chapter 3.2.

Speaking of stratocumulus penetrating inland, it should be noted that stratocumulus clouds necessarily mix with dry inland air. This

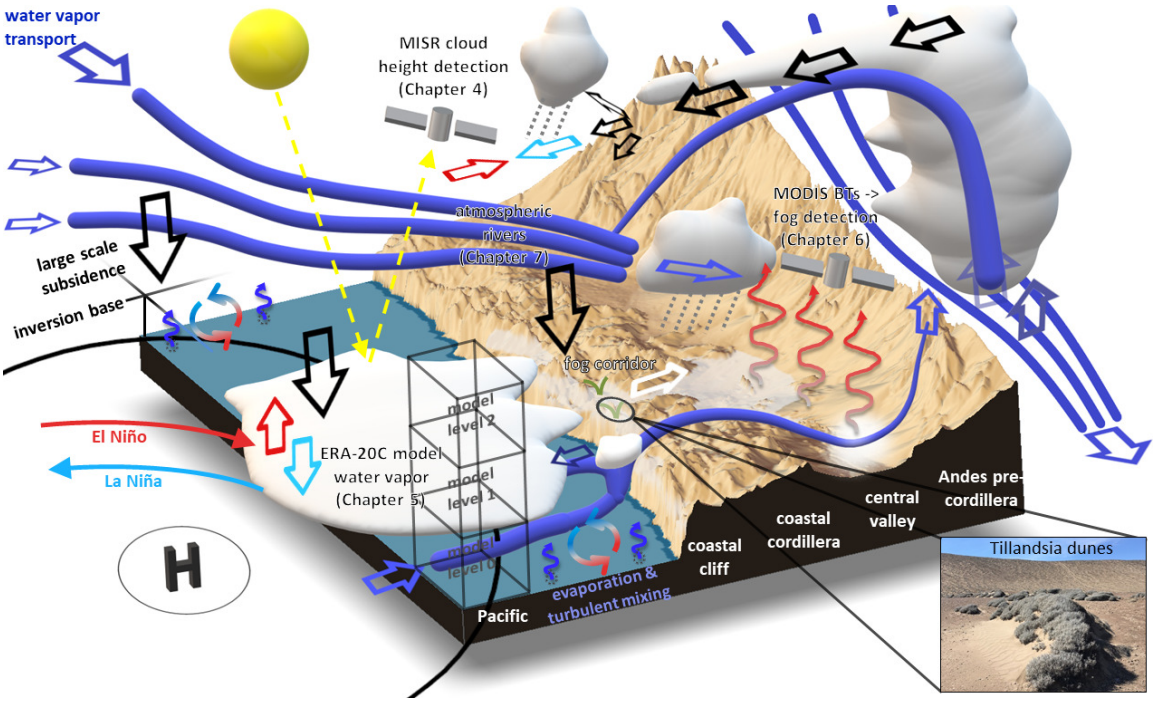

Figure 3.1: Schematic overview of atmospheric water cycle elements for the Atacama Desert which are addressed in this thesis. The studies are explicitly annotated with respective Chapter numbers. Bottom right: Photograph of a *Tillandsia* dune at Cerro Oyarbide taken by Andrea Jaeschke (Jaeschke et al., 2019, Fig. 1c, used under the Creative Commons CC-BY-NC- ND license).

leads to an evaporation of the cloud liquid water, resulting in water vapor enriched air. Aside from the advection of evaporating stratocumulus, water vapor can be transported to the core desert region directly from the Pacific. But also other sources of water vapor, such as ground water evaporation or the interior of the South American continent, might be at play.

Whether easterly or westerly source regions dominate the moisture supply is closely linked to large-scale patterns, such as ENSO and PDO. Therefore, this thesis includes a study which investigates the influence of these large-scale climate modes and which shows how potential source regions for water vapor can be attributed to specific phases of large-scale patterns. Furthermore, some indications of the seasonality of this dependance are discussed. An introduction to this study is given in Chapter 3.3.

Another form of water supply to the Atacama is fog. While the inland penetration of maritime stratocumulus constitutes one source of fog enabling moisture, results from the study introduced in Chapter 3.3 indicate that the continental interior might be another source of water vapor contributing to fog formation depending on the ENSO state. To determine whether moisture sources other than the Pacific

stratocumulus and maritime water vapor are involved in fog formation, it is necessary to identify the spatial distribution of fog over time.

To achieve these goals, a novel satellite-based fog retrieval method is developed. Furthermore, this new algorithm enables to determine the influence of large-scale and synoptic patterns on fog occurence frequency. The corresponding study is introduced in Chapter 3.4.

Besides fog and water vapor, precipitation is a third form of water supply to the Atacama Desert. Precipitation events are associated with cut-off lows (e. g. Bozkurt et al., 2016; Reyers and Shao, 2019) and by an episodically southward displaced Bolivian High during austral summer (Garreaud et al., 2003). To add a so far unattended synoptic feature to this range of phenomena, this thesis aims to determine the role of ARs for the moisture supply to the Atacama Desert. Furthermore, the related origin of the precipitation water is assessed. A brief overview of the corresponding study is provided in Chapter 3.5.

3.2 CLOUD HEIGHTS

Clouds heights over the near coastal Pacific determine where fog can be formed at the cloud-intercepting coastal cliff and mountain range. Therefore, monitoring the cloud top and base heights allows to estimate potential habitats for plants which adapted to water and nutrition supply by fog (e. g. *Tillandsia*; Rundel et al., 1997; Pinto et al., 2006; Westbeld et al., 2009; González et al., 2011). Very few studies exist regarding cloud heights of the near coastal maritime stratocumulus. These are limited to individual coastal sites based on human observers at three sites (Muñoz et al., 2016) and indirectly by determining fog water supply along a height transect at Alto Patache research station (Cereceda et al., 2008b).

Quantifying the spatio-temporal cloud height distribution and variability region-wide would enable relating these stratocumulus properties to biological and geological processes, such as desert soil formation (Voigt et al., 2020), plant growth within coastal fog oases (Muñoz-Schick et al., 2001) or the *Tillandsia* nitrogen cycle within the coastal Atacama Desert (Latorre et al., 2011; Díaz et al., 2016). Furthermore, the heights of the coastal clouds determine the regions where fog water can be collected most efficiently. Fog water collection is of great social and economic impact in this region (e. g. Schemenauer and Cereceda, 1994b; Osses et al., 2000; Larraín et al., 2002).

To achieve region-wide coverage, satellite remote sensing of cloud heights would be beneficial. While several satellite-based data products are available for the cloud top height, no operational retrieval algorithm for the cloud base height exists to date. In Chapter 4, a novel satellite-based cloud base height retrieval method is developed and applied to the southeast Pacific stratocumulus deck.

3.3 WATER VAPOR VARIABILITY

The influence of low frequency climate modes, such as ENSO or the PDO, on the water supply has almost solely been assessed for the Altiplano (e. g. Vuille, 1999; Garreaud and Aceituno, 2001; Garreaud et al., 2003; Canedo-Rosso et al., 2019) and very few coastal station records (Vuille, 1999; Houston, 2006; Vargas et al., 2006; Schulz et al., 2012; Valdés-Pineda et al., 2018). Few region-wide studies are based on gridded precipitation data sets or interpolations derived from sparsely distributed gauge measurements (Vuille, 1999; Andreoli and Kayano, 2005; Houston, 2006; Kayano and Andreoli, 2007; Garreaud et al., 2009).

To determine the role of large-scale climate modes for the Atacama Desert in more detail, a long-term data record is needed which can adequately represent decadal-scale variability as expected, for example, from the PDO. While satellite-based observations provide sufficient spatial coverage, available records are too short, since the satellite era only began in the 1980s. Spatially and temporally consistent data sets suitable for the assessment of longer time scales are provided by reanalysis data, such as the European Centre for Medium-Ranged Weather Forecast's (ECMWF) twentieth century reanalysis (ERA-20C; Poli et al., 2016). After a validation of ERA-20C's moisture representation, the interannual variability of moisture supply is determined and linkages to large-scale drivers are assessed. Besides, the length of the data record allows to distinguish between long-term trends and decadal scale oscillations.

For the validation, a time series analysis is carried out comparing satellite-based IWV retrievals to the reanalysis. As a validated century-long IWV record for the Atacama Desert, ERA-20C can be considered for other applications within this region. For instance, long-term IWV trend analysis is crucial for the development of astronomical observations which are routinely carried out at observatories in the study area (Kerber et al., 2014; Otarola et al., 2019). The validation and application of the 20th century IWV is presented in Chapter 5.

3.4 FOG

Near coastal inland fog (Fig. 3.2) is the life vein for many local plant species (e. g. Muñoz-Schick et al., 2001; Pinto et al., 2006) and is of great social and economic importance (Schemenauer and Cereceda, 1994b; Osses et al., 2000; Larraín et al., 2002) as mentioned above. The abundance of fog drives the state of the coastal vegetation, such as the isotopic composition of *Tillandsia* (Latorre et al., 2011; Jaeschke et al., 2019). Furthermore, the amount of organic traces (Mörchen et al., 2019) and the activity of microbial communities (Jones et al., 2018) depend

Figure 3.2: Photograph of retreating inland fog taken by a trail camera on 29 June 2017 at 8 CLT (about 30 minutes after sunrise) at weather station 13 – Cerros de Calate, Chile facing the Pacific Ocean (May and Hoffmeister, 2018).

on water sources, such as fog. However, a region-wide fog climatology is still missing.

Regional studies of fog occurrence and water supply are typically based on either in-situ fog water collection at few individual sites (e. g. Cereceda et al., 2002; Osses et al., 2005; Cereceda et al., 2008b; Lobos Roco et al., 2018; Río et al., 2018) or on satellite remote sensing (e. g. Farías et al., 2005; Cereceda et al., 2008b; Lehnert et al., 2018b). While studies on fog water collection typically suffer from coarse temporal resolution, i. e. weekly to monthly measurements of water accumulation (Río et al., 2018), to date satellite-based studies are limited to short time periods or small regions.

Moreover, the applied satellite-based fog detection techniques do not distinguish between low clouds and fog. Only few attempts have been made to carry out a thorough ground validation. These approaches involve extensive field work in remote hardly accessible locations to create ground-based reference data and are limited in space and time (Cereceda et al., 2002; McIntyre et al., 2005; Osses et al., 2005).

The previously introduced study (Chapter 3.3 Böhm et al., 2020a, Chapter 5) demonstrates that ENSO modulates IWV across the Atacama Desert and over the bordering Pacific depending on region and season. In particular for the austral summer season, the observed IWV modulation can be explained by a link between ENSO and the frequency of easterlies which advect moist air from the interior continent across the Altiplano. The way this mechanism might impact the variability of fog occurrence has not been assessed yet. In the central parts of the Atacama Desert, where radiation fog might dominate compared to other fog types (Cereceda et al., 2002; Westbeld et al., 2009), enhanced IWV may result in enhanced nocturnal condensation. However, in order

to assess a potential relationship, a fog detection method is required which can be applied region-wide with daily resolution.

Such a detection method would also allow to relate the variability of fog occurrence to specific synoptic conditions. Thereby, potential moisture sources could be identified more accurately. A robust relation between the maritime stratocumulus and fog occurrence clearly establishes the Pacific as the moisture source for the coastal desert (e. g. Farías et al., 2005; Cereceda et al., 2008b; Río et al., 2018). Farther inland, where radiation fog might be more dominant compared to the stratocumulus advection, the source of moisture is less clear.

In order to build a new fog detection method, appropriate data are required for validation. The study presented within this thesis (Chaper 6) utilizes in-situ observations from a network of climate stations (Hoffmeister, 2017b; Schween et al., 2020) to derive a ground-based reference data set. These stations have been installed recently by the CRC 1211. Furthermore, Moderate Resolution Imaging Spectroradiometer (MODIS) brightness temperatures from thermal emissive bands are applied to train a neural network to detect fog. Thereby, all spectral information which is available from the spaceborne instrument can be exploited by the detection algorithm.

3.5 ATMOSPHERIC RIVERS

The studies introduced above focus on cloud and fog water supply, which become dominant moisture sources due to extremely low precipitation rates. Additionally, water vapor is studied as a representative variable which allows to investigate longer time series. Nevertheless, precipitation events are the dominant driver for fluvial alterations of the landscape through land slides and debris flow (e. g. Vargas et al., 2006; Haug et al., 2010; Jordan et al., 2019; Walk et al., 2020), which are manifested and preserved in climate archives (Ritter et al., 2019; Diederich et al., 2020; Bartz et al., 2020). To interpret such archives, it is necessary to identify synoptic drivers for precipitation and understand the underlying mechanisms.

While synoptic drivers, such as cut-off lows and troughs, have been discussed in previous studies (e. g. Montecinos and Aceituno, 2003; Vargas et al., 2006; Jacques-Coper et al., 2015; Bozkurt et al., 2016; Marín and Barrett, 2017; Reyers and Shao, 2019), the role of ARs for the Atacama Desert has not been investigated so far. Intriguingly, a study revealing AR landfall frequencies on a global scale indicates that a few of theses systems reach the coast of northern Chile each year (Guan and Waliser, 2015). As precipitation events are extremely rare in this environment, a few ARs may already account for a substantial fraction of the total rainfall. The presented study (Chapter 7) quantifies this impact on a region-wide scale for the Atacama Desert using a state-of-the-art AR catalog (Guan and Waliser, 2015; Guan et al., 2018)

together with a previously validated precipitation data set from a highly resolved long-term Weather Research and Forecasting Model (WRF) simulation (Reyers, 2018; Reyers et al., 2020).

As the Atacama Desert is characterized by strong topographical gradients, the vertical structure of regionally occurring ARs is investigated by means of a case study and a composite analysis. This allows a better understanding of the relevant processes, such as moisture advection and cloud formation. Furthermore, the role of the MBL and potential orographic effects are discussed. Ultimately, potential pathways of the moisture can be estimated. To further identify the origin of the moist air mass, back trajectories are calculated as part of the case study.

Part II

METHOD DEVELOPMENT AND ANALYSIS

4

CLOUD HEIGHTS

CLOUD BASE HEIGHT RETRIEVAL FROM MULTI-ANGLE SATELLITE DATA

Böhm, C., O. Sourdeval, J. Mülmenstädt, J. Quaas, and S. Crewell (2019). "Cloud base height retrieval from multi-angle satellite data." In: *Atmospheric Measurement Techniques* 12.3, pp. 1841–1860. DOI: 10.5194/amt-12-1841-2019.

ABSTRACT. Clouds are a key modulator of the Earth energy budget at the top of the atmosphere and at the surface. While the cloud top height is operationally retrieved with global coverage, only few methods have been proposed to determine cloud base heights (z_{base}) from satellite measurements. This study presents a new approach to retrieve cloud base heights using the Multi-angle Imaging SpectroRadiometer (MISR) on the Terra satellite. It can be applied if some cloud gaps occur within the chosen distance of typically 10 km. The MISR cloud base height (MIBase) algorithm then determines z_{base} from the ensemble of all MISR cloud top heights retrieved at a 1.1-km horizontal resolution in this area. MIBase is first calibrated using one year of ceilometer data from more than 1500 sites within the continental United States of America. The 15th percentile of the cloud top height distribution within a circular area of 10 km radius provides the best agreement with the ground-based data. The thorough evaluation of the MIBase product z_{base} with further ceilometer data yields a correlation coefficient of about 0.66 demonstrating the feasibility of this approach to retrieve z_{base}. The impacts of the cloud scene structure and macrophysical cloud properties are discussed. For a three year period, the median z_{base} is generated globally on a $0.25° \times 0.25°$ grid. Even though overcast cloud scenes and high clouds are excluded from the statistics, the median z_{base} retrievals yield plausible results in particular over ocean as well as for seasonal differences. The potential of the full 16 years of MISR data is demonstrated for the southeast Pacific revealing interannual variability in z_{base} in accordance with reanalysis data. The global cloud base data for the three year period (2007–2009) are available at https://doi.org/10.5880/CRC1211DB.19.

4.1 INTRODUCTION

As Boucher et al. (2013) state in the IPCC Assessment Report 5, clouds and aerosols continue to contribute the largest uncertainty to estimates and interpretations of the Earth's changing energy budget. To describe the effect of clouds on the radiation energy budget, the geometric thickness, the vertical location of clouds and, therefore, the cloud base height (z_{base}) are crucial parameters. Furthermore, long term observations of cloud heights would be beneficial to assess the contribution and the response of clouds to climate change. z_{base} is a key parameter for the radiative energy budget at the Earth surface. z_{base} may also have an impact on ecosystems which are supplied with water by the immersion of clouds (Van Beusekom et al., 2017). Aviation is another field which benefits from information on z_{base}.

Various methods to retrieve the z_{base} have been proposed applying different physical concepts, such as active measurements, spectral methods, approaches using an adiabatic cloud model (e.g., Goren et al., 2018), and in-situ measurements.

From the ground, the most accurate and well-established method to derive z_{base} is the backscatter information from a lidar ceilometer, also providing crucial information on visibility for aircraft safety. Thus, ceilometers are employed at airports. Their number has increased in particular in Europe and North America during the past couple of years. A dedicated web page hosted by the Deutscher Wetterdienst shows the distribution of ceilometer stations around the world (http://www.dwd.de/ceilomap, last access: 13 March 2019). Radiosondes provide in-situ measurements of thermodynamic variables. Costa-Surós et al. (2014) compare different methods to infer z_{base} from radiosonde data. For the best of these methods, 67 % of the considered profiles agree with the utilized reference data regarding number of cloud layers and height category (distinguished are low, middle and high). Cloud radar transmits microwave radiation to derive vertical profiles of radar reflectivity. However, this signal strongly depends on the particle size. Therefore, the occurrence of a few drizzle drops can mask cloud base. Measurements with radiosondes and cloud radars are even less common than ceilometers, global coverage cannot be achieved from the ground today.

From space, active measurements are carried out by CALIOP (Cloud Aerosol Lidar with Orthogonal Polarization) on the CALIPSO (Cloud Aerosol Lidar and Infrared Pathfinder Satellite Observations) satellite (Winker et al., 2010). A valid retrieval of the z_{base} can only be ensured if the signal of CALIOP reaches the Earth's surface, which is only possible in case of low optical thickness. Optically thick clouds will lead to attenuation of the signal. The spatial coverage is limited to the narrow laser beam of CALIOP. The CALIOP cloud base determination has been revisited by Mülmenstädt et al. (2018). They developed an algorithm to extrapolate cloud base retrievals for thin clouds into locations where the CALIOP signal is attenuated within a thicker cloud before it reaches the cloud base.

Passive measurements in the near-infrared exploiting spectral information have been proposed by Ferlay et al. (2010). They suggest an approach to infer the cloud vertical extent from multi-angular POLDER (POLarization and Directionality of the Earth's Reflectances) oxygen A-band measurements. As they point out, the penetration depth of photons into a cloud, and, hence, the height of the reflector, depends on the cloud vertical extent and the viewing geometry. Exploiting the different viewing angles provided by POLDER, Desmons et al. (2013) apply this approach to infer the vertical position of clouds. Their comparison to retrievals from the cloud profiling radar on CloudSat and CALIOP shows that this method works best for liquid clouds over ocean with a retrieval bias of 5 m and a standard deviation of the retrieval differences of 964 m. However, this approach has not been carried out operationally yet. Moreover, an estimate of the cloud top

height is required to retrieve the cloud base height from the cloud vertical extent, which introduces additional uncertainty.

Meerkötter and Zinner (2007) suggest a method to derive z_{base} of convective clouds which are not affected by advective motions. An adiabatic cloud model incorporating measurements of cloud optical depth and effective radius is used to calculate the geometric extent of the cloud from the retrieved cloud top height. By introducing a subadiabatic factor, Merk et al. (2016) investigate the adiabatic assumption in more detail. By additionally introducing a factor into the calculations, they account for subadiabaticity due to entrainment of dry air through the cloud edges. As a reference, the cloud vertical extent is derived as the difference between z_{top} (radar) and z_{base} (ceilometer) from ground based measurements. The authors conclude that for their two year data set neither the assumption of an adiabatic cloud nor the assumption of a temporally constant subadiabatic factor is fulfilled.

Lau et al. (2012) suggest a new approach to determine z_{base} utilizing the Multi-angle Imaging SpectroRadiometer (MISR) on the Terra satellite. For a preliminary case study, they chose the observations from island Graciosa, Azores, Portugal, for which they compared cloud top height (z) retrievals from MISR to collocated and coincidental lidar measurements. Under the assumption that the cloud vertical extent varies horizontally within the cloud, they retrieve z_{base} by identifying the lowest cloud top height in the height profile provided by MISR. The reference cloud base height ($\hat{z}_{base}$) is retrieved from the lidar signal by visual inspection of the backscatter coefficient in a time-height cross section over a period of about five hours. They selected 12 cases which show a promising agreement between MISR and lidar retrievals.

We build on the approach proposed by Lau et al. and develop an automatic retrieval method to derive z_{base} from MISR measurements. Parameters employed in the retrieval scheme are derived from coincident ceilometer measurements over one year in the continental United States of America (USA). The performance of the z_{base} algorithm is demonstrated by an evaluation with ceilometer over a longer time period and the potential for application on the global scale and for longer time series is explored.

The paper is structured as follows. In Section 4.2, the utilized data from MISR and from ceilometers are described. Section 4.3 introduces the new retrieval method along with a case study for illustration. In Section 4.4, the evaluation of the algorithm against the ceilometer measurements is shown and the effect of the cloud vertical extent on the performance of the algorithm is discussed. Section 4.5 includes two applications of the algorithm: the median z_{base} is presented globally for a three-year period, and regionally over the southeast Pacific for a 16-year period. Finally, Section 4.6 concludes the study.

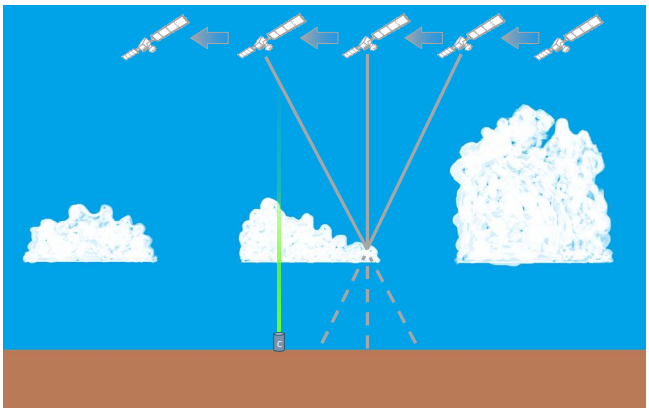

Figure 4.1: Schematic depiction of a cloud field observed from different viewing angles during the satellite overpass. Ceilometers, here represented as a cylindrical box, provide ground-based measurements of cloud base heights which can be used as reference.

4.2 DATA

4.2.1 *MISR cloud product*

MISR is carried on board the Terra satellite and provides sun-synchronous (equatorial overpass at around 10:30 local solar time) global products of cloud properties with a 1.1 km horizontal resolution. With an across-track swath width of 380 km, MISR takes two (poles) to nine (equator) days for repeated observations of the same site. The MISR Level 2TC Cloud Product (MIL2TCSP; Diner, 2012; Moroney and Mueller, 2012; Mueller et al., 2013) is used in this study to provide retrievals of cloud top height z and a stereo-derived cloud mask (SDCM). Three years of global data (2007–2009) are utilized here. The MISR Ancillary Geographic Product (Bull et al., 2011) is additionally used to assign corresponding spatial coordinates and the average scene elevation for each pixel. Here, we give a brief summary on how the operational MISR z product is derived. More in-depth descriptions can be found in Moroney et al. (2002) and in Marchand et al. (2007).

A cloud field is schematically depicted in Fig. 4.1. MISR hosts cameras providing a total of nine viewing angles. Besides the nadir viewing camera (0°), there are four forward and four aftward viewing cameras set up at 26.1°, 45.6°, 60.0° and 70.5° angles, respectively. During an overpass, each camera of MISR records the reflected radiances at its particular viewing angle. A pattern matching routine which compares the radiances recorded at a wavelength of 670 nm identifies equal cloud features in the images of the different viewing angles. Pixels with the least deviation from each other are matched. This way, a detected cloud feature is observed from multiple satellite positions with its respective time and viewing angle. If at least three images can be attributed to the same cloud feature, the cloud motion vector

along with the horizontal and vertical position of the cloud feature can be inferred geometrically. This process is not sensitive to absolute values of the radiances so that this retrieval method is not sensitive to calibration.

The cloud motion vector is determined at a 17.6 km resolution. For each of these coarser grid boxes, the cloud motion vector is then used to determine z at 1.1 km resolution, which is carried out for two camera pairs individually: one pair (FWD) consisting of the nadir and 26.1° forward viewing cameras and the other (AFT) consisting of the nadir and 26.1° aftward viewing cameras. This way, two z values for the same location are available, and the mean of the two values yields the final z. In case only one camera pair provides a valid z, it is taken as the final z at its specific location. To derive the stereo-derived cloud mask, the two individual z values undergo the following comparison. The retrieval of each camera pair is classified as surface or cloud retrieval according to the threshold height h_{min} (Equation 4.1). This is Equation 59 in the Algorithm Theoretical Basis documentation by Mueller et al., 2013, where the threshold height for flat terrain H_{SDCM} is 560 m, H is the terrain height and σ_{h} is the variance of the the terrain height listed in the Ancillary Geographic Product. Within the MISR Level 2TC Cloud Product, the cloud top height and the stereo-derived cloud mask are also provided without wind correction. Here, we use the the wind corrected data sets.

$$h_{\text{min}} = H_{\text{SDCM}} + H + 2\sigma_{\text{h}} \tag{4.1}$$

The use of two camera pairs allows attribution of a confidence level to the retrieved z. If the mean of the two values is above or below the threshold, the pixel will be classified as cloud or surface, respectively. If only one camera pair provides a valid retrieval, it is tested against the threshold and classified accordingly. In case only one camera pair provides a valid retrieval and in case of two valid retrievals which disagree upon their individual classification, the z retrieval is marked low confidence. If two retrievals are available which agree upon their individual classification, the z retrieval is marked high confidence. Any other case leads to a non-retrieval. Table 4.1 summarizes possible combinations of retrievals from the two camera pairs and their corresponding attribution within the stereo-derived cloud mask.

MISR z is given in meters above the World Geodetic System 1984 (WGS 84) surface. To calculate the height above ground level, we subtract the average scene elevation which is provided within the Ancillary Geographic Product for each pixel.

The MISR z product is expected to be superior to z products from other passive instruments. It does not depend on any auxiliary data and it is not sensitive to calibration. Therefore, it is not granted that

Table 4.1: Classification scenarios of MISR retrievals. The cloud height obtained using the nadir and the 26.1° forward viewing camera pair (denoted by FWD) and the cloud height obtained using the nadir and the 26.1° aftward viewing camera pair (AFT) are tested against the threshold height $h_{\min}$ (Equation 4.1) individually and then compared to one another to determine the Stereo-Derived Cloud Mask (SDCM) attribute.

condition	**SDCM attribute**
FWD and AFT above threshold	high confidence cloud
FWD and AFT disagree, mean(FWD, AFT) above threshold	low confidence cloud
only one camera pair, retrieval above threshold	low confidence cloud
FWD and AFT below threshold	high confidence surface
FWD and AFT disagree, mean (FWD, AFT) below threshold	low confidence surface
only one camera pair, retrieval below threshold	low confidence surface

the application of MIBase to z retrieved by techniques other than the geometric approach would yield similar results.

4.2.2 *METAR data*

Aerodrome routine meteorological reports (METAR) (WMO; World Meteorological Organization, 2013) contain weather observations at airports worldwide, including measurements of z_{base}. METARs from airports from the continental USA provide z_{base} determined by the Automated Surface Observing System (ASOS; National Oceanic and Atmospheric Administration, Department of Defense, Federal Aviation Administration, and United States Navy, 1998). ASOS utilizes lidar ceilometers which operate at a wavelength of 0.9 µm and have a vertical range of 12000 ft (≈ 3700 m). Cloud base heights are routinely retrieved by evaluating the vertical gradient of the detected backscatter profile with a temporal resolution of 30 seconds. These individual retrievals are stored in different bins by rounding to the nearest 100 ft (≈ 30 m) for heights between the surface and 5000 ft (≈ 1500 m); to the nearest 200 ft (≈ 60 m) for heights between 5000 ft (≈ 1500 m) and 10000 ft (≈ 3000 m); and to the nearest 500 ft (≈ 150 m) for heights above 10000 ft (≈ 3000 m). If there are more than five bins filled with measurements during a 30 minute period, the cloud heights are clustered into layers until only five cluster remain. Finally, all cluster heights are rounded according to the rules given in Tab. 4.2. The lowest three layers are passed on to the METAR message.

We extract the ceilometer cloud base height $\hat{z}_{\text{base}}$ from METAR data for a total of 1510 ceilometer sites around the continental USA to benefit from the homogeneity of the automated measurements and the

Table 4.2: The ceilometer $\hat{z}_{\text{base}}$ retrievals are rounded to different values depending on their height window according to ASOS User Guide (National Oceanic and Atmospheric Administration, Department of Defense, Federal Aviation Administration, and United States Navy, 1998). The values are originally given in feet and are converted to meters here.

height [ft]	**rounded to nearest value [ft]**	**rounded to nearest value [m]**
< 5000	100	30.5
5000 to 10000	500	152
> 10000	1000	305

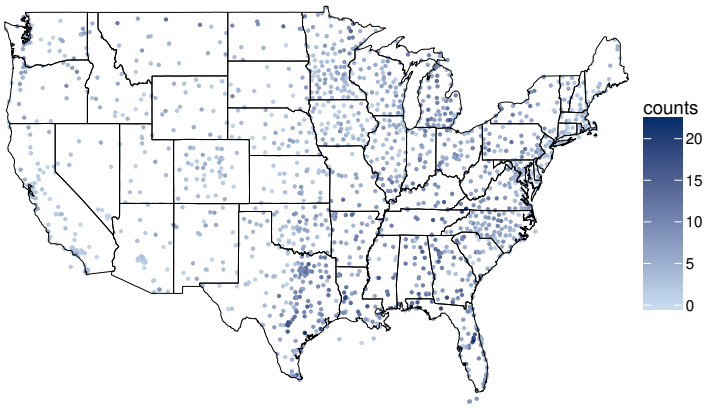

Figure 4.2: Locations of ceilometer stations utilized in this study across the continental USA. Data from these stations for the years 2008 and 2007 are used for the calibration of the z_{base} retrieval algorithm and a subsequent evaluation, respectively. Blue shading indicates the number of valid coincidental retrievals from MISR and ceilometers which have been utilized for the calibration (year 2008) and are within the constraints described in the text.

standardized reporting range. $\hat{z}_{\text{base}}$ serves as reference data to which the z_{base} derived from the satellite cloud heights is compared. First, METAR data from 2008 are used to estimate parameters used in the z_{base} retrieval algorithm to create the MISR Cloud Base height algorithm (MIBase). Second, to validate the "tuned" algorithm, METAR data from 2007 are applied for a statistically independent comparison. For a total of 1510 ceilometer stations, collocated and coincidental satellite-based z_{base} retrievals could be found (see below for exact definition). A distribution of the locations can be seen in Fig. 4.2.

4.3 CLOUD BASE HEIGHT RETRIEVAL

The MISR Cloud Base height retrieval (MIBase) algorithm, which derives z_{base} from the MISR z product, is developed and calibrated

with collocated METAR data for defining the involved parameters and preconditions. The first section of this chapter introduces the retrieval principle on the basis of a case study. By comparison with METAR ceilometer measurements from 2008, parameters used within MIBase are estimated, namely the radius R_c of the MIBase retrieval cell, the minimum number of valid cloud pixel N and the percentile P of the z distribution.

4.3.1 *Method*

We assume that the information on the z_{base} is included in the distribution of the z retrievals from the MISR cloud product for a specific area of limited size. This assumption is valid in a cloud scene with a homogeneous z_{base} and a heterogeneous z similar to the one schematically depicted in Fig. 4.1. Especially at the edge of a cloud where the cloud is thinner, z can serve as a proxy for z_{base}. To ensure that the thinner edge of the cloud is within the observed MIBase retrieval cell, the considered area needs to be large enough and the cloud field needs to be broken. The inherent assumption of a homogeneous z_{base} over a certain area presupposes a horizontally constant lifting condensation level. This is a valid approximation in particular for a well mixed boundary layer or a homogeneous air mass away from the proximity of a frontal zone, where advective motion could introduce temperature or humidity gradients across the horizontal plane.

In order to derive z_{base} from the z product, the following steps, which are outlined in Fig. 4.3, are undertaken. First, a retrieval cell has to be defined. For the comparison to the ceilometer measurements, we consider a circular area with the radius R_c around its midpoint at a ceilometer station. In order to estimate the magnitude of R_c, we consider the following: METAR $\hat{z}_{base}$ retrievals are representative for a time window of 30 minutes. Within this time window and at a typical wind speed of approximately $10\,\mathrm{ms}^{-1}$, a cloud would shift its position about 20 km in the wind direction. Therefore, the magnitude of R_c should be on the order of kilometers. The impact of R_c on the retrieved z_{base} and, therefore, the deviation from the ceilometer $\hat{z}_{base}$ is discussed below. When we apply the algorithm to retrieve a global estimate of z_{base}, we use a regular lat-lon grid of 0.25° (see Section 4.5). This grid size corresponds to a meridional length of the grid boxes of about 28 km and a zonal length ranging between 25 km (25°N) and 18 km (50°N), taking the continental U.S.A. as an example. A greater MIBase cell increases the chance of seeing the thinner part of the cloud. This could lead to a more realistic z_{base} retrieval. In turn, for a smaller MIBase cell the assumption of a homogeneous z_{base} is more realistic.

For each grid cell or circular MIBase cell, the enclosed z retrievals from the MISR cloud product are processed further. MIBase only selects those z retrievals which are marked high confidence cloud

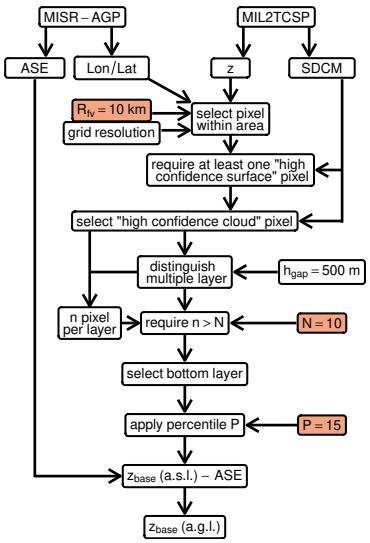

Figure 4.3: Flow chart of the z_{base} retrieval algorithm. MISR's MIL2TCSP cloud product provides z and the Stereo-Derived Cloud Mask (SDCM). MISR's Ancillary Geographic Product (MISR-AGP) provides the Average Scene Elevation (ASE) and the longitude and latitude coordinates for each pixel. Starting from these products, the depicted processing steps are undergone to derive z_{base}. The parameters which have been optimized during the calibration are highlighted in orange.

(hcc) according to the stereo-derived cloud mask. A consideration of retrievals marked low confidence cloud has shown a decrease of the correlation with the ceilometer $\hat{z}_{base}$. An example of a cloud field with z retrievals and the corresponding stereo-derived cloud mask for 21 August 2015 at the International Airport of Atlanta, Georgia, USA, is presented in Fig. 4.4 (left, middle).

For some scenes, the distribution of z reveals extended height ranges with no z retrievals between two or more local maxima. Such cases suggest multi-layer cloud scenes if the apparent gap between adjacent z retrievals is of sufficient size. If such a gap h_{gap} is greater than 500 m, the algorithm distinguishes between the cloud layer above and below the gap (see Fig. 4.4 for the aforementioned example). The value for this threshold has been chosen to be close to the specified accuracy of MISR (560 m). By evaluating different vertical cloud layers individually, a z_{base} retrieval for each layer can be derived. Since for most applications the lowest z_{base} is of interest, the lowest detected cloud layer is processed here. For the comparison with $\hat{z}_{base}$, we restrict ourselves to scenes for which MISR detects only one cloud layer.

The occurrence of a broken cloud field is a basic requirement of MIBase. Therefore, at least one z retrieval marked high confidence surface needs to be within the MIBase cell. A complete cloud cover or a high rate of non-retrievals can prevent this criterion from being met.

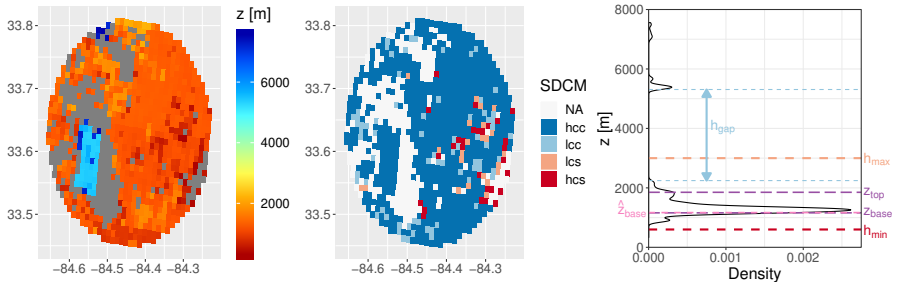

Figure 4.4: MISR observations within a 20 km radius within the vicinity of Atlanta, Georgia, USA (ICAO:KATL) on 21 August 2015 at around 16:30 UTC. Left: z. Middle: Corresponding Stereo Derived Cloud Mask (SDCM) distinguishing non-retrievals (NA), high confidence cloud (hcc), low confidence cloud (lcc), low confidence surface (lcs) and high confidence surface (hcs). Right: Density of z measurements with illustration of certain parameters: height between two layers (h_{gap}) which is the height difference between the highest retrieval of the bottom layer and the lowest retrieval of the next higher layer (dashed blue lines), upper cut-off height (dashed orange) for z_{base} retrievals (h_{max}) which is based on the ceilometer granularity, lower cut-off height (dashed red), which is based on the MISR threshold height to distinguish between cloud and surface retrieval (h_{min}), and the ceilometer retrieval $\hat{z}_{base}$ from 16:52 UTC (dashed pink). z_{top} and z_{base} (dashed purple) are inferred by applying the 15th and 95th percentile to the distribution of z of the lowest cloud layer, respectively. Heights are above sea level.

Both scenarios suggest doubtful z_{base} retrievals. Hence, they are not considered.

For each grid cell or circular cell surrounding the ceilometer station, z_{base} is diagnosed from the height distribution of z using a certain percentile P. In principle, P should be as low as possible. However, as a certain measurement noise is expected and a robust result should be achieved, a choice substantially larger than zero is necessary. Another parameter which describes the distribution of z for each scene is the number of valid z retrievals marked high confidence cloud n. A higher n implies a higher observed cloud cover within the MIBase cell. In order to take a meaningful percentile of the z distribution, a minimum $n > N$ is required. A cloud which is horizontally more extended (higher cloud cover) is more likely to pass over the ceilometer, so that there is a higher chance that both instruments observe the same cloud. Therefore, the deviation of z_{base} from $\hat{z}_{base}$ is expected to decrease for a higher n. The impact of the threshold for N is studied later on.

For certain applications, the cloud vertical extent Δz might be of interest. Therefore, an estimate of the cloud top height z_{top} is required. In principle, $P = 100$ should yield the highest point of the cloud. However, analogously to the retrieval of z_{base}, a certain measurement noise is expected, so that P is not chosen to be the extreme value.

Without further validation, we apply the 95th percentile rather than the median, as we do not want a height which might be representative for the whole area, but rather an estimate of the highest top of the cloud especially for a heterogeneous cloud top height to estimate Δz at its most extensive point.

4.3.2 *Case study*

One of the utilised ceilometer stations is located at the Hartsfield–Jackson Atlanta International Airport. To illustrate the functionality of the presented algorithm, we investigate a particular MISR overpass over this station on 21 August 2015 at around 16:30 UTC. Figure 4.4 shows the z retrievals for all pixels which are within the circular MIBase cell defined by R_c. Here, we exemplarily use $R_c = 20\,\text{km}$ with its midpoint at the ceilometer station. z is given above the WGS 84 surface, which is approximately equal to sea level. The spatial distribution shows a low cloud layer with z between 800 m and 2000 m, which covers most of the area. Another cloud layer appears between 5 km and 6 km. Some pixels with heights above 7 km indicate the presence of a third layer (Fig. 4.4, left). For a few pixels, MISR was not able to determine z. This might be due to the viewing geometry. A retrieval requires valid images from two different cameras, one camera viewing nadir and the other viewing at a 26.1° angle. In the case studied here, the most missing retrievals are closely attached to high clouds which might lead to shading effects (Fig. 4.4, middle).

The density of the z distribution shows the aforementioned three cloud layers. They are distinguished according to the threshold value for h_{gap} (Fig. 4.4, right) as illustrated for the bottom and middle layer. For the bottom layer, which is selected for further processing, the number of z retrievals marked high confidence cloud is determined to be $n = 621$. This number is well above the threshold N which is defined later. z_{base} is then calculated using $P = 15$ as the preliminary percentile of the z distribution. This yields $z_{\text{base}} \approx 1160\,\text{m}$ above the WGS 84 surface. The mean average scene elevation for the given area is subtracted from the retrieval to obtain $z_{\text{base}} \approx 927\,\text{m}$ above ground level. The closest METAR report for this day is from 16:52 UTC. Three heights were reported at 2800 ft (≈ 853 m), 7500 ft (≈ 2286 m) and 23000 ft (≈ 7010 m) above ground level. By adding the station elevation (315 m), the corresponding height above sea level is obtained. This yields $\hat{z}_{\text{base}} \approx 1168\,\text{m}$ and is denoted in Fig 4.4 (right). In conclusion, using the preliminary values for P the z_{base} retrieval from MISR is about 927 m above ground level which is 74 m higher than the ceilometer retrieval ($\hat{z}_{\text{base}} = (853 \pm 15)\,\text{m}$). The given uncertainty solely represents the resolution of the METAR reports (Tab. 4.2). Note that the third layer detected around 7000 m by MISR has also been detected by the ceilometer.

Table 4.3: Slope, intercept, correlation coefficient r, RMSE E, bias B and number of samples n_s resulting from comparing z_{base} and $\hat{z}_{base}$ retrievals for different radii of the MISR circular area around the ceilometer stations. These values are obtained for the year 2008 applying a required minimum number of cloud pixels of $N = 10$ and the 15th percentile to the z distribution.

R_c [km]	slope	intercept [m]	r	E [m]	B [m]	n_s
5	0.65	371	0.66	392	-71	3059
10	0.62	412	0.66	404	-75	5120
15	0.60	433	0.65	413	-77	6140
20	0.58	464	0.63	423	-74	6895
30	0.54	515	0.60	437	-71	7772

4.3.3 *Parameter optimization*

For each considered ceilometer station (Fig. 4.2), collocated and coincidental MISR overpasses from the year 2008 are identified. The algorithm is then applied as described in the case study (Sec. 4.3.2 to retrieve z_{base}. All pairs of MIBase z_{base} and ceilometer $\hat{z}_{base}$ are evaluated to investigate the influence of R_c, N and P on the performance of the z_{base} retrieval algorithm and to estimate the most suitable values. For this purpose, the following statistical measures are considered: the slope and intercept of a linear regression, which are ideally 1 and 0, respectively; the Pearson correlation coefficient r (ideally unity); the root mean square error (RMSE) E defined as

$$E = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(z_{base,i} - \hat{z}_{base,i})^2}; \tag{4.2}$$

and the retrieval bias B defined as

$$B = \frac{1}{n}\sum_{i=1}^{n}(z_{base,i} - \hat{z}_{base,i})\,. \tag{4.3}$$

MISR can only detect clouds above the threshold height according to Equation 4.1. To prevent this obvious limitation from introducing a bias into the statistics, we only consider cloud scenes for which the ceilometer retrieval is above h_{min}. In addition, only z_{base} retrievals below a maximum height h_{max} of 3000 m are considered to focus on a cloud range for which the ceilometer retrievals are more finely granulated (below 10000 ft according to Tab. 4.2).

First, we investigate the influence of the size of the MIBase cell on the comparison of MIBase and ceilometer retrievals. For this purpose, R_c is varied between 5 and 30 km while the other parameters are set to

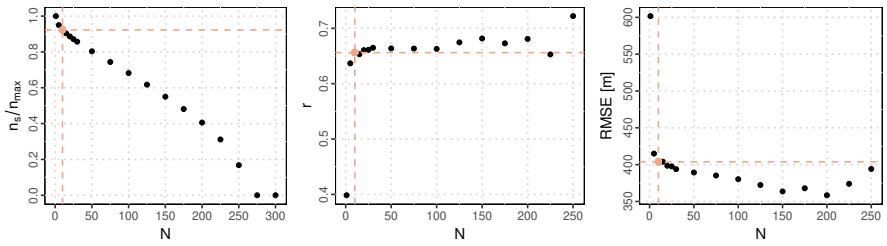

Figure 4.5: Evaluation of minimum number of valid pixels N within a cloud layer detected by MISR for the year 2008. Left: The normalized number of events $\frac{n_s}{n_{max}}$ for which z_{base} and $\hat{z}_{base}$ could both be retrieved. n_{max} is the maximum number of events, which is found for $N = 1$. Middle: The linear correlation coefficient r between z_{base} and $\hat{z}_{base}$. Right: The RMSE between z_{base} and $\hat{z}_{base}$. MISR z_{base} is retrieved using the 15th percentile of the z distribution for a 10 km radius around the individual ceilometer measurements. The chosen value for N is highlighted in orange. For further details see text.

the preliminary values $P = 15$ and $N = 10$. With a decreased R_c, the correlation between z_{base} and $\hat{z}_{base}$ increases and E decreases (Tab. 4.3). This is to be expected as the representativity should increase. However, for a lower R_c, the retrieval algorithm encounters more situations where at least one of the requirements (at least one high confidence surface pixel is visible and at least 10 valid cloud pixel per layer) cannot be fulfilled, as the decrease in the total number of retrievals indicates. The better agreement between z_{base} and $\hat{z}_{base}$ for lower R_c might be due to a relatively larger overlap of the measurement sampling areas of the two instruments and to a better fulfilment of the assumption of a homogeneous z_{base} over smaller areas. For further evaluation, a radius of 10 km is chosen as a compromise between a good agreement in terms of r and E and without having to discard too many retrieval scenes.

Second, the effect of the minimum number of of valid z_{base} retrievals is studied which strongly limits the number of samples for the comparison (Fig. 4.5). With increasing N, initially a slight increase to $N = 10$ improves the correlation between z_{base} and $\hat{z}_{base}$ and E significantly to a correlation coefficient of about 0.66. A further increase only yields slight improvement of the correlation and E. This slight increase can be explained by the elimination of more complex scenes from the comparison. However, for a higher N the trade off is a lower total number of z_{base} retrievals. For instance, for $N = 50$ only 80 % of possible retrievals yield a valid z_{base} (Fig. 4.5, left). Therefore, we select $N = 10$.

Finally, we consider the percentile threshold used to diagnose z_{base} from the z distribution. Figure 4.6 shows an evaluation of different percentiles which are applied to derive z_{base} . Percentiles between the 10th and the 15th give the best correlation. The lowest E is achieved

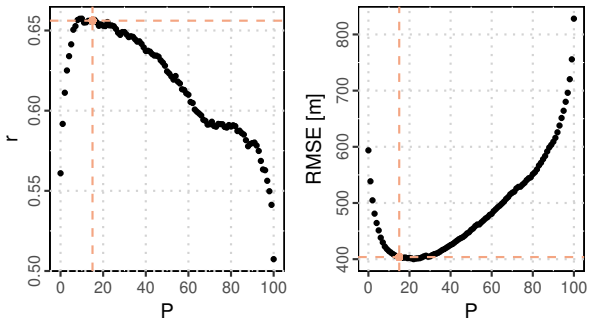

Figure 4.6: Evaluation of the percentile P which is applied to retrieve z_{base} from the distribution of z for the year 2008 with $N = 10$ and $R_c = 10\,\text{km}$. Left: The linear correlation coefficient r between z_{base} and $\hat{z}_{\text{base}}$. Right: The RMSE between z_{base} and $\hat{z}_{\text{base}}$. The chosen value for P is highlighted in orange.

Table 4.4: Slope, intercept, correlation coefficient r, RMSE E, bias B and number of retrievals n_s resulting from a comparison of z_{base} and $\hat{z}_{\text{base}}$ for data obtained 2008 (calibration) and 2007 (validation). These values are obtained with $N = 10$ and $P = 15$.

data	pixel/grid definition	slope	intercept [m]	r	E [m]	B [m]	n_s
2008	$R_c = 10\,\text{km}$	0.62	412	0.66	404	-75	5120
2007	$R_c = 10\,\text{km}$	0.61	419	0.66	385	-59	6801
2007	$0.25° \times 0.25°$	0.58	455	0.64	398	-60	7970
2007	$0.75° \times 0.75°$	0.49	579	0.55	446	-56	10474

for percentiles between the 15th and the 25th. Therefore, $P = 15$ is chosen for further processing. The fact that very clear and localised minima (maxima) for E (r) are found supports the hypothesis that the z distribution contains information on z_{base}.

In summary, the comparison yields the estimated parameters $R_c = 10\,\text{km}$, the minimum number $N = 10$ and the percentile $P = 15$. While the latter two are kept fixed in MIBase, R_c is optimised for the intercomparison with point data, i.e. ceilometer measurements. The algorithm can also be applied to larger grids. However, no data for validating extended areas are available.

4.3.4 *Scene limitations*

This section investigates the applicability of MIBase by quantifying the amount of cases for which the concurrent conditions allow the successful derivation of a z_{base} retrieval. First, we filter for cases which fulfill the following two conditions: i) The number of valid z retrievals within the MIBase cell N_{val} must be > 0 and ii) METAR data must

Table 4.5: Number of cases for different conditions of the cloud field observed by MISR and reported in METAR messages for the considered METAR sites. The number of z retrievals labeled "high confidence cloud" (N_{HCC}) or "high confidence surface" (N_{HCS}) according to MISR's stereo-derived cloud mask is used to characterize the cloud field. The size of the scene is defined by $R_{\mathrm{c}} = 10\,\mathrm{km}$.

description of the situation		2008	2008	2007	2007
MISR	METAR		[%]		[%]
overpasses over METAR sites		80454	154.1	89782	145.9
valid z retrievals	message available	52215	100.0	61531	100.0
$N_{\mathrm{HCC}} = 0$; $N_{\mathrm{HCS}} > 0$ (clear sky*)		19507	37.4	20300	33.0
	clear sky*	26983	51.7	30037	48.8
clear sky*	clear sky*	16982	32.5	17374	28.2
$N_{\mathrm{HCC}} = 0$; $N_{\mathrm{HCS}} > 0$ (clear sky*)	$\hat{z}_{\mathrm{base}}$ retrieval	2525	4.8	2926	4.8
$N_{\mathrm{HCC}} = 0$; $N_{\mathrm{HCS}} > 0$ (clear sky*)	$\hat{z}_{\mathrm{base}} > h_{\mathrm{min}}$	2106	4.0	2520	4.1
$N_{\mathrm{HCC}} > 0$; $N_{\mathrm{HCS}} > 0$	clear sky*	6800	13.0	8511	13.8
$N_{\mathrm{HCC}} > 0$; $N_{\mathrm{HCS}} = 0$ (overcast*)		15945	30.5	19725	32.1
$N_{\mathrm{HCC}} > 0$; $N_{\mathrm{HCS}} = 0$ (overcast*)	$\hat{z}_{\mathrm{base}}$ retrieval	12769	24.5	15600	25.4
$N_{\mathrm{HCC}} > 0$; $N_{\mathrm{HCS}} = 0$ (overcast*)	clear sky*	3176	6.1	4125	6.7
$N_{\mathrm{HCC}} = 0$; $N_{\mathrm{HCS}} = 0$		51	0.1	51	0.1
$N_{\mathrm{HCC}} > 0$; $N_{\mathrm{HCS}} > 0$	$\hat{z}_{\mathrm{base}}$ retrieval	9912	19.0	12944	21.0
$N_{\mathrm{HCC}} \geq N = 10$; $N_{\mathrm{HCS}} > 0$	$\hat{z}_{\mathrm{base}}$ retrieval	8603	16.5	11387	18.5
z_{base} retrieval	$\hat{z}_{\mathrm{base}}$ retrieval	8535	16.3	11319	18.4
z_{base} retrieval; single layer	$\hat{z}_{\mathrm{base}}$ retrieval	7863	15.1	10251	16.7
$z_{\mathrm{base}} < h_{\mathrm{max}} = 3\,\mathrm{km}$; single layer	$\hat{z}_{\mathrm{base}}$ retrieval	7206	13.8	9407	15.3
$z_{\mathrm{base}} < h_{\mathrm{max}}$; single layer	$\hat{z}_{\mathrm{base}} < h_{\mathrm{max}}$	7043	13.5	9227	15.0
$z_{\mathrm{base}} < h_{\mathrm{max}}$; single layer	$h_{\mathrm{min}} < \hat{z}_{\mathrm{base}} < h_{\mathrm{max}}$	5120	9.8	6801	11.1

* indicates apparent conditions. See text for details.

be available for the calibration and validation. These requirements are fulfilled for about two thirds of a all considered MISR overpasses over the ceilometer sites (Table 4.5). Furthermore, there are two main conditions which prevent the derivation of a z_{base} retrieval. These are namely apparent clear sky conditions and apparent overcast which is only a limitation for MIBase. Here, we use the phrases "apparent clear sky" and "apparent overcast" rather than "clear sky" and "overcast", respectively, to account for the fact that this attribution is based on instrumental indications rather than known actual sky condition.

For METAR, apparent clear sky is indicated if a METAR message is available, but does not provide a valid retrieval. Note that in case the lowest cloud is above the METAR reporting range (typically 3700 m),

it is possible that no retrieval is issued. Here, such cases would also be attributed apparent clear sky.

For MIBase, we attribute apparent clear sky to the following configuration of the SDCM: MISR sees the surface with high confidence ($N_{\mathrm{HCS}} > 0$), and has no high confidence cloud in the view ($N_{\mathrm{HCC}} = 0$). This does not have to be an actual clear sky case since it could include low confidence surface or low confidence cloud retrievals for which the declaration is less certain. In case of invalid z retrievals, it is also uncertain whether clouds are present or not.

Out of all MISR apparent clear sky cases, 87 % are also classified as clear sky by METAR while the remaining 13 % yield a METAR cloud height retrieval. Mismatches in attributing apparent clear sky cases are due to METAR retrievals below the threshold height h_{min} (17 %) and other reasons, such as the temporal offset between MISR and METAR measurement. The METAR reports comprise retrievals over a 30 minute period. During this time, cloud formation and cloud dissipation can alter the cloud scene and cause mismatches between MISR and METAR retrievals.

Furthermore, for MIBase, we attribute apparent overcast to the following configuration of the SDCM: MISR observes a cloud with high confidence ($N_{\mathrm{HCC}} > 0$) and does not observe any surface retrievals with high confidence ($N_{\mathrm{HCS}} = 0$). Again, the scene could include invalid retrievals, or retrievals of low confidence. In about 20 % of all the MISR apparent overcast cases, the corresponding METAR report yields an apparent clear sky case. These could be cases where the cloud cover is mainly above the reporting range of the ceilometer.

Out of all cases with valid z retrievals within the MIBase cell ($N_{\mathrm{val}} > 0$) and a corresponding METAR retrieval, 19 % are processed further. The main reasons why cases are excluded are apparent clear sky scenes for MISR (37.4 %), apparent overcast for MISR (30.5 %) and apparent clear sky for METAR when valid z retrievals are within the MIBase cell (13 %). Additional requirements, such as the minimum number of z retrievals marked high confidence cloud ($N_{\mathrm{HCC}} > N$), single layer situations, z_{base} and $\hat{z}_{\mathrm{base}}$ retrievals below h_{max} and METAR retrievals above the MISR threshold height ($\hat{z}_{\mathrm{base}} > h_{\mathrm{min}}$), lead to a further reduction of the number of cases which are used to derive the statistics. Further numbers for specific cases are presented in Table 4.5.

4.4 MIBASE EVALUATION

With the parameters $R_{\mathrm{c}} = 10\,\mathrm{km}$, $N = 10$ and $P = 15$ derived in the previous section, MIBase is applied to MISR retrievals which are coincident with ceilometer retrievals from the year 2007. These data have not been used for calibration. The joint density of z_{base} retrieved from MISR and ceilometer is shown in Fig. 4.7. For lower z_{base}, MISR yields higher heights than the ceilometers. This can possibly be attributed

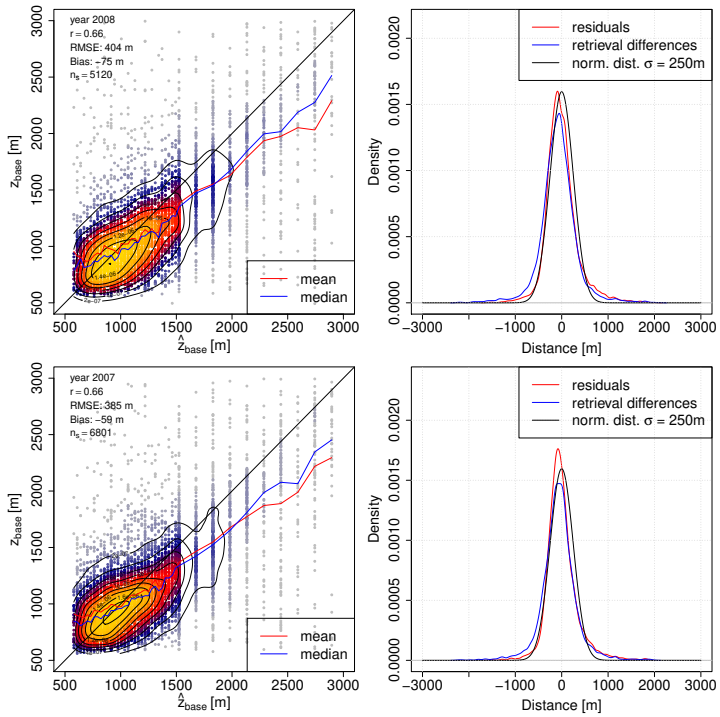

Figure 4.7: Left: Joint density of z_{base} and $\hat{z}_{base}$ for the year 2008 (top) which is used to estimate parameters of the algorithm and for the year 2007 (bottom) which is used to validate the stability of the algorithm with the estimated parameters. The value of the normalized density is indicated by color (maximum values in light yellow) and contour lines with corresponding values on them (linear scale). For each ceilometer height bin, the mean (red) and median (blue) of the MISR z_{base} is shown. Right: Probability density functions of the residuals after a linear fit (red), the retrieval differences (blue) and a normal distribution with a standard deviation of 250 m (black).

to the threshold height (Equation 4.1) constraining z_{base} retrievals at the lower end of the height distribution. For z_{base} greater than 1000 m, mean and median MISR heights are lower than the ceilometer. Overall, the bias B is slightly negative (about 60 m; cf. Tab. 4.4) and the density of the retrieval differences is shifted slightly towards negative values (Fig. 4.7d). Thus, MISR z_{base} retrievals are generally lower than the ceilometer retrievals. This could be due to the different sample volumes. On the one hand, the ceilometer only records point measurements over a period of time, so that the measured sample of the cloud depends on the velocity of the wind. On the other hand, MISR observes the entire circular area defined by R_c around the ceilometer location. Chances are that MISR can observe a cloud with a lower base which does not pass over the ceilometer.

The joint density and the density of the retrieval differences appear similar for both the 2007 and the 2008 data sets (Fig. 4.7). Slope, intercept, r^2, E, and B resulting from the z_{base} retrieval comparisons for

the year 2008 (calibration) and the year 2007 (validation) appear very similar, demonstrating the stability of the algorithm with the chosen parameters (Tab. 4.4) to interannual variability in cloud properties. Changing the MIBase cell to a $0.25^\circ \times 0.25^\circ$ latitude–longitude grid results in a slightly lower correlation coefficient accompanied by a higher E. An even coarser grid size of $0.75^\circ \times 0.75^\circ$, which is applied later for a comparison with ERA-Interim cloud heights, results in an even lower correlation and higher E. A decreasing agreement between z_{base} and $\hat{z}_{\text{base}}$ for a larger MIBase cell has already been described when studying the influence of R_c (see discussion in Section 4.3.3).

4.4.1 *Scene structure influence*

To estimate the influence of the the scene structure on the performance of MIBase, we further exploit the MISR cloud top height product and the MISR Ancillary Geographic Product to investigate characteristics of the terrain height and the cloud field.

To derive a quantity to estimate the variability of the terrain height, we calculate the standard deviation of the average scene elevation, which is provided by the ancillary product at 1.1 km resolution. For each METAR site, the standard deviation is calculated for an area defined by different R_c (5 km, 10 km, 15 km, 20 km and 30 km). Typical standard deviations range around a few tens of meters with overall higher standard deviations for greater R_c (Fig. A.1a in the Supplement). When METAR sites with a higher standard deviation of the average scene elevation are excluded from the comparison of MIBase and METAR cloud base height retrievals, the RMSE decreases slightly, the bias slightly increases (towards 0), while the correlation is hardly affected (Fig. A.1b,c,d). Thus, the variability of the terrain height has a very small effect on the accuracy of the MIBase algorithm, with a slightly better performance over more homogeneous terrain.

To further investigate the performance of the MIBase algorithm as a function of parameters related to cloud types, we determine RMSE, bias, and the correlation coefficient as a function of z_{top} and the cloud vertical extent Δz (Fig. A.2). The best correlation is obtained for cloud vertical extents up to 1000 m. The RMSE is also smaller for lower Δz and for lower z_{top}. However, the RMSE increases with decreasing z_{top} below about 1000 m. We conclude that MIBase performs best for shallow low clouds. However, further analyses are necessary to increase the sample size of thicker clouds and to include more medium high and high clouds for a more robust analysis of such cloud types. Furthermore, the increased RMSE for very low z_{top} indicates that, for very shallow low clouds in the proximity of the threshold height, MIBase retrievals do not agree as well with the METAR retrievals. This might be due to cases for which MIBase detects a shallow low cloud with z_{base} and z_{top} close the h_{min} when, in fact, the actual cloud base is

below h_{min}. MIBase would miss this actual cloud base height because the retrievals below h_{min} would not be marked high confidence cloud. For that matter, we require that the ceilometer retrieval is above the threshold height ($\hat{z}_{base} > h_{min}$). However, if such a near surface cloud was not detected by the ceilometer, a mismatch would result leading to a higher RMSE.

Additionally, we exploit the stereo-derived cloud mask as a proxy of cloud cover fraction to investigate the sensitivity of the MIBase performance to the number of valid z retrievals N_{val}, the number of z retrievals marked high confidence surface N_{HCS}, and the number of z retrievals marked high confidence cloud N_{HCC} within the MIBase cell. We determine RMSE, bias, and the correlation coefficient as a function of N_{val}, N_{HCS} and N_{HCC} normalized by the total number of pixels N_{tot} which the MIBase cell encloses (Fig. 4.8). For example, for $R_c = 10\,km$, a total of $N_{tot} = 265$ pixel is processed by MIBase to obtain a unique z_{base} retrieval. For the continental USA, most cases comprise a high portion of valid z retrievals within the MIBase cell. The RMSE, bias, and the correlation coefficient are robust under different choices of N_{val} and N_{HCS}. This suggests that MIBase generally does not depend much on cloud cover fraction. However, for cases which suggest almost apparent clear sky, indicated by high N_{HCS}, RMSE increases and r decreases. This could be due to a lower chance of observing the same cloud in case of less extended clouds. This bias appears to strongly depend on the portion of z retrievals marked high confidence cloud (Fig. 4.8). The increased bias for higher N_{HCC} could be explained by the decreasing portion of the thin edge of the cloud compared to the thicker part of the cloud with greater horizontal extent. For instance, the edge of a larger cloud might only be partly within the MIBase cell, whereas the edge of a smaller cloud might be fully processed by MIBase. The clear increase of the bias with increasing N_{HCC} shows potential for a bias correction in the future after a better understanding of the underlying reasons. The bias obtained in this study can have different sources: the different sample volumes of the defined MIBase cell and the ceilometer, biased MISR z retrievals, various scene characteristics.

4.5 MIBASE APPLICATION

4.5.1 *Global cloud height distribution*

MIBase has been applied for a three year period between 2007 and 2009 to determine the z_{base} from MISR globally. Herein, z data from each individual orbit have been sorted into a $0.25^\circ \times 0.25^\circ$ longitude by latitude grid. For each orbit and each grid box z_{base} has been retrieved as described above and the median over the three year period has been calculated. Only cloud height retrievals below 5000 m are considered

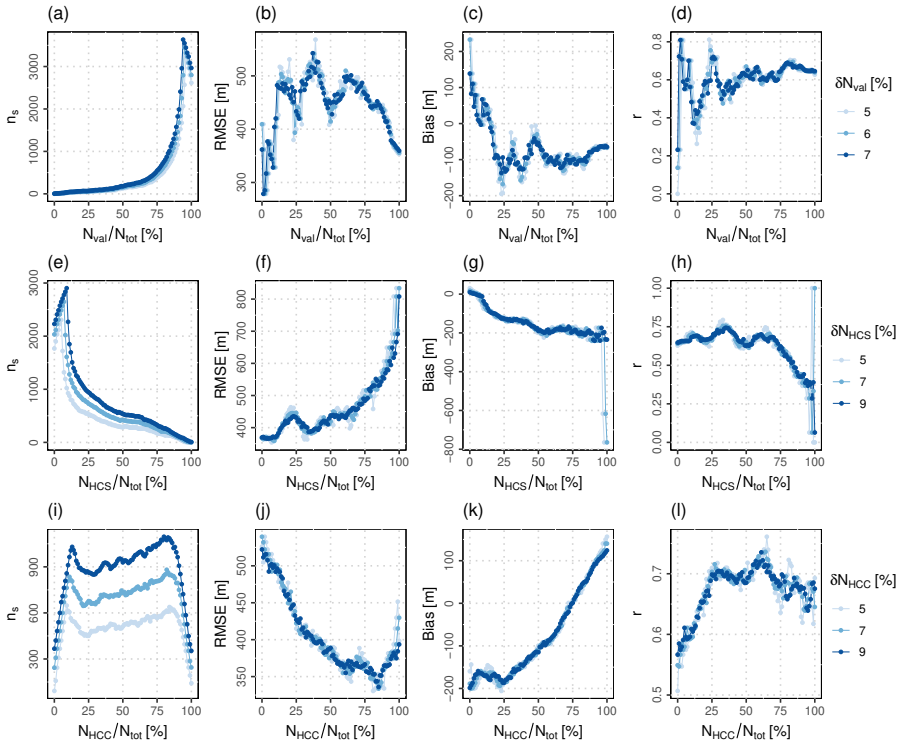

Figure 4.8: From left to right: number of samples n_s, RMSE, bias and correlation coefficient r for the comparison of MIBase and ceilometer retrievals as a function of the number of valid z retrievals N_{val} (top row), the number of retrievals marked high confidence surface N_{HCS} (middle row) and the number of retrievals marked high confidence cloud N_{HCC} (bottom row). Each data point is calculated for a sub sample which includes only $N_{val} \pm \delta N_{val}$, $N_{HCS} \pm \delta N_{HCS}$ and $N_{HCC} \pm \delta N_{HCC}$, respectively. The various widths of the considered N_{val} and N_{HCC} windows are indicated by the blue shading. All values are normalized by the total number of pixels within the MIBase cell N_{tot}. Data are for the year 2008 with $R_c = 10\,\text{km}$, $P = 15$ and $N = 10$.

to exclude cirrus clouds from the statistics. z_{top} is retrieved analogously to z_{base} by applying the 95th percentile on the z distribution. Taking the difference between z_{top} and z_{base} for each observed cloud scene yields Δz. The medians of these measures are shown in Fig. 4.9.

A sharp and steep gradient of the z_{base} can be seen at most coast lines with a higher z_{base} over land. This seems plausible as boundary layers above oceans are known to be shallower. Exceptions to this rule are the Congo Basin and the Amazon Basin. These regions are moisture sinks characterized by high precipitation and excessive surface run-off. The maritime stratus cloud regions are clearly visible at the subtropical eastern boundaries of the Pacific, Atlantic and Indian ocean. These regions are characterized by prevailing high pressure due to the location at the subsiding branch of the Hadley circulation and cold ocean currents creating a temperature inversion on top of

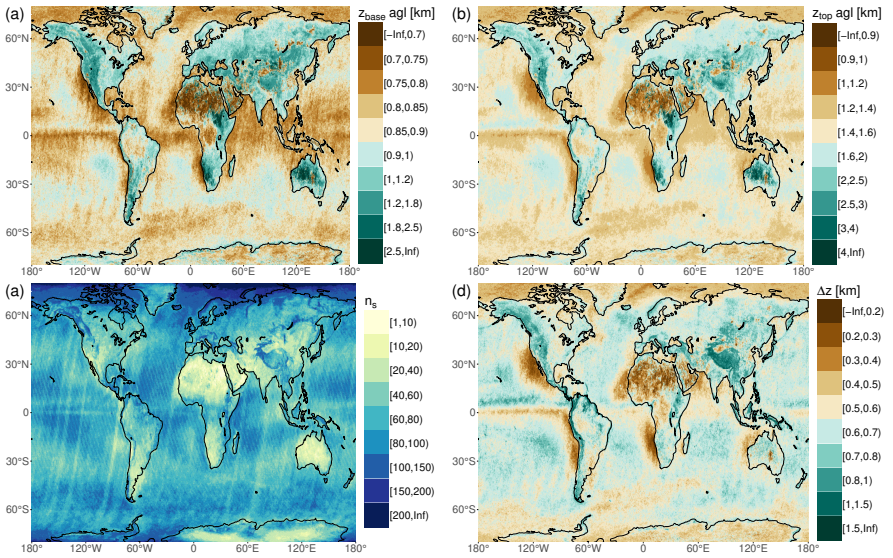

Figure 4.9: Global distribution of median cloud heights for a 3-year period (2007–2009). Shown are z_{base} (a), z_{top} (b), and cloud vertical extent (d) on a $0.25° \times 0.25°$ latitude–longitude grid. z_{base} and z_{top} are above ground level (agl). z_{base} and z_{top} retrievals are only included in the statistic if z_{base} is below 5000 m. The number of retrievals n_s (c) represents the number of valid z_{base} retrievals within this 3-year period.

the boundary layer. For these regions cloud formation is limited to the well mixed maritime boundary layer. The Intertropical Convergence Zone (ITCZ) is clearly visible in particular for the tropical Pacific ocean with a higher z_{base} and even higher z_{top} yielding an overall higher Δz slightly north of the equator. Over land, this phenomenon is not as clear. There, the diurnal cycle of surface heating becomes important. MISR on the Terra satellite has a morning overpass over the equator when cloud formation just begins. Taylor et al., 2017 show the diurnal cycle of cloud top temperature (CTT) derived from SEVIRI measurements indicating that the lowest z_{top} occurs between 9:00 and 13:00 local time with the lowest mean CTT at 11:00. and the lowest median CTT at 12:00, close to the overpass time of MISR.

The sampling size varies spatially with a higher number of retrievals in the Arctic region. (Fig. 4.9 (c)). This is expected for a polar orbiting satellite with more frequent MISR overpasses in polar regions (Fig. 4.10 (a)). Generally, the causes for retrieval failure are apparent clear sky and apparent overcast situations as discussed in Section 4.3.4. The frequency of occurrence of such situations varies spatially. For continental dry regions in the subtropics and continental polar regions apparent clear sky conditions predominantly limit the number of z_{base} retrievals (Fig. 4.10 (c)). The continental polar regions yield a high number of cases for which the grid cell comprises only high confidence surface retrievals ($N_{HCS} = N_{tot}$, Fig. A.3). This poses an even

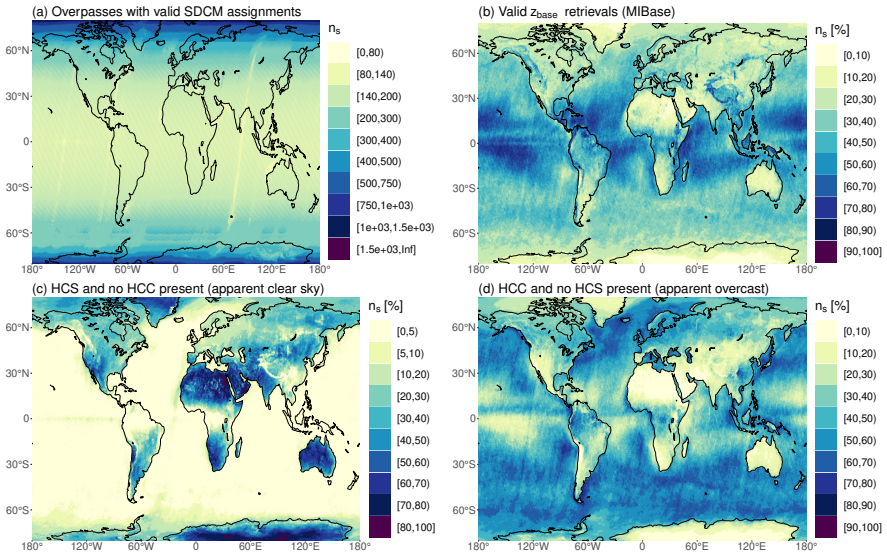

Figure 4.10: Relative occurrences of different stereo-derived cloud mask (SDCM) configurations within the three-year period (2007–2009). The reference sample size n_s given in (a) corresponds to 100 % and includes all overpasses per grid cell which contain valid z retrievals. (b) Relative number for which MIBase successfully retrieved z_{base}. (c) through (d) show the relative number of occurrence of cloud scenes which include z retrievals of specific SDCM labels within a grid cell. These configurations are: (c) No high confidence cloud (HCS). These cases are apparent clear sky cases. (d) No high confidence cloud (HCS). These cases are apparent overcast cases.

more robust indication of apparent clear sky conditions. However, the boundary layer is typically shallower in polar regions. Therefore, boundary layer clouds occur likely below h_{min}, so that z_{base} cannot be retrieved by the MIBase algorithm. Predominant apparent overcast conditions limit the number of z_{base} retrievals for midlatitude regions over ocean and stratocumulus regions on the western boundaries of continents in the subtropics. In midlatitude continental regions, a mix of apparent clear sky and apparent overcast conditions limits the number of z_{base} retrievals. In the trade cumulus regions within 30°N and 30°S, very high success rates occur (Fig. 4.10 (b)). A visual comparison to the 2011 mean cloud cover fraction derived from MODIS (Suen et al., 2014) indicates the plausibility of the attribution of apparent clear sky and apparent overcast.

To further investigate the plausibility of the seasonal variability of cloud heights, composites over the three year period are presented in Fig. 4.11. We distinguish boreal winter season comprising December, January and February (DJF) and boreal summer season comprising June, July and August (JJA). Over land and between 30°N and 70°N, z_{base} and z_{top} are lower during winter, when stratiform clouds prevail. In contrast, z_{base} and z_{top} are higher during summer, when more

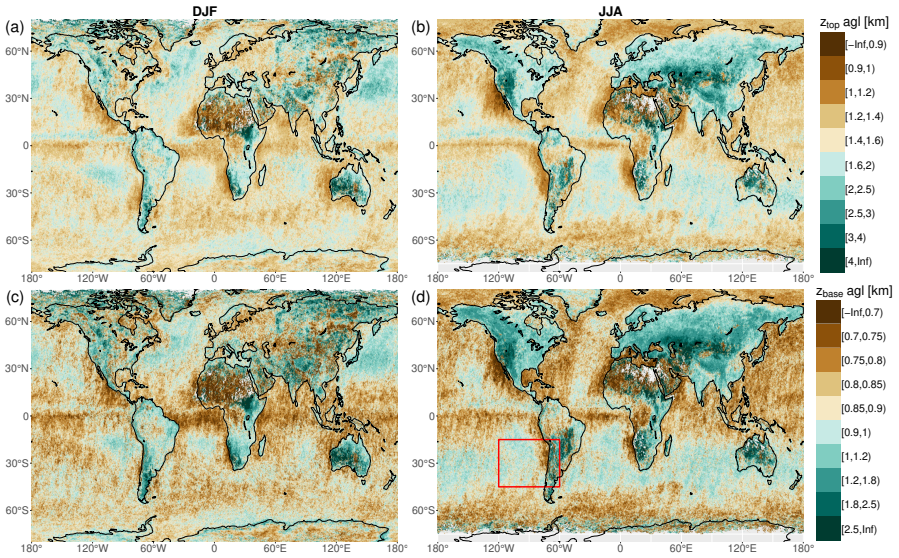

Figure 4.11: Global distribution of seasonal median cloud heights for a 3-year period (2007–2009). Shown are z_{top} (a, b), and z_{base} (c, d) for December, January, February (a, c) and June, July, August (b, d) on a $0.25° \times 0.25°$ latitude–longitude grid. z_{base} and z_{top} are above ground level (agl). z_{base} and z_{top} retrievals are only included in the statistic if z_{base} is below 5000 m. The red rectangle in (d) frames the region for which results over a 16-year period are presented in Fig. 4.12.

convective clouds are typically present. Boundary layer clouds are also lower during winter season since the boundary layer is shallower during the cold season. Over ocean an inverse pattern can be observed on both hemispheres. During winter z_{base} and z_{top} are higher than during the summer. Sea surface temperatures show less seasonal variation than air temperatures due to the higher heat capacity of the water. This causes additional instability during winter enhancing convective cloud formation which can result in higher cloud heights. Additionally, the instability during winter can be attributed to storm tracks. During summer, the influence of high pressure systems can limit convection to the maritime boundary layer causing cloud heights to be lower.

4.5.2 *Southeast Pacific*

The southeast Pacific hosts one of the largest and most persistent stratocumulus cloud decks on Earth as shown by Wood, 2012 using data from the combined land-ocean cloud atlas database (Hahn and Warren, 2007). In this region, cloud cover and cloud thickness have major impacts on the net cloud radiative effect, which raises the importance of studying the heights of these clouds.

Orographically induced fog at the coastal cliff ranging from Peru to northern Chile is the major source of moisture for this region (Pinto et al., 2006). z_{base} and z_{top} of the stratocumulus clouds near the coast determine the areas where fog can provide water to the environment at the coastal cliff. The cloud heights also affect the ability of the fog to be advected further inland across the cliff. Here, we apply the z_{base} retrieval algorithm to determine the spatial and seasonal variability of z_{base} and z_{top} for the region (see red rectangle in Fig. 4.11d). We extend the time window to the full 16-year record of available MISR data (2001–2016). Furthermore, we investigate how well the temporal changes are represented in the global reanalysis ERA-Interim.

4.5.2.1 *Spatial and seasonal variability of z_{base} and z_{top}*

For the 16-year period, the medians of z_{base} and z_{top} over the southeast Pacific are shown in Fig. 4.12. Distinguished are summer and winter season. Over ocean the median z_{base} ranges from 600 m near the the coast to about 1200 m further west. During austral summer (DJF) the lowest z_{base} is observed near the coast between 30°S and 35°S. During austral winter the region of low z_{base} shifts to the north between 20°S and 30°S. This shift is in phase with the direction of the seasonal shift of the Hadley cell. It appears that the region of lowest z_{base} corresponds to the strongest subsidence. During austral summer the highest z_{base} clearly appear in the north, whereas during austral winter a north–south gradient is hardly visible between 120°W and 80°W. Over land, z_{base} is generally higher except for the coastal line north of 35°S, where cloud heights are even lower than over ocean. There, the prevailing maritime stratocumulus clouds form orographic fog as they reach the coastal cliff. Similar spatial and seasonal patterns are apparent for z_{top}. Over ocean, the highest z_{top} is about 2500 m, which is observed during austral summer in the northwest of the region. The lowest z_{top} is about 1000 m, which is observed during winter and closer to the coast of northern Chile.

4.5.2.2 *Cloud height comparison between MISR and ERA-Interim*

In order to preliminarily assess how well clouds are represented in common reanalysis, we compare MISR derived z_{base} and z_{top} to cloud heights derived from ERA-Interim (Dee et al., 2011a) which is provided by the European Centre for Medium-Range Weather Forecasts (ECMWF). Cloud heights are not a direct output variable of ERA-Interim. Therefore, the cloud liquid water content is used to infer the cloud base height $\tilde{z}_{base}$ and cloud top height $\tilde{z}_{top}$. For each grid point, the vertical column is scanned for model levels with a specific cloud liquid water content greater than $10^{-18}\,\mathrm{kg\,kg^{-1}}$ (≈ 0). The bottom height of the lowest of such levels is taken as $\tilde{z}_{base}$. Moving higher in the column, $\tilde{z}_{top}$ is given by the bottom height of the next higher

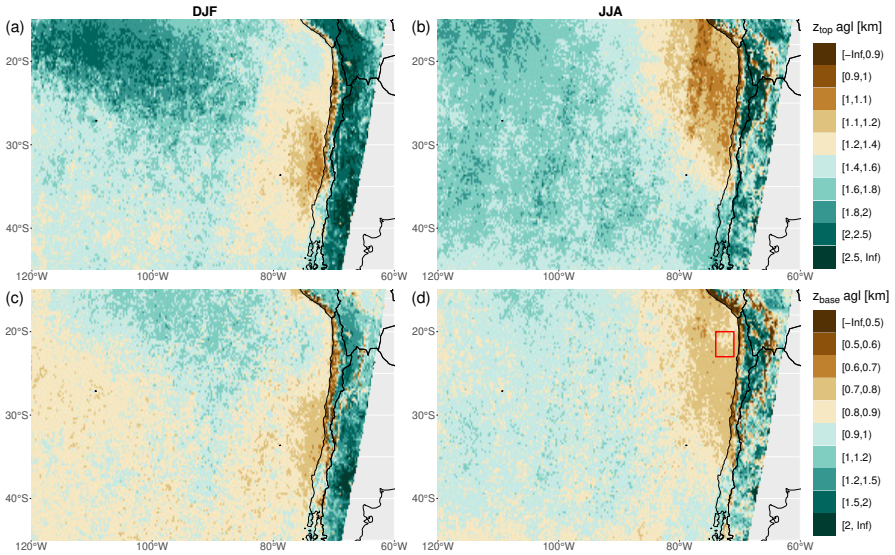

Figure 4.12: Median of z_{top} (a, b), and z_{base} (c, d) over a 16-year period (2001–2016) for austral summer (DJF, (a) and (c)) and austral winter (JJA, (b) and (d)) on a $0.25° \times 0.25°$ longitude by latitude grid at the southeast Pacific. z_{base} and z_{top} are given above ground level (agl). The red rectangle (d) frames the region for which a time series of cloud heights is presented in Fig. 4.13.

model level which has a cloud liquid water content equal to zero. We use data with a $0.75° \times 0.75°$ resolution, which is similar to the native grid of ERA-Interim, over a region between 20°S and 23°S and 74°W and 71°W as indicated by the red rectangle in Fig. 4.12. ERA-Interim data is provided 6-hourly. The comparison is performed using the 18:00 UTC output which corresponds to 14:00 Chile Standard Time (CLT). Note, MISR overpass times range around 10:51 CLT to 11:29 CLT for this particular region.

For each MISR overpass and ERA-Interim 18:00 UTC output, the median cloud heights are used to calculate the median cloud heights of each month over the whole 16-year period. The mean difference of the monthly cloud heights is roughly 500 m for both cloud base height and cloud top height, with ERA-Interim yielding lower cloud heights than MISR. That $\tilde{z}_{base}$ is lower than z_{base} could be due to the threshold height used to determine the MISR stereo derived cloud mask (Equation 4.1) which leads to a cut-off of z_{base} retrievals at h_{min}. At the same time the same bias is found between z_{top} and $\tilde{z}_{top}$. This could be an indicator that clouds are systematically placed too low by ERA-Interim. Hannay et al. (2009) mentioned several studies which conclude that models typically underestimate the height of the planetary boundary layer (PBL) in the southeast Pacific area. This would cause boundary layer clouds to appear lower than observed. Their study compares the PBL height retrieved from in-situ measurements and remote sensing to different models. While the observations show

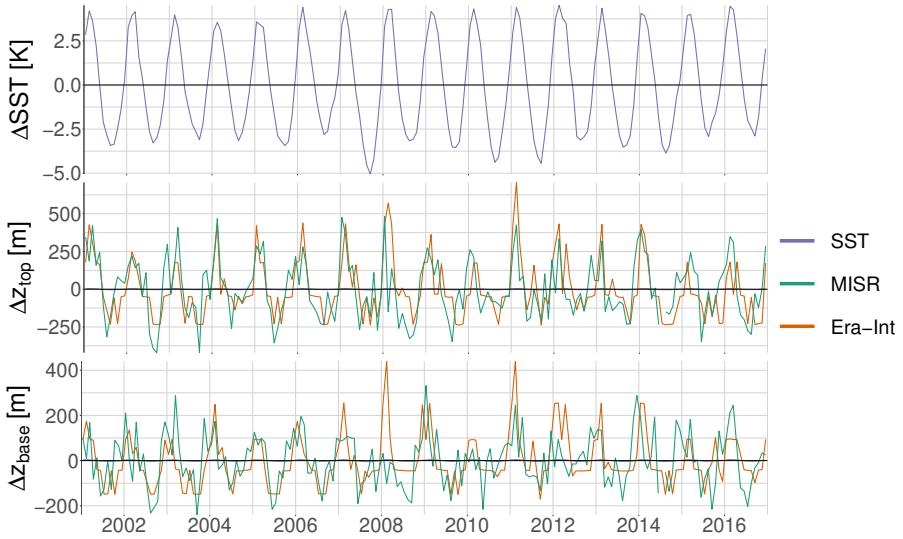

Figure 4.13: Time series of deviations of sea surface temperature ΔSST (top), cloud top height Δz_{top} (middle), cloud base height Δz_{base} (bottom) from the corresponding mean over the entire period from 2001 through 2016. Cloud heights are derived from MISR (green) and ERA-Interim (orange). SST is derived from ERA-Interim.

a PBL height of 1100 m, the models produce a PBL height between 400 m and 800 m, hence an underestimation of 700 m to 300 m. This is in accordance with the bias found here.

To reveal the annual cycle of the cloud heights, we look at anomalies from the 16-year mean of each time series (Fig. 4.13). These anomalies of z_{base} and $\tilde{z}_{\text{base}}$ as well as z_{top} and $\tilde{z}_{\text{top}}$ from their respective mean values agree rather well, thus the amplitude of the annual cycle appears very similar. Figure 4.13 also shows the anomaly of the sea surface temperature (SST) from its 16-year mean value. SSTs are taken from ERA-Interim as well. The peaks of the cloud heights correspond to the maxima of the SSTs. While the highest SSTs coincide with the highest cloud heights during austral summer, the lowest SSTs coincide with the lowest cloud heights during austral winter.

4.6 CONCLUSION

Here, we present a new method to determine z_{base} over a spatial region from satellite based measurements. The MIBase algorithm derives z_{base} from the high spatial resolution MISR cloud top height product z if some preconditions, such as a broken cloud scene, are met. Validation against 1510 ceilometer stations in the continental USA results in a correlation coefficient of 0.66 and a RMSE of 385 m for the validation data set (year 2007). The bias of -59 m even states that MISR sees a slightly lower z_{base} on average. This is possibly due to the larger

retrieval cell which is set up for the retrievals from MISR as opposed to the point measurements provided by the ceilometer.

Very few attempts to derive z_{base} from satellite have been performed and evaluated before. Desmons et al. (2013) retrieve Δz from POLDER measurements. The standard deviation of the difference between their Δz retrieval and reference data from CPR and CALIOP is about 964 m. However, their method is hard to compare to the MIBase algorithm, since they retrieve Δz and make a distinction of different types of clouds which is not done in this study. The CBASE algorithm (Mülmenstädt et al., 2018) derives z_{base} from CALIOP measurements even for optically thick clouds. Depending on the circumstances different retrieval uncertainties can be derived. Similar to the study presented here, they compare their z_{base} retrievals with ceilometer data over the continental USA. They obtain RMSEs between 404 m and 720 m depending on the concurrent local conditions of the individual retrievals. The RMSE we obtain for the MIBase algorithm is slightly lower. Even though the two studies make use of a similar reference data base, they measure cloud heights at different times of the day. While CALIOP has an afternoon overpass, MISR has a morning overpass, when more clouds of lesser extent are present. For a more in-depth comparison and validation of the presented algorithm, more cloud height reference observations would be desirable including observations in different climate zones and especially over ocean.

Within Europe, the European Cooperation in Science and Technology (COST) activity is expected to harmonise the networks of the different weather services (e.g., Haeffelin et al., 2016; Illingworth et al., 2019), enabling more intercomparisons in the future.

An important strength of MIBase is the geometric approach which is applied to create the z product from MISR measurements. Neither a calibration nor auxiliary data are necessary to obtain the z product which is the starting point for the z_{base} retrieval algorithm presented here. In consequence, retrievals are possible over all kinds of terrain even above ice. A disadvantage is the threshold height which MISR requires to create the stereo derived cloud mask. Therefore, depending on the terrain variability in the vicinity of the measurement, this new z_{base} retrieval method is not capable of deriving z_{base} below at least 560 m (flat terrain). The algorithm requires a broken cloud scene. For complete overcast within the chosen MIBase cell, z_{base} cannot be retrieved. Therefore, climatologies derived from this algorithm would be biased towards cloud types for which MISR is able to observe the surface through cloud gaps.

Depending on the application, the MIBase uncertainty and the missing coverage of the diurnal cycle can be a limitation. However, in combination with ceilometer networks, both temporal and spatial patterns can be investigated. The application of MIBase over a three-year period reveals plausible patterns in the global distribution and

seasonal variability of z_{base}. A first analysis over the 16-year MISR time series in the southeast Pacific shows the potential to investigate the interannual variability of z_{base}. This makes MIBase a promising tool for the evaluation of climate models on seasonal and interannual time scales in data sparse regions if for example the climate model output is limited to clouds below 5 km and cloud fractions below 1 and if a sufficient amount of MIBase retrievals is provided within the considered region and time period.

4.7 APPENDIX: SENSITIVITY TO THRESHOLD HEIGHT

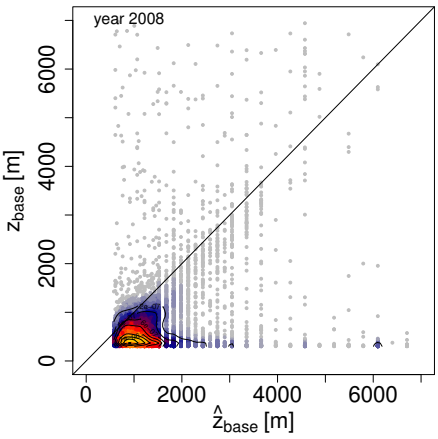

Figure 4.14: Joint density of z_{base} and $\hat{z}_{\text{base}}$ for the year 2008 applying a lower threshold height $h_{\text{min}} = 300\,\text{m} + H + 2\sigma_{\text{h}}$ (Equation 4.4) for the distinction between surface and cloud pixels in contrast to Equation 4.1.

The distinction between surface and cloud retrieval according to the threshold height described by Equation 4.1 introduces a constraint to the z_{base} retrieval algorithm. Below a height of 560 m for flat terrain, or higher for more complex terrain, z_{base} retrievals are not possible. As an attempt to lower this threshold height, we adjusted H_{SDCM} in Equation 4.1, so that:

$$h_{\text{min}} = 300\,\text{m} + H + 2\sigma_{\text{h}} \tag{4.4}$$

This modification results in a bimodal retrieval density clearly showing a mode consisting of surface retrievals (Fig. 4.14). Therefore, the original threshold height given by MISR has to be applied, in order to ensure that only cloud retrievals are utilized during data processing.

4.8 REMARKS

SUPPLEMENT Supplement material is available in the Appendix A.1 or at: https://doi.org/10.5194/amt-12-1841-2019-supplement

DATAAVAILABILITY Multiple archives providing METAR data are available. The data utilized here were downloaded from the Weather Underground archive (https://www.wunderground.com/history/airport/, last access: 4 December 2018). The MISR Level 2TC Cloud Product data were downloaded from the NASA Langley Research Center Atmospheric Science Data Center (ftp://l5ftl01.larc.nasa.gov/MISR/MIL2TCSP.001/, last access: 17 October 2017). ERA-Interim data were downloaded from the ECMWF data server via Web-API. The MIBase cloud base dataset (Böhm, 2019) is freely available at the Collaborative Research Centre 1211 database under the DOI https://doi.org/10.5880/CRC1211DB.19. It comprises z_{base} retrievals globally on a $0.25^\circ \times 0.25^\circ$ grid for a three year period (2007–2009). Daily files include z_{base} retrievals derived from the MISR MIL2TCSP product for about 14 respective Terra revolutions around the Earth. Cloud base altitudes are given above the WGS 1984 ellipsoid. Furthermore, the surface altitude is provided to derive the cloud base height above ground level.

ACKNOWLEDGMENTS The MISR Level 2TC Cloud Product data were obtained from the NASA Langley Research Center Atmospheric Science Data Center (Mueller et al., 2013). We gratefully acknowledge financial support by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – Projektnummer 268236062 – SFB 1211. We thank the seven reviewers for their constructive feedback.

5

WATER VAPOR VARIABILITY

WATER VAPOR VARIABILITY IN THE ATACAMA DESERT DURING THE 20TH CENTURY

Böhm, C., M. Reyers, J. H. Schween, and S. Crewell (2020). "Water vapor variability in the Atacama Desert during the 20th century." In: *Global and Planetary Change* 190, p. 103192. ISSN: 0921-8181. DOI: 10.1016/j.gloplacha.2020.103192.

ABSTRACT. This study focuses on integrated water vapor (IWV) which is the main source for precipitation, fog and dew formation in the Atacama Desert in northern Chile. In order to study its long-term variability, a consistent meteorological record is needed. Here, we utilize the European Centre for Medium-Range Weather Forecasts' reanalysis ERA-20C which provides IWV among other atmospheric variables over the course of the entire 20th century (1900–2010). In this two fold study, we first present a validation of ERA-20C IWV for the Atacama and the bordering southeast Pacific region. Comparisons to satellite observations, i.e. the Hamburg Ocean Atmosphere Parameters and Fluxes from Satellite data record and the Moderate Resolution Imaging Spectroradiometer measurements, for overlapping time periods prove the suitability of ERA-20C to study IWV variability. Assessment of the observation feedback in ERA-20C reveals a higher uncertainty for the beginning of the 20th century when fewer observations are assimilated. Nevertheless, departures of the assimilated observations do not show a systematic bias in space or time supporting suitability of ERA-20C for long-term investigations. In the second part of the study, we describe the IWV variability over the course of the 20th century. Deviations from the long-term mean greater than 30 % are found on an inter-annual time scale over the continental Atacama. Furthermore, we investigate potential drivers of the IWV variability such as the Pacific Decadal Oscillation (PDO) and the El Niño Southern Oscillation (ENSO) phenomenon. The relationship between the local IWV and these large scale indices depends on region and season. For instance, during austral summer, La Niña conditions yield overall greater IWV variability in the Atacama allowing both drier and even more pronounced wetter extremes than El Niño conditions.

5.1 INTRODUCTION

The Atacama Desert in northern Chile is one of the driest places on Earth. Nevertheless, it hosts a variety of species and microorganisms which adapted to the concurrent hyper-arid conditions. Their spatial appearance is not well understood but it is likely connected to the availability of water. For instance, Pinto et al. (2006) found that the geographical distribution of Tillandsia lomas is associated with fog corridors. Furthermore, events of extreme precipitation or wetter time periods on geological time scales can leave long lasting traces in the landscape and impact biological evolution and colonization. Characterizing the moisture supply to the Atacama Desert in the context of the recent climate is essential in order to establish thresholds for growth and development of the local biota and for surface alterations.

Water vapor, which amounts to about 99.5 % of the total water in the atmosphere (Stevens and Bony, 2013), is the most important source for precipitation and is the key variable for fog formation and dew. Aside

from these obvious sources of liquid water for plants and surfaces, water vapor itself constitutes a direct source of water for soils in arid regions via water vapor adsorption and thereby stimulating microbial activity (McHugh et al., 2015). Furthermore, relative humidity along with temperature determines the phase transitions between gypsum, anhydrite and their intermediate phases which has been demonstrated in theory by Tang et al. (2019). Relative humidity and the isotopic composition of the water vapor which is related to its source and pathway are essential variables in order to develop a paleo-humidity proxy (Surma et al., 2018). A better knowledge of the spatiotemporal distribution of water vapor over a longer time period could help improve the accuracy of such a proxy.

Another field of application for water vapor in the Atacama Desert is Astronomy. The region is home to multiple astronomical facilities, such as the European Southern Observatory (ESO) which operates for instance the Very Large Telescope at the summit of Cerro Paranal. Even though, the Atacama provides a hyper-arid environment, water vapor is still a limiting factor of atmospheric transparency in the millimeter and submillimeter wavelength spectral window. Characterizing the variability of water vapor and identifying potential drivers benefits the development of the observatories and planning the conduction of very demanding observations (Kerber et al., 2014; Otarola et al., 2019). Water vapor is one of the key factors which determine the surface solar radiation budget (Rondanelli et al., 2015). The higher elevated parts of the Atacama were found to be the most likely location with the highest downwelling solar radiation at the surface on Earth (Rondanelli et al., 2015). The extreme solar radiation also results in an extreme exposure of the surfaces to ultraviolet radiation (Cordero et al., 2016) which is a limiting factor to bacterial life (Cockell et al., 2008).

The influence of internal climate modes such as the El Niño Southern Oscillation (ENSO) pattern or the Pacific Decadal Oscillation (PDO) on precipitation and water vapor content has been assessed for the Atacama region and the bordering Altiplano in various studies (e.g. Garreaud, 1999; Vuille, 1999; Vargas et al., 2006; Houston, 2006; Garreaud et al., 2009; Marín and Barrett, 2017). For instance, through evaluation of gauge measurements it was demonstrated that the warm phase of ENSO (El Niño) leads to more precipitation at coastal stations and less precipitation in the Altiplano during austral summer (Houston, 2006). Furthermore, the PDO, which which shows similar patterns in SST anomaly as ENSO but acts on a much longer time scale (Garreaud et al., 2009), appears to amplify the ENSO signal during its warm phase (Andreoli and Kayano, 2005). On a synoptic scale, cut-off lows over the adjacent southeast Pacific are associated with increased moisture supply to the Atacama Desert (e.g. Bozkurt et al., 2016; Reyers and Shao, 2019).

Water vapor is a major part of the water cycle being the dominant phase which is subject to transport. More details on the atmospheric water cycle in the Atacama region are given in Section 5.2.6. Due to the scarcity of in-situ measurements especially over longer time periods, region-wide studies on a climatological scale are limited. Satellite observations can provide greater spatial coverage but are temporally limited to the satellite era which started in the 1980s. For long term trend analyses and to study dependencies on low frequency internal climate modes such as the ENSO pattern or the PDO, longer time series are beneficial. Such time series with broad spatial and temporal resolution including multiple atmospheric parameters can only be provided by reanalyses data. Reanalyses combine model simulations and observations to obtain the best estimate of the true state of the atmosphere. To our knowledge four reanalyses data sets which cover the entire 20th century are available as of today, namely the National Oceanic and Atmospheric Administration's (NOAA) Twentieth Century Reanalysis (20CR) and its successor 20CRv2c (Compo et al., 2011) and the European Centre for Medium-Ranged Weather Forecasts (ECMWF) twentieth century reanalysis (ERA-20C Poli et al., 2016) and a later release with a coupled ocean model (CERA-20C Laloyaux et al., 2018). In this study, we utilize ERA-20C (see Section 5.2.1). The accuracy of reanalyses depends on the model representation of the physical processes and on parameter schemes for processes which happen at scales below the model resolution. Furthermore, the accuracy of the assimilated observations and the abundance of observations plays a role as well as the accuracy of the prescribed forcing data, e.g. sea ice concentration, sea surface temperature, aerosols, etc. Additionally, complex orography with large variability on scales not resolved by the model can decrease the accuracy. Therefore, the accuracy of a reanalysis typically varies in space and time, so that a validation for the particular study area and time period is inevitable to determine the suitability for an application.

Integrated water vapor (IWV) is the moisture related variable which can be observed most accurately from satellite measurements compared to liquid water path or precipitation as the latter show much higher spatiotemporal variability. Thus, due to its role as the major storage term in the atmospheric water cycle and its good measurability, we focus this study on IWV.

Here, we present a two fold study, which firstly investigates the capabilities of ERA-20C to represent IWV in the Atacama region and secondly provides an analyses of the 20th century IWV. The paper is structured as follows. In Section 5.2, we describe ERA-20C, and the satellite data products which are utilized for a comparison. Section 5.3 presents the validation of the reanalysis by comparing the different satellite products for four different regions. In Section 5.4, we discuss the spatiotemporal variability of the IWV over the entire time period

of 111 years covered by ERA-20C and its relation to large scale indices such as ENSO and PDO. Finally, Section 5.5 concludes the study.

5.2 DATA AND FOCUS REGIONS

5.2.1 *ERA-20C*

The ECMWF reanalysis ERA-20C (Poli et al., 2016) provides a consistent meteorological record spanning 111 years (1900-2010) with a horizontal resolution of about 125 km. Therefore, for this study area, data on a $1.25^\circ \times 1.25^\circ$ longitude by latitude grid are investigated. It is based on ECMWF's weather forecasting system IFS cy38r1 and run with time varying prescribed forcing data such as sea ice concentration, sea surface temperature (SST), solar radiation, tropospheric and stratospheric aerosols, ozone and greenhouse gases. To approximate the concurrent synoptic conditions, the assimilation of surface pressure and marine surface wind are essential. To avoid introducing break points and trends in the representation of the atmosphere, observations which are only available for more recent years, such as vertical profiles and satellite observations, were left out. Atmospheric reanalyses typically assimilate vertical humidity profiles from radiosonde data. Humidity from radiosondes is usually not homogenized, so that artificial trends of water vapor can be introduced into these reanalyses (Dai et al., 2011) among other complications (Elliott and Gaffen, 1991). Since ERA-20C does not consider radiosonde data, it is not affected by any of such inhomogeneities.

Poli et al. (2016) compared integrated water vapor (IWV) spatially averaged over the tropical oceans (20°S–20°N) from ERA-20C to two observational products, i.e. the Remote Sensing Systems version 7 IWV product (RSS; Wentz (2013)) and the Hamburg Ocean Atmosphere Parameters and Fluxes from Satellite data record version 3.2 (HOAPS; Fennig et al. (2012)). Their resulting time series reveal a dry bias of ERA-20C of approximately 2 to 3 $\mathrm{kg\,m^{-2}}$. However, when anomalies with respect to a long term mean are considered, ERA-20C IWV agrees better with the observations than the Japanese 55-year Reanalysis (JRA-55) and the widely used ERA-Interim from ECMWF. Therefore, it seems well suited for the analysis of climate variability.

5.2.2 *HOAPS4*

In order to validate IWV values in ERA-20C, we utilize IWV estimates from the most recent version 4 of the Hamburg Ocean Atmosphere Parameters and Fluxes from Satellite data record (HOAPS; Andersson et al. (2017)). This data record provides satellite based retrievals of IWV as 6-hourly composites on a $0.5^\circ \times 0.5^\circ$ longitude by latitude grid for the period between July 1987 to December 2014. The retrievals are

derived from Special Sensor Microwave/Imager (SSM/I, Hollinger et al., 1990) measurements. They are only provided over ice free ocean where the microwave emission by water vapor can be separated well from the surface signal. By utilizing radiation in the microwave spectrum, retrievals are possible for all sky conditions except for heavy precipitation which can lead to strong scattering.

Generally, SSM/I based data records, such as HOAPS, provide IWV retrievals of similar quality regarding stability and homogeneity whereas non-SSM/I data records contain relatively large break points which coincide with changes in the observational set up. This can lead to different trend estimates from such records which are not in line with theoretical expectations (Schröder et al., 2016; Andersson et al., 2017b).

5.2.3 *MODIS*

To allow an evaluation of ERA-20C IWV over land, we further utilize IWV retrievals from the Moderate Resolution Imaging Spectroradiometer (MODIS) which is installed aboard the Terra satellite and the Aqua satellite. The MOD05_L2 (on Terra; Borbas et al., 2017a)) and the MYD05_L2 (on Aqua; Borbas et al., 2017b) products provide IWV retrievals which are derived from near infrared (NIR) and infrared (IR) channels for both satellites at 1 km (NIR) and 5 km (IR) horizontal resolution. Here, we utilize collection 6.1 which is the newest MODIS collection. We only use the near infrared retrievals because they show higher accuracy than the infrared retrievals (Steinke et al., 2015). These are available for surfaces which are highly reflective in the near infrared such as clear land areas, clear ocean areas with sun glint, or clouds above ocean or land. However, if the reflector is a cloud, the retrieved IWV is not representative of the entire atmospheric column because the water vapor path between the surface and the cloud is not fully captured. For the calculation of spatial means of the MODIS IWV, we omit scenes for which the reflector type is a cloud according to the MODIS water vapor product. A potential "clear sky" bias is assumed to be negligible considering the reflector type is a cloud for only 7 % out of of all retrievals for the inland regions.

MODIS retrievals are available from February 2000 (Terra) and from July 2002 (Aqua) to present. For the climatology study (Section 5.3.2, we only utilize Terra MODIS to benefit from the longer data record. The 10-year period between 2001 and 2010 is considered. For the analysis of the variability over time (Section 5.3.3), we consider the 8-year period between 2003 and 2010 to benefit from additional Aqua MODIS retrievals. The equatorial overpass time of the satellites are around 10:30 (Terra) and 13:30 (Aqua) local solar time. The uncertainty for MODIS IWV retrievals typically ranges between 5 % and 10 % (Gao and Kaufman, 2003). A previous study revealed a spatially varying dry

bias between 1.6 mm and 3.5 mm of the MODIS NIR IWV compared to GPS derived IWV within the Atacama region (Remy et al., 2011). Furthermore, the authors report good agreement of the IWV variability (1.3 mm standard deviation between GPS and MODIS retrievals after a linear correction). A more recent study carried out over North America reveals a wet bias of the MODIS NIR water vapor product of 4.1 mm, an RMSE of 5.6 mm and coefficient of determination of 0.964 over land(He and Liu, 2019). Another study in a drier continental environment (Iran) revealed a wet bias of 2.4 mm, an RMSE of 3.4 mm and a correlation coefficient of 0.95 (Khaniani et al., 2020). Both of these studies used Global Positioning System IWV retrievals as reference and considered a one year period for their comparisons. Judging from these studies, the bias seems to depend on the study area. However, suitable representation of IWV variability is proven by all of these comparisons.

5.2.4 *Large scale indices ENSO and PDO*

The ENSO pattern is a variability in the climate system manifested in alternating cold and warm phases in the surface temperature of the eastern tropical pacific which has strong influence on weather all over the globe (e.g. Timmermann et al., 2018). During the warm phase (El Niño), above average sea surface temperatures close to the western coast of South America lead to locally decreased stability of the troposphere and changes of the large scale circulation. For instance, enhanced upper tropospheric westerlies due to a northward displacement of the Bolivian High lead to drier conditions in the Altiplano (Vuille, 1999). On the other hand, during the cold phase (La Niña) enhanced tropospheric stability over the southeast Pacific favors drier conditions at the southern coast of the Atacama (Houston, 2006). To quantify the ENSO state, various indices have been created. Here, we apply the commonly used Niño 3.4 index which represents the SST anomaly of the equatorial Pacific between 5°S–5°N and between 170°W–120°W. To calculate this index, we obtained SST data from https://www.esrl.noaa.gov/psd/gcos_wgsp/Timeseries/Data/nino34.long.data (last access: 20 May 2019) which provides spatially averaged SST values for the Niño 3.4 region with a monthly resolution. These data are sampled from the Met Office Hadley Centre's sea ice and sea surface temperature data set, HadISST1 (Rayner et al., 2003). The HadISST1 provides a long continuous SST time series (1870-present). To derive the Niño 3.4 index, we calculate the monthly SST anomalies. Only the years of our study period, which is determined by ERA-20C (1900-2010), are considered. The two different versions of HadISST for ERA-20C (Version 2.1) and the Niño 3.4 index (Version 1) are highly correlated ($r = 0.96$). Therefore, no noticeable implications on the analysis are expected.

In contrast to ENSO, which is focused on the equatorial Pacific, the PDO represents a major climate variability pattern centered over the midlatitude North Pacific basin. It has been defined as the principle component of the leading Empirical Orthogonal Function (EOF) of the SST anomalies (e.g. Mantua et al., 1997; Zhang et al., 1997) for the Pacific north of 20°N. Here, we apply the PDO index according to Mantua et al. (1997) which is based on the U.K. Meteorological Office Historical Sea Surface Temperature data set (Folland and Parker, 1990; Folland and Parker, 1995) for the years between 1900-1981 and on the Reynold's Optimally Interpolated SST from 1982 to present. The PDO index was downloaded from http://research.jisao.washington.edu/pdo/PDO.latest.

Furthermore, local SST anomalies which are determined from the ERA-20C SST as a spatial mean between 18.125°S–29.375°S and 75.625°W–71.875°W on a monthly resolution are considered as local SST "index". The chosen region lies in close proximity to the coast off the Atacama (black rectangle in Fig. 5.1b).

5.2.5 *Cut-off lows*

To evaluate the capabilities of ERA-20C to represent synoptic features which potentially impact IWV variability, we consider cut-off lows off the coast of the Atacama (Section 5.3.2). Therefore, we utilize the cut-off low data set derived by Reyers and Shao (2019). They define cut-off lows as a local minimum of the geopotential height in 500 hPa within the area ranging from 85°W to 70°W and from 30°S to 15°S (black rectangle in Fig. 5.2 c, f, i). Geopotential heights were taken from the ECMWF reanalysis ERA-Interim (Dee et al., 2011b). To derive a list of days which featured a cut-off low, the 12 UTC ERA-Interim output was considered. More details on cut-off lows in the Atacama can be found in Reyers and Shao (2019).

5.2.6 *Focus regions and local atmospheric water cycle*

A simplified water cycle for the Atacama region is depicted in Fig. 5.1. The maritime boundary layer (MBL) is fueled with moisture by evaporation from the Pacific. Turbulent mixing results in a vertically almost constant specific humidity within the MBL which has an average height between 800 m and 1100 m (Rutllant et al., 2003; Muñoz et al., 2011). Adiabatic cooling with increasing height causes the water vapor to reach saturation, so that condensation leads to cloud formation. Due to the location at the subsiding end of the Hadley circulation, the MBL is topped by a strong temperature inversion which prevents the exchange between the MBL and the free troposphere above. Therefore, the cloud top height is limited to the base height of the temperature inversion. Longwave radiative cooling at cloud top maintains the stratocumulus or leads to further expansion. During the day, the surface

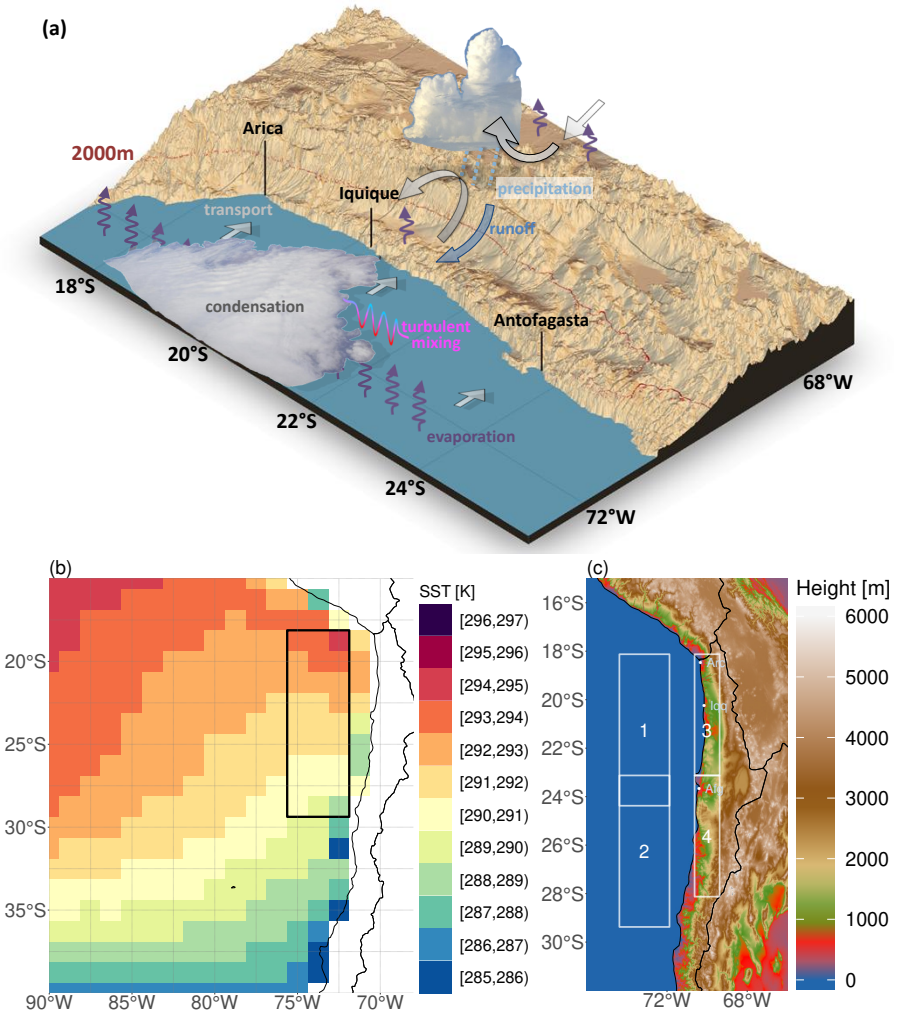

Figure 5.1: (a) Schematic depiction of the simplified atmospheric water cycle for northern Chile and its surrounding area. Over land, day time conditions are reflected. Strong heating of the Andean slopes leads to a rising branch of the flow in the high Andes and a return flow at around 5000 m which causes additional subsidence above the coastal region. Below this upper circulation cell, another circulation cell typically establishes in the maritime boundary layer. Here, early night time condition are depicted. A stronger cooling of the coastal cliff compared to the ocean surface causes a weak flow towards the coast so that the stratocumulus clouds can spread towards the cliff and penetrate the coastal mountain range. (b) Mean ERA-20C SST averaged between 1900–2010. (c) Topography derived from the Shuttle Radar Topographic mission (SRTM) data set. The white rectangles denote the four focus regions which are studied. Their sizes are given by the size and position of the ERA-20C grid boxes (1.25° × 1.25°). Regions 1 (ocean-N) and 2 (ocean-S) span 5 × 2 grid boxes, regions 3 (land-N) and 4 (land-S) span 4 × 1 grid boxes.

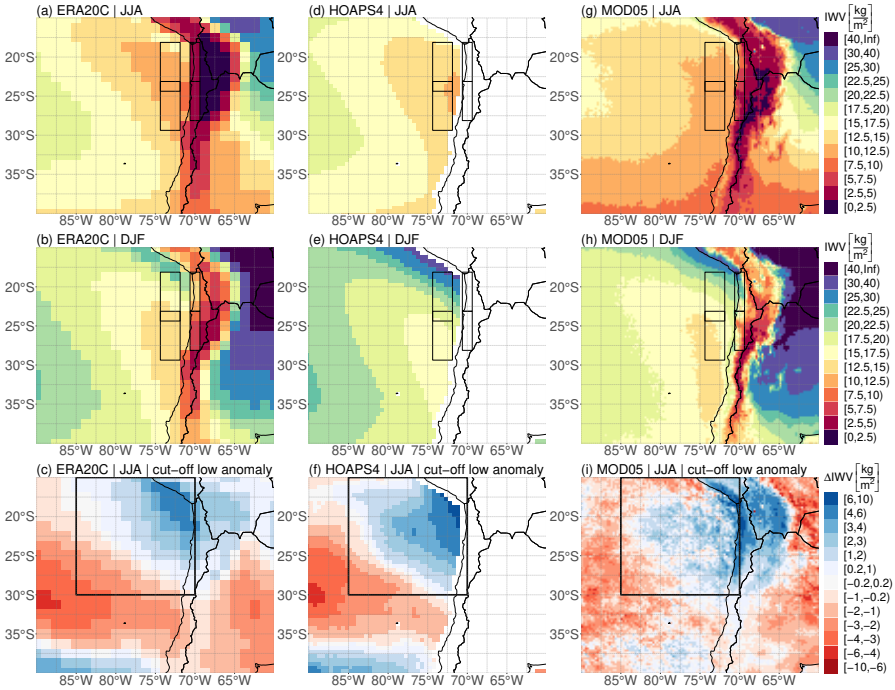

Figure 5.2: Composite means of IWV retrieved from ERA-20C (a,b,c), HOAPS (d,e,f) and Terra-MODIS (g,h,i) for winter season (JJA, top row), summer season (DJF, middle row) and composite anomalies of IWV for days for which the presence of a cut-off low was identified in ERA-Interim (bottom row). Here, cut-off lows are defined as a local minimum of the geopotential height in 500 hPa (Reyers and Shao, 2019) within the region from 85°W to 70°W and 30°S to 15°S which is denoted by the black rectangle (c,f,i). Composites are taken for a 10-year period (2001–2010).

at the coast and coastal cliff warm more than the ocean water leading to ascending motion at the cliff. This drives a circulation with a flow towards the coast above the surface and a return flow beneath the inversion base height, causing the stratocumulus cloud to be advected away from the coast. At night, though less pronounced, the circulation is reversed. Along with an even more effective radiative cooling, the strocumulus cloud deck can then penetrate the coastal cliff area. Even stronger heating of the Andean slopes during the day creates a strong inland flow above the coastal cliff which end in an upward motion in the high Andes creating convective clouds and precipitation. A return flow at higher altitudes leads to increased subsidence over the ocean and coastal area. Therefore, two circulation cells are established (Rutllant et al., 2003). In case the top of the coastal cliff, which has altitudes ranging between 400 m and 1500 m, is below the inversion base height, the MBL cell and the cell atop are coupled and Pacific moisture can be transported inland. When the stratocumulus is advected inland the mixing with warmer and drier inland air leads to dissipation of the cloud. However, the air with enhanced water vapor can still be

transported further inland where night time cooling might lead to the formation of radiation fog (Cereceda et al., 2002).

The inland regions are characterized by fog at the coastal cliff and coastal cordillera which typically forms at night and dissipates during the morning hours after sunrise (Cereceda et al., 2008b). Further inland, clouds are rarely observed. Due to these differences, we distinguish between inland and ocean for the validation. Furthermore, inland precipitation decreases from south to north in particular during winter season (Houston, 2006) which indicates the influence of the midlatitude storm tracks in the southern part. The stronger influence of frontal systems in the southern part calls for an additional distinction in northern and southern regions. Therefore, we focus on four individual regions (Fig. 5.1). Region 1 (18.125°S–24.375°S) and 2 (23.125°S–29.375°S) are over ocean (74.375°W–71.875°W), and region 3 (18.125°S–23.125°S) and 4 (23.125°S–28.125°S) are mainly over land (70.625°W–69.375°W).

5.3 VALIDATION OF ERA20C

To estimate how well IWV is represented in ERA-20C, we carry out a comparison with HOAPS and MODIS IWV. By choosing the Terra-MODIS IWV, a 10-year period (2001–2010) with retrievals from all three data sets can be compared. Average IWV values for each data record and each of the defined focus region are given in Tab. 5.1 revealing a dry bias of ERA-20C between 18 % (north) and 14 % (south) for the ocean regions (compared to HOAPS) and between 45 % (north) and 46 % (south) for the inland regions (compared to Terra-MODIS). Potential origins of this bias are discussed in section 5.3.1. Henceforward, spatial patterns are discussed for composites and seasonal means of the IWV (Section 5.3.2) followed by an analysis of decomposed IWV time series (Section 5.3.3).

5.3.1 *Bias assessment*

Since a substantial bias between the satellite data sets and ERA-20C is evident from the climatologies (Tab. 5.1), we investigate whether the reasons of this bias potentially hamper ERA-20C's suitability for the assessment of long term IWV variability within the study region. Several factors can contribute to the bias, e.g. instrumental factors, such as omitting cloudy scenes for the MODIS NIR retrievals, or problems inherent to the reanalysis.

Starting with the ocean regions, we approximate the expected IWV by the following conceptual considerations. Assuming a well mixed boundary layer with a vertically constant specific humidity and a negligible water vapor content within the free troposphere aloft, then the IWV is only a function of SST and the inversion height. The specific

Table 5.1: 10-year (2001–2010) mean IWV values for the focus regions indicated in Fig. 5.1 for all months, austral winter months, i.e. June, July, August (JJA), and austral summer months, i.e. December, January, February (DJF). Values are given in $kg\,m^{-2}$.

		ocean		inland	
	season	north	south	north	south
HOAPS	all	17.46	14.65		
	JJA	13.38	12.91		
	DJF	22.21	16.95		
Terra MODIS	all	15.99	12.96	14.68	8.90
	JJA	12.26	11.63	9.23	5.66
	DJF	18.57	15.17	20.87	12.90
Terra MODIS land only	all			13.9	8.5
	JJA			7.1	5.2
	DJF			20.8	12.2
ERA-20C	all	14.33	12.55	8.13	4.83
	JJA	12.00	11.77	4.52	3.75
	DJF	17.15	13.78	12.91	6.60

humidity is approximated, assuming a relative humidity of 80 % at the surface. According to ERA-20C, the mean SST for the northern ocean region is approximately 22°C and 17°C for the summer and winter season, respectively. According to vertical profiles acquired by radio soundings near Antofagasta (70.4°W; 23.4°S), the inversion base height ranges between 900 m (winter) and 1100 m (summer) (Muñoz et al., 2011). Towards the northwest, the inversion base height is increasing (Rahn and Garreaud, 2010), so that inversion base heights about 100–200 m above the heights at Antofagasta appear realistic. Such SSTs and inversion base heights amount to theoretical IWV values of $20\,kg\,m^{-2}$ and $12\,kg\,m^{-2}$ for summer and winter season, respectively (cf. Fig. A.5). Considering the coarse resolution of the utilized SST which might cause an underestimation of the strong SST gradient at the northern tip of the focus region and a missed component of the free troposphere, these theoretical values constitute a lower bound of the expected IWV. In this light, the slightly higher HOAPS IWV appears realistic while ERA-20C underestimates the IWV.

Judging from the ERA-20C temperature profiles within the lower troposphere, an underestimation of the inversion height of about 50 % is revealed compared to results from the Variability of the American Monsoon Systems Ocean-Cloud-Atmosphere-LandStudy Regional Experiment (VOCALS-REx; Wood et al., 2011; Rahn and Garreaud, 2010) (cf. Fig. A.5a). Such underestimations near the coast of subtropical

stratocumulus regions are typical for numerical models (Rahn and Garreaud, 2010; Hannay et al., 2009; Wyant et al., 2010). For an inversion height of about 600 m, the theoretical boundary layer contribution to the IWV amounts to about 7 $kg\,m^{-2}$ and 10 $kg\,m^{-2}$ for winter and summer season, respectively (cf. Fig. A.5). Furthermore, for the time period in the case study, it was found, that about 20 % of the total IWV are contributions from between height levels between the modeled and the observed inversion base height, partially compensating for the missing moisture below the inversion base. Therefore, we conclude that the dry bias originates from an underestimation of the inversion base and a false representation of the vertical humidity profile within the boundary layer.

For the two inland regions, we investigate the bias between ERA-20C and the MODIS near-infrared retrievals. While Remy et al. (2011) reveal a dry bias of the MODIS NIR IWV, ERA-20C shows even lower values in our study. This bias can be partially attributed to the coarse height representation in ERA-20C (Fig. A.6). The average height of the northern inland region is about 1297 m in ERA-20C which is about 342 m above the height according to Shuttle Radar Topography Mission (SRTM; Farr et al., 2007) data (955 m). For the southern inland region the overestimation is about 191 m (ERA-20C: 1778 m; SRTM: 1587 m). An approximation of specific humidity of 4 $g\,kg^{-1}$ (JJA) and 9 $g\,kg^{-1}$ (DJF) yield a theoretical difference in IWV of 1.5 $kg\,m^{-2}$ (JJA) and 3.3 $kg\,m^{-2}$ (DJF) for the northern region and 0.8 $kg\,m^{-2}$ (JJA) and 1.8 $kg\,m^{-2}$ (DJF) for the southern region. However, in particular for the northern region a higher bias is expected because the northern tip of the region has an actual even lower altitude allowing the penetration of moist air from the MBL inland more frequently which is manifested in a higher fog and low cloud occurrence in this region. This contribution to the region wide mean IWV is not represented in ERA-20C due to its coarse resolution. The same holds true for coastal strips and valleys such as the Rio Loa valley which are not represented in ERA-20C.

According to the performed case study including theoretical considerations, the bias between ERA-20C and the satellite-based IWV can be attributed mainly to the reanalysis. The reasons are a false representation of the MBL (ocean) and topography (inland). Both effects constitute systematic errors and thus should not interfere with the representation of temporal variability. Composite IWV anomalies for cut-off low situations show similar magnitudes and spatial patterns for both ERA-20C and the satellite-based data sets, demonstrating the ability to capture IWV variations on a synoptic scale (Section 5.3.2).

5.3.2 *Climatologies and composites*

Across a wider area, the seasonal means over this 10-year period reveal a meridional pattern over ocean with higher IWV to the west and lower

IWV closer to the coast for all three data sets (Fig. 5.2). This can be attributed to the lower SST in closer proximity to the coast where the Humboldt current is strongest (Fig. 5.1). During the summer season, i.e. December, January, February (DJF), IWV observed by all three data sets is higher than during winter season, i.e. June, July August (JJA) (Tab. 5.1). While HOAPS shows an increase from winter to summer of about 66 % and 31 % for the northern and southern regions, respectively, the increase in ERA-20C is about 43 % and 17 %, respectively. Again, a higher SST causes enhanced evaporation so that IWV increases in particular over ocean. The southeast Pacific anticyclone typically shifts towards the south in the summer season causing the surface winds which drive the Humboldt current to shift south as well. This way, the SST reaches maximum values in particular ahead of the Peruvian coast (15°S–20°S). Enhanced IWV is found for this region in all three data sets. Aside from higher SSTs, the summer season is also characterized by higher air temperatures providing a higher water vapor holding capacity and advection of already water vapor enriched air. Accordingly, MODIS and ERA-20C show an increased IWV over land during the summer season. Between 20°S–26°S, cloud heights are typically higher during the warm season (Böhm et al., 2019) which indicates a higher inversion base height. This could enable more Pacific moisture to cross the coastal cliff and be transported further inland leading to increased IWV in this region. On the eastern side of the Andes, a zonal pattern is revealed by ERA-20C and MODIS IWV with maximum values over the Amazon Basin and decreasing values towards the midlatitudes. Overall, ERA-20C is able to represent the patterns and seasonal changes which are observed by the satellite based data sets.

Furthermore, we investigated the influence of cut-off lows on the spatial distribution of IWV. Reyers and Shao (2019) found a higher moisture availability in the Atacama coinciding with these synoptic features. In this study, composite anomalies for cut-off low situations during the winter season show an enhanced IWV for the Atacama and for an extended area over the ocean off the Atacama coast (Fig. 5.2 c, f, i). In addition to the previously discussed agreement between the IWV patterns on a climatological scale, ERA-20C is even capable of representing the pattern of enhanced IWV for these synoptic features.

5.3.3 *Comparison of decomposed IWV time series*

To obtain time series for each data record, we calculate spatial and monthly means for each of the four focus regions. In order to evaluate the ability of ERA-20C to represent the IWV variability, we investigate the IWV deviation from the respective temporal mean of each time series. Furthermore, we decompose the time series into a 12-month centered moving average (CMA) a seasonal component and the resid-

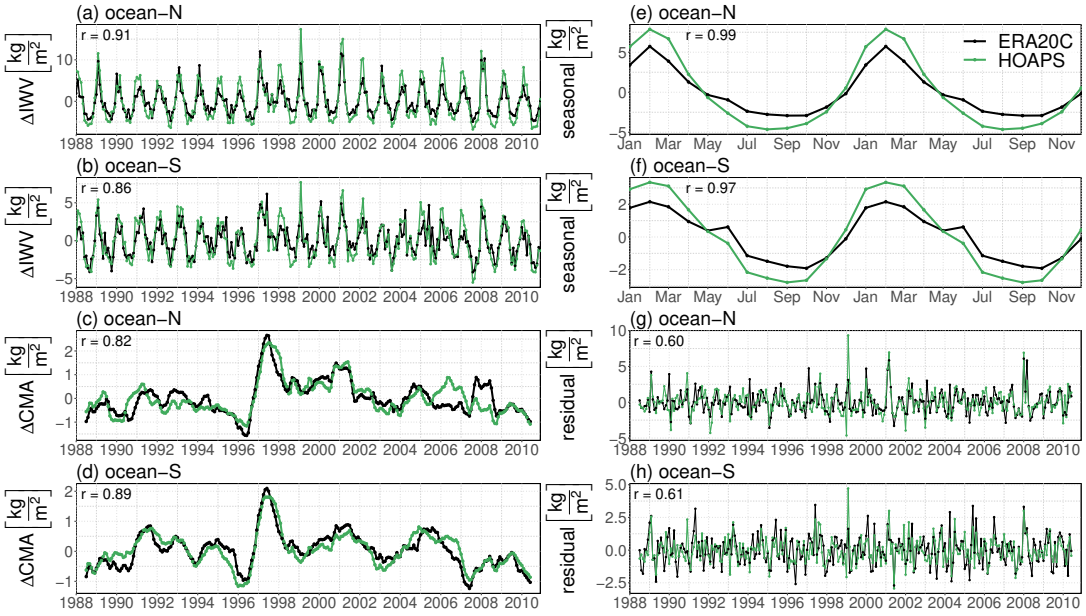

Figure 5.3: IWV time series of ERA-20C and HOAPS for focus regions 1 and 2 (over ocean, denoted in Fig. 5.1). (a,b) Deviations from the respective means. (c,d) 12 months centered moving average of the ΔIWV time series. (e,f) seasonal cycle. (g,h) residuals. Correlation coefficients *r* are denoted at the top left of each panel. By adding the centered moving average, the seasonal cycle and the residual together, the overall time series of the deviation of the IWV results.

ual. The seasonal component is derived by subtracting the 12-month CMA from the time series and then calculating the mean value for each month. Subtracting the 12-month CMA and the seasonal component from the time series yields the residuals. While the 12-month CMA represents the year to year variability, the residuals are a measure of the month to month variability.

The unfiltered monthly time series of IWV for ERA-20C and HOAPS between 1988 and 2010 are highly correlated due to a consistent representation of the annual cycle and year to year variability (1-year CMA) for both the northern and the southern ocean regions (Fig. 5.3). The month to month variability (residuals) yield slightly lower correlations. A reason could be that synoptic features which pass through the regions might be slightly offset in time and space. The strong 1997 El Niño event is captured is equally apparent in both data sets. Only a small period (2005–2008) of ongoing disagreement between HOAPS and ERA-20C is apparent for the northern ocean region (1-year CMA, Fig. 5.3c).

The short period of disagreement and offset of the CMA between 2005 and 2008 within the northern region becomes a more dominant feature for the comparison between ERA-20C and MODIS IWV because the overlapping time period is much shorter (2003–2010). A missed peak during 2006 is followed by an offset period between 2007

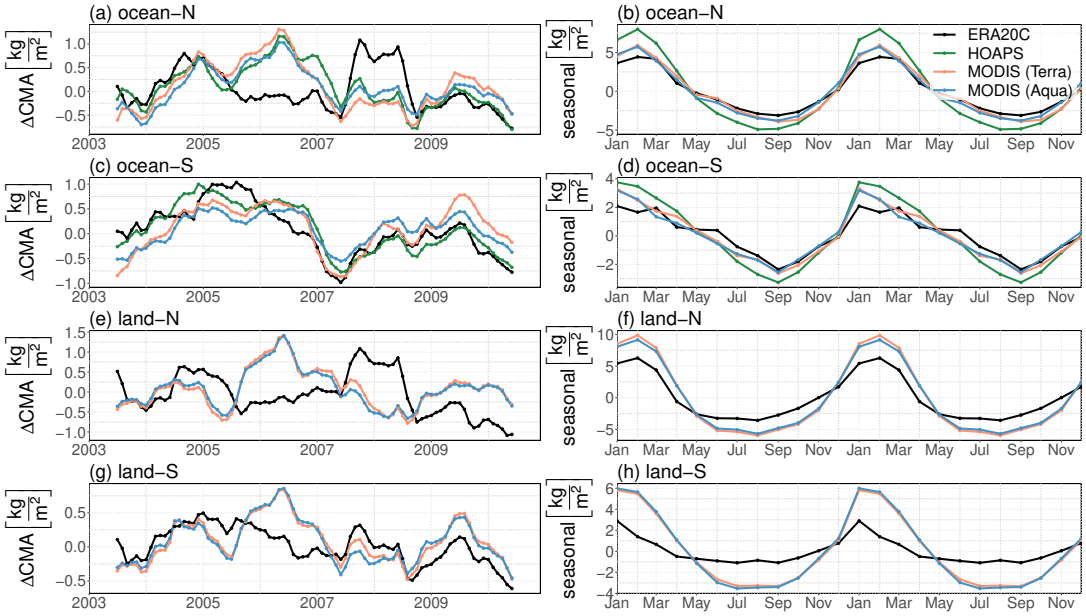

Figure 5.4: IWV time series of ERA-20C, HOAPS, Terra-MODIS and Aqua-MODIS for region 1 (a,b), region 2 (c,d), region 3 (e,f), and region 4 (g,h) which are indicated in Fig. 5.1. Shown are the 1-year centered moving average (CMA) of the IWV anomalies (left) and the seasonal components (right).

and 2008 which is apparent in the 1-year CMA for both the northern ocean and inland region (Fig. 5.4). Before and after this period, ERA-20C IWV is in agreement with HOAPS and MODIS regarding both amplitude and phase. However, this short period of disagreement which is less pronounced for the southern regions results in overall low correlations of the 1-year CMA (Fig. 5.5 c, d). The month to month variability (residual) shows similar agreement between ERA-20C and MODIS (Fig. 5.5 e, f) as for the comparison between ERA-20C and HOAPS for the longer period. This indicates that ERA-20C is able to represent synoptic features although a substantial noise is present. This is expected given the coarse resolution of the model.

The phase of the seasonal cycle is captured accurately by ERA-20C which is indicated by correlations reaching almost unity for all regions 5.5 a, b). The amplitude of the seasonal cycle is underestimated by ERA-20C, when absolute numbers are considered (Tab. 5.2). However, relative to its overall mean, the amplitudes are closer to the satellite-based estimates. These discrepancies are consistent with the detected misrepresentation of the maritime boundary layer (ocean) and topographic heights (inland) which result in an overall dry bias of the ERA-20C IWV (see Section 5.3.1).

The seasonal cycle is the main driver of variability according to the amplitudes (Tab. 5.2). Differences between summer and winter peaks range between 7.0 $\mathrm{kg\,m^{-2}}$ (HOAPS, southern ocean region) and 12.9 $\mathrm{kg\,m^{-2}}$ (HOAPS, northern ocean region). The month to month

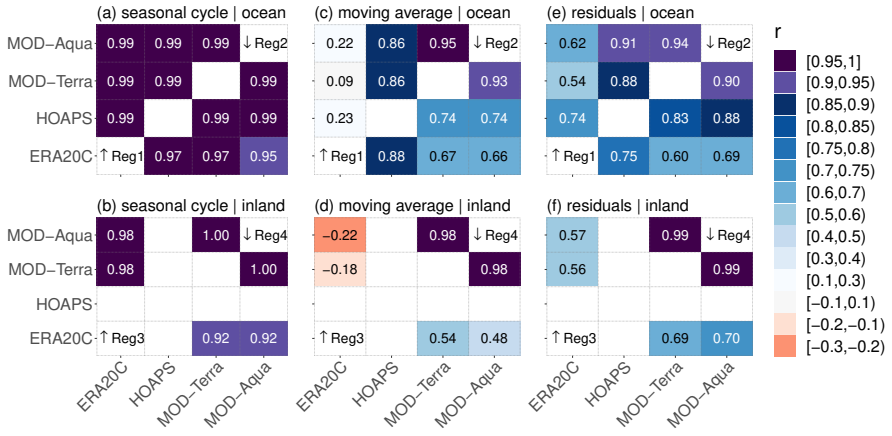

Figure 5.5: Correlation of the decomposed time series of IWV between the different sources: Aqua MODIS NIR (MOD-Aqua), Terra MODIS NIR (MOD-Terra), HOAPS, and ERA-20C. (a,b) correlation of the seasonal cycle. (c,d) correlation of the centered moving average. (e,f) correlation of the residuals. Regions 1 and 2 (a,c,e) and 3 and 4 (b,d,f) are distinguished by showing the northern region in the upper left corner and the southern region in the lower right corner of each panel as indicated. The underlying time series comprise 8 years (2003–2010). Note that HOAPS is only available over ocean.

variability (residuals) superimposes another $5\,\mathrm{kg\,m^{-2}}$ (north) and $3\,\mathrm{kg\,m^{-2}}$ (south) of variability over ocean according to the difference between most upper and most lower peaks (c.f. Fig. 5.3, g, h). Similar values are observed for the inland regions with about $4\,\mathrm{kg\,m^{-2}}$ (north) and $3\,\mathrm{kg\,m^{-2}}$ (south).

A measure of uncertainty of the IWV representation in ERA-20C can be derived by calculating the root mean square error (RMSE) for the 3 components (seasonal cycle, moving average and residuals) of the IWV anomalies (Fig. 5.6). For the moving average, the RMSE between HOAPS and ERA-20C is about $0.56\,\mathrm{kg\,m^{-2}}$ and $0.25\,\mathrm{kg\,m^{-2}}$ for the northern and southern ocean regions, respectively. In relation to the long term mean values (Table 5.1) these values range around 3 % and 2 %. For the residuals, the RMSE between HOAPS and ERA-20C is about $1.09\,\mathrm{kg\,m^{-2}}$ and $0.80\,\mathrm{kg\,m^{-2}}$ for the northern and southern ocean regions, respectively. In relation to the long term mean values (Table 5.1) these values range around 6 % and 5 %.

An extension of the analysis beyond the four focus regions to a wider area reveals that the RMSE values are highest in the northern part of the study area where the variability is found to be highest (Fig. 5.7 e, f, g, h for ERA-20C against HOAPS and Fig. 5.8 e, f, g, h for ERA-20C against Terra-MODIS). While higher correlations regarding to the moving averages and the residuals are revealed in the south, higher correlations regarding to the seasonal cycles are revealed in the north (Fig. 5.7 a, b, c, d and Fig. 5.8 a, b, c, d).

Table 5.2: Assessment of the seasonal cycles for the focus regions indicated in Fig. 5.1. Differences between summer and winter peaks $P_S - P_W$ for the seasonal cycles shown in Fig. 5.4 and long term mean IWV w (c.f. Tab. 5.1) are given in $\mathrm{kg\,m^{-2}}$. The ratio between half of the peak difference and the long term mean IWV is given in percent.

		ocean		inland	
	data	north	south	north	south
$P_S - P_W$	ERA-20C	7.6	4.4	9.8	4.0
	HOAPS	12.9	7.0		
	Terra MODIS			15.8	9.1
mean IWV w	ERA-20C	14.33	12.55	8.13	4.83
	HOAPS	17.46	14.65		
	Terra MODIS			14.68	8.90
$\frac{\frac{1}{2}(P_S - P_W)}{w}$ [%]	ERA-20C	26.5	17.5	60.3	41.4
	HOAPS	36.9	23.9		
	Terra MODIS			53.8	51.1

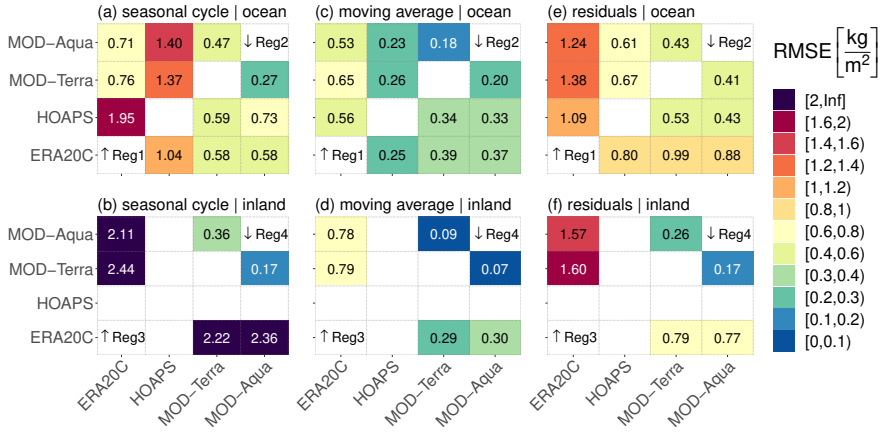

Figure 5.6: Same as Fig. 5.5 but for root mean square error (RMSE).

5.3.4 *Representativity for the 20th century*

Thus far, the validation is carried out for a time period for which satellite-based IWV retrievals are available. To further assess how representative the validation is for the whole time period, we analyze the observations which were assimilated into ERA-20C. We utilize the number of assimilated surface pressure and marine surface wind observation along with their individual departures from the analyses after their assimilation. These data are provided by the ECMWF observation feedback archive (Hersbach et al., 2015). We consider any assimilated observations that are located within the region bounded by 90°W to 60°W and 40°S to 15°S. The total number of assimilated

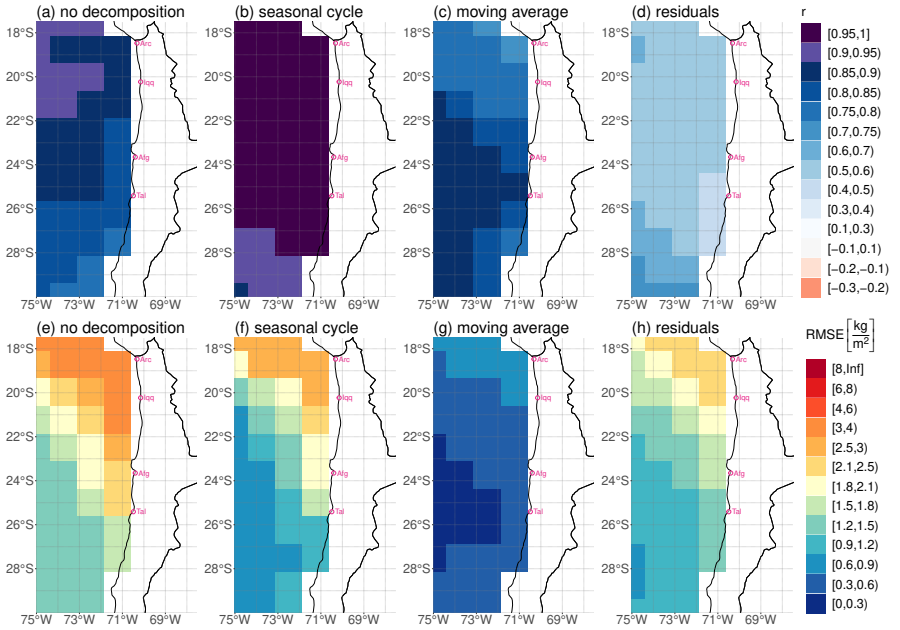

Figure 5.7: Spatial distribution of correlation r (a–d), and RMSE (e–h) between HOAPS4 and ERA-20C on a 1.25° × 1.25° grid for a 23-year period (1988–2010). Shown are the parameters for the time series with no decomposition, the seasonal cycle, the moving average, and the residuals (left to right).

observations per year stayed below 5000 prior to the late 1950s except for a local maximum around 1910 (Fig. A.7a). During the 1970s this number increased by a factor of greater than 25. The enormous increase is mainly due to additional observations over land (Fig. A.7b). On a global scale, the number of assimilated observations increased continuously despite a few interruptions (e.g. after World War II) (Poli et al., 2016). Concurrently, the mean bias between ERA-20C and the assimilated observations decreased to almost zero by the late 1990s (Fig. A.7b). The bias corrected mean absolute error decreased from about 1 hPa to at the beginning of the modeled period to about 0.5 hPa at the end of the period (Fig. A.7b). This is not surprising, considering a denser observation network can add more realism to the model. This means, the uncertainties are higher for the earlier part of the time series, in particular prior to 1935. However, the spatio-temporal variability of the bias does not reveal any systematic pattern in space or time (Fig. A.9). Therefore, there is no indication that the circulation pattern could be systematically offset for the earlier part of the 20th century.

As of today, no data with higher resolution than ERA-20C is available to study IWV for the entire 20th century. Overall, ERA-20C reveals a dry bias which has been shown already by Poli et al. (2016). However, the spatial distribution patterns agree well with satellite observations.

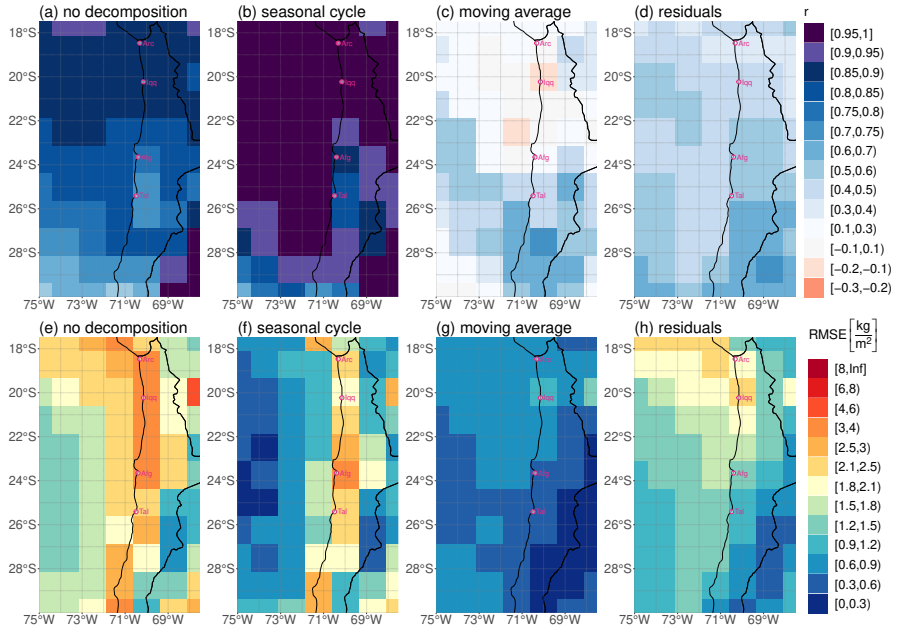

Figure 5.8: Spatial distribution of correlation r (a–d), and RMSE (e–h) between Terra-MODIS and ERA-20C on a $1.25° \times 1.25°$ grid for a 10-year period (2001–2010). Shown are the parameters for the time series with no decomposition, the seasonal cycle, the moving average, and the residuals (left to right). Here, the IWV from the 15 UTC output of ERA-20C is utilized.

Agreement of anomaly patterns for cut-off low situations even prove a good representation on a synoptic scale. The phase of the seasonal cycle is represented very accurately, even though the amplitude is underestimated. Year to year variability is found to be represented very well. However, a short time period was identified for which the moving averages of ΔIWV from ERA-20C diverged from the satellite based retrievals. Even for this period, the month to month variability is still in rather good agreement. Therefore, we conclude that ERA-20C represents the IWV variability sufficiently well for the Atacama and the bordering southeast Pacific region so that it can be utilized to study the IWV within this region over the course of the 20th century.

5.4 20TH CENTURY IWV

To investigate the long term IWV variability in the Atacama Desert, we consider the IWV deviations from the respective mean values for the entire period which is covered by ERA-20C (1900-2010) with a monthly resolution. First, the inter-annual to inter-decadal variability is discussed for the focus regions. Second, the relationship between the observed variability of IWV and the PDO, ENSO and local SST are investigated. Third, a seasonal distinction is shown.

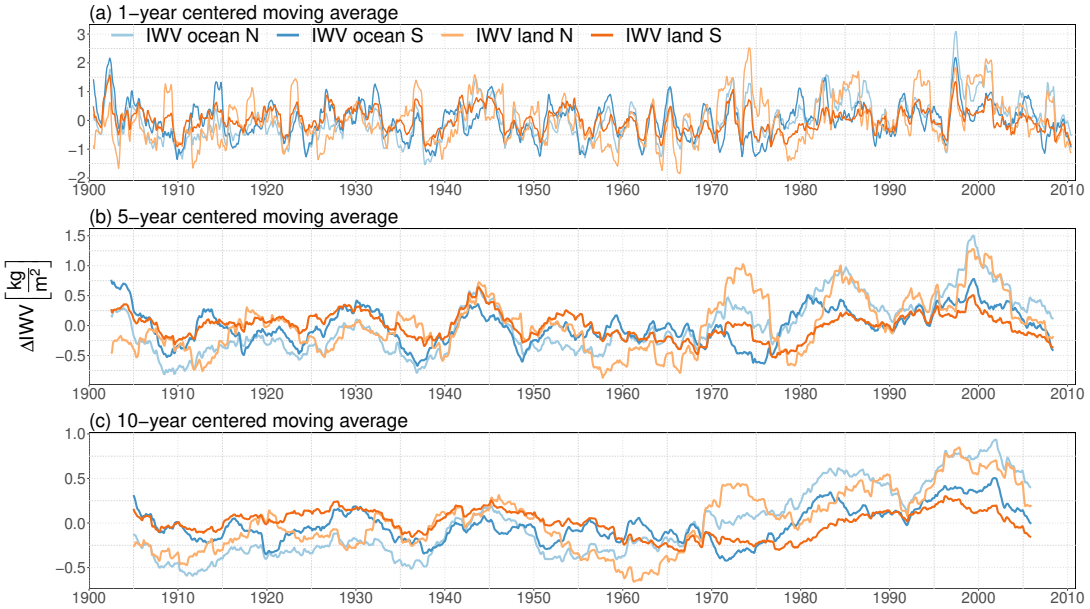

Figure 5.9: Time series of IWV anomalies from ERA-20C for the northern and southern ocean and inland regions indicated in Fig. 5.1. Shown are the 1-year (a), 5-year (b) and 10-year (c) centered moving averages between the years 1900 and 2010.

5.4.1 *Regional IWV variability*

The 1-year, 5-year, and 10-year centered moving averages (CMA) for each of the four designated focus areas show different IWV variations on different time scales (Fig. 5.9). For all regions, a substantial inter-annual to inter-decadal variability is observed. For the ocean regions, IWV relative variations between −11 % and +18 % (ocean S) to +22 % (ocean N) are detected. For the inland regions, even higher variations between −20 % (land S) to −24 % (land N) and +30 % (land S) to +31 % (land N) are detected. The absolute numbers of positive and negative deviations and the distribution of the 1-year CMA IWV (Fig. 5.10) show that wet extremes are less frequent but more pronounced. This means the dry state is closer to the mean condition. A similar effect is generally observed for the ENSO state with El Niño typically showing stronger departures from the mean compared to La Niña (see Fig. 5.11).

The highest peak within the 111-year record occurs for the northern ocean region during the 1997 El Niño. This El Niño yields a peak for all focus regions for the 1-year CMA and is still a prominent feature in the 5-year CMA. Thereafter, IWV has been decreasing in all regions until the end of the covered period. Therefore, compared to the late 1990s, the years around 2010 constitute a dry period. However, since the 1970s, IWV has been increasing in all four focus regions until the end of the 20th century (Fig. 5.9c). The latter coincides with a shift

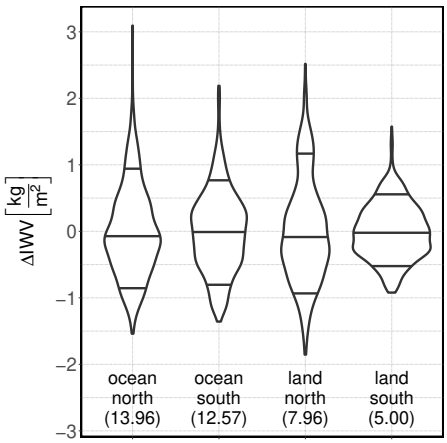

Figure 5.10: Violins of IWV anomalies based on monthly values of the 1-year centered moving average (time series shown in Fig. 5.9a) between 1900–2010 for the four focus regions indicated in Fig. 5.1. Horizontal lines within the violins represent the 10th, 50th and 90th percentile of the distribution. The overall mean value is given in parenthesis for each region in $\mathrm{kg\,m^{-2}}$.

from cold to warm phase of the PDO which will be discussed in more detail in Section 5.4.2.

Correlations between the individual regions (Fig. 5.12) range around 0.65 to 0.7 for the 1-year CMA indicating rather pronounced coupling between the regions. An exception is the northern inland region which correlates less to the southern inland region (0.58) and the southern ocean region (0.16). This indicates that the northern inland region responds differently to short term variability and is less connected in particular to the southern regions. On the other hand, the northern inland region shows higher correlations on longer time scales (10-year CMA) with the southern ocean region (0.47) and the northern ocean region (0.85). On shorter time scales, periods with more frequent wet episodes might still cause a distinct signal in the 1-year CMA which is smoothed on longer time scales. During summer, such wet episodes can occur in the northeastern Atacama (Altiplano) when easterly winds advect moist air from Amazon basin into the region. These easterlies appear in connection to a southward displacement of the Bolivian High which is a seasonal upper tropospheric high pressure system. It is present during summer with a mean climatological position above Bolivia (Garreaud et al., 2009). The origin of the moisture is the continental boundary layer in central South America. Diurnal heating of the eastern Andean slopes along with entrainment of easterly momentum from upper levels results in strong upslope transport of moisture (Garreaud, 1999; Garreaud et al., 2003). Enhanced summer precipitation (Vuille, 1999) and enhanced summer IWV (Marín and Barrett, 2017) over the Altiplano are associated with this synoptic pattern. While individual wet episodes connected to

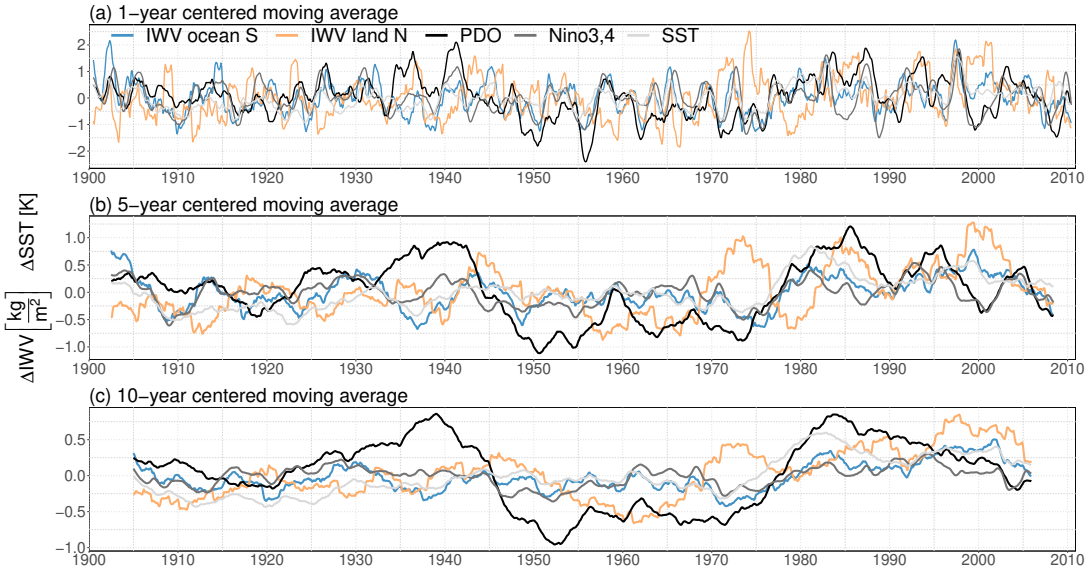

Figure 5.11: Same as Fig. 5.9 but only for the southern ocean and northern inland regions. Additional lines of different shades of gray represent the PDO index, Niño 3,4 index, and the local SST anomalies. Local SST is taken from ERA-20C averaged over the region ranging from 75.625°W to 71.875°W, and from 29.375°S to 18.125°S)

upper level easterlies typically last for a couple of days up to a week, their frequency of occurrence can vary on inter-annual time scales (Garreaud et al., 2003). Therefore, the moisture variability due the described mechanism could be visible in 1-year CMA of the IWV. However, the described impacts of a southward displacement of the Bolivian High are constraint to the northern part of the Atacama. This could be an explanation why the northern inland focus region might be less connected to the other focus regions, especially the southern regions. A still higher correlation with the northern ocean region could be an indication that the easterly moisture transport might even reach the Pacific. How this phenomenon relates to longer term climate variability (ENSO, PDO) is discussed in Section 5.4.2 and Section 5.4.3 with seasonal distinction.

5.4.2 *IWV relationship with ENSO, PDO and local SST*

To assess how the described variability of regional IWV in the Atacama and bordering southeast Pacific is connected to larger scale climate variations, we investigate its relationship to ENSO and the PDO. Furthermore, local SST anomalies determined from the ERA-20C SST (Section 5.2.4) are considered. ENSO, PDO and the local SST are all positively correlated (Fig. 5.12). This is not surprising since all these indices usually feature warm (cold) SST anomalies in the southeast Pacific during their warm (cold) phase. However, ENSO, PDO and

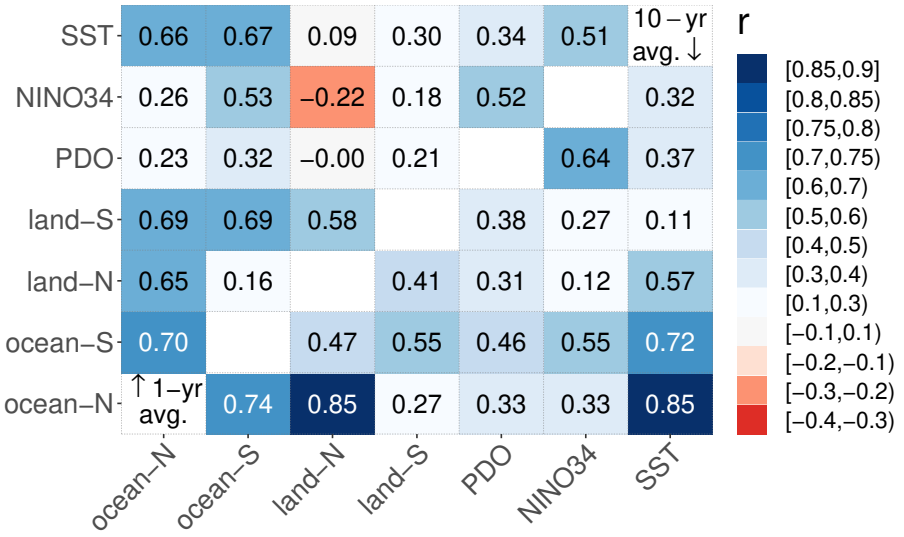

Figure 5.12: Correlations between the different time series, namely the ERA-20C IWV for 4 different regions (ocean and inland, north and south as indicated in Fig. 5.1), PDO index, Niño 3,4 index, and the local SST. Shown are results for a 1-year moving average (upper left) and a 10-year moving average (lower right). Correlation coefficients r are indicated by color and are given as numbers.

local SST act on different time and spatial scales, so that different relationships to the local IWV can be expected. In particular ENSO and PDO feature more pronounced teleconnection patterns compared to local SST. Therefore, a more complex relationship to local IWV can be expected for these two large scale features.

All indices feature mainly positive relationships with the IWV of the focus regions, indicating higher IWV during the respective warm phases. An exception is the northern inland region which does not show a correlation to local SST and PDO, but a slight negative correlation to ENSO ($r = -0.22$) for the 1-year CMA. This means, the region is drier during El Niño and wetter during La Niña. Previous studies have shown this effect for the Altiplano during the austral summer season mainly for precipitation (Vuille, 1999; Garreaud et al., 2003; Canedo-Rosso et al., 2019) but also for specific humidity (Vuille, 1999). During El Niño, which typically reaches its mature state during austral summer, the Bolivian High weakens and retreats to the north, so that pronounced upper level westerlies steer dry air towards the Altiplano (Vuille, 1999). On the contrary, during austral winter an opposite relationship has been reported. Enhanced ridging over the Pacific in higher latitudes and troughing over the southeastern Pacific result in a northward displacement of midlatitude disturbances yielding wetter winters mainly in the southern Atacama (Vargas et al., 2006; Houston, 2006). Extended northward troughing with anomalous northwesterly midtropospheric flow are also related to higher IWV over the Altiplano during austral winter (Marín and Barrett, 2017).

Therefore, opposite responses to the same phase of ENSO are expected for summer and winter season. This could explain generally low correlations of ENSO with the inland IWV in particular for the northern inland region. A more detailed analysis of seasonal effects is presented in Section 5.4.3.

The mechanism causing anticorrelation between IWV and ENSO during austral summer is more pronounced in the northern region. Therefore, the southern regions are not expected to show such strong seasonal dependence of the relationship. This is manifested in higher correlations between IWV and ENSO for the southern regions (Fig. 5.12). The PDO is expected to have similar impacts in the study area compared to ENSO but with smaller amplitude (Garreaud et al., 2009). Moreover, PDO and ENSO phases can have constructive interference (Andreoli and Kayano, 2005).

Correlations between indices and regional IWV are generally lower for the 1-year CMA compared the 10-year CMA (Fig. 5.12). In particular the PDO shows stronger correlations on an inter-decadal time scale compared to an inter-annual time scale. ENSO and to a lesser extent the PDO favor certain weather patterns but are not expected to directly cause them. Therefore, not every El Niño is associated with increased precipitation at the southern coast of the Atacama (Houston, 2006) or anomalous dry episodes in the Altiplano (Vuille, 1999; Garreaud et al., 2003). The variability is still related to particular synoptic patterns. For instance, an extreme precipitation event occurred in the Atacama in March 2015 which was associated with an El Niño. Bozkurt et al. (2016) demonstrate that only the interplay of the positive SST anomalies and a concurrent cut-off low enabled the enhanced moisture transport which ultimately fueled the unusual precipitation inland. Especially on shorter time scales, the synoptic variability influences the IWV resulting in lower correlations for these time scales.

For the southern inland region, the PDO explains more variability compared to ENSO and local SST. This raises the hypothesis that during the warm phase of the PDO the frequency of blocking situations in higher latitudes with an accompanied northward shift of midlatitude disturbances is increased. The latter pattern yields wetter conditions in the Atacama during austral winter and was found to occur more often during the developing stage of El Niño (Vargas et al., 2006).

Over ocean and also for the northern inland region, the local SST appears to have the strongest relationship with the IWV on an inter-decadal time scale compared to ENSO and PDO. This implies that the direct connection from an increased SST to a higher IWV is stronger than the more complex relationship with ENSO and PDO. An increased SST causes a warming and moistening of the MBL. An increase of the water holding capacity of warmer air by about 7 % per Kelvin is expected according to the Clausius-Clapeyron relation so that the IWV would most likely increase for a higher SST.

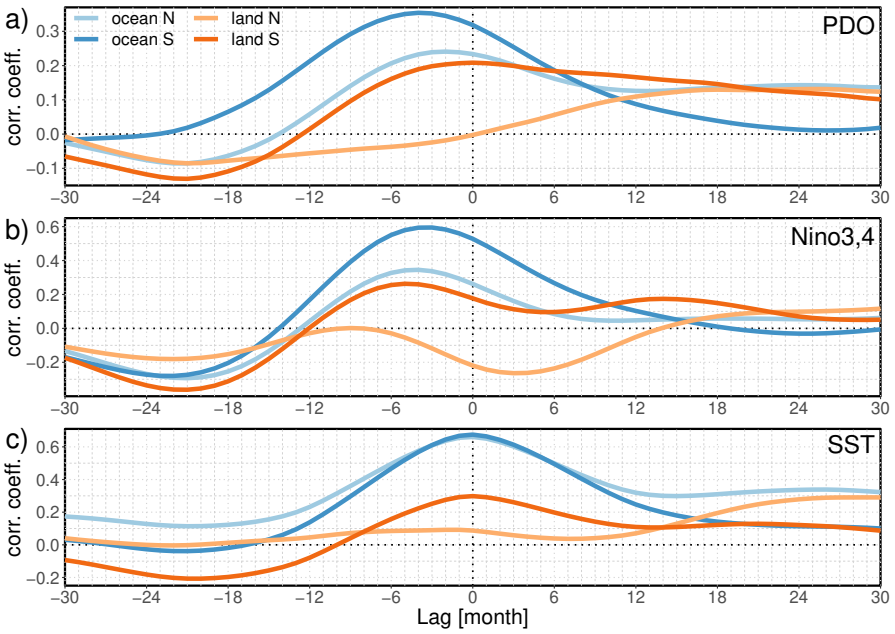

Figure 5.13: Correlation coefficients for lagged correlations between ERA-20C IWV and PDO (a), Niño3,4 index (b) and local SST (c) for the 1-year centered moving averages. Distinguished are the 4 focus regions (ocean north, ocean south, land North, land south). The lag k indicates that the correlation is calculated between the time series IWV_{t+k} and the respective index ind_t (PDO, Niño3,4 or local SST).

This direct relationship between local SST and IWV is also manifested by the results of lagged correlations. The highest correlation is yielded for lag 0 (Fig. 5.13c). ENSO and PDO yield higher correlations if the IWV is shifted back in time by 2 to 5 months (Fig. 5.13a and 5.13b). This supports the results from Vargas et al. (2006) who reported that the pattern of anomalous ridging over the Pacific in higher latitudes and troughing in subtropical latitudes is favored during the developing stage of El Niño. Therefore, enhanced wetness would be expected a few month prior to the fully developed El Niño.

5.4.3 *Seasonal dependencies*

To investigate the seasonal dependence, we calculate the normalized density of the monthly IWV distinguished between austral winter (JJA) and summer (DJF) season for each of the four focus regions. Further, we distinguish between different states of the Niño 3.4 index and the local SST by defining intervals with break points resulting from different quartiles of the respective index (Fig. 5.14). El Niño like conditions, i.e. the top quartile of the Niño 3.4 index, lead to above average IWV during winter (JJA, Fig. 5.14 a–d) and decreased IWV variability (narrower distribution) during summer (DJF, Fig. 5.14

e–h) with a shift to the dry end of the overall IWV range. La Niña like conditions, i.e. the lowest quartile of the Niño 3.4 index, lead to below average IWV during winter (JJA, Fig. 5.14 a–d) and only show a small difference compared to the neutral ENSO state (i.e. interquartile range of the Niño 3.4 index). For the northern inland and ocean region, la Niña allows slightly higher wet extremes of IWV compared to neutral or El Niño conditions during the summer. However, the drier conditions during El Niño are a lot more pronounced. A mechanism behind this could be enhanced upper tropospheric easterly wind flow during La Niña. Vuille (1999) demonstrate that a high southern oscillation index (SOI), which represents La Niña conditions, is associated with a more pronounced and southward shifted Bolivian high which leads to increased easterly winds over the Altiplano. On the other hand, El Niño conditions result in a weakened Bolivian High and stronger westerlies. Garreaud et al. (2003) also show that while enhanced easterlies strengthen the upslope flow of moist continental air on the eastern side of the Andes, weakened easterlies strengthen the upslope flow of dry desert air on the western side of the Andes. La Niña conditions, which are typically associated with intensified subsidence over the eastern Pacific, potentially trigger enhanced easterly airflow over the Altiplano and control the position of the Bolivian high. These moist upper tropospheric easterlies associated with La Niña conditions cause increased specific humidity even at Antofagasta (Vuille, 1999), a coastal city in the Atacama region.

Analogous results have been obtained for precipitation by Houston (2006). He reports that higher winter precipitation in the Atacama occurs during El Niño conditions, whereas enhanced summer precipitation occurs during La Niña conditions. The principle agreement of our findings with the results from Houston (2006) indicates the close relationship between IWV and precipitation. Thus, long term IWV retrievals might link to precipitation estimates for the past.

For the winter season, the effects of the local SST appear similar compared to the effects of the Niño 3.4 index. This means that higher local SSTs correspond to a shift towards higher IWV values whereas lower local SSTs correspond to a shift towards lower IWV values. For the summer season, only the ocean regions show a dependence on local SST. Moreover, only the upper quartile of the SST range appears to have an effect on the IWV, by shifting the distribution towards higher values.

5.5 CONCLUSION

We investigated the IWV within the Atacama Desert and the bordering southeast Pacific over the course of the 20th century. For IWV estimates, we utilized the reanalysis ERA-20C. In order to assess its suitability, we carried out a validation study first by comparing ERA-20C IWV to

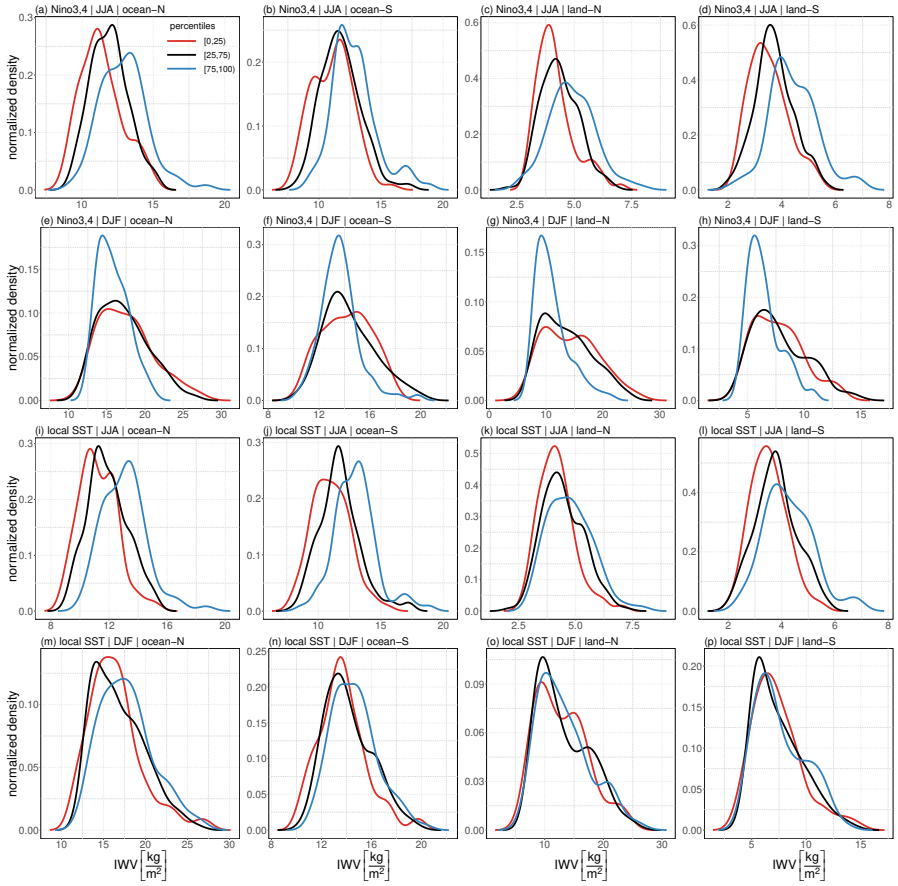

Figure 5.14: Composite normalized density of IWV retrieved from ERA-20C for the covered time period (1900-2010) for different conditions of the Niño 3,4 index (a-h) and of the local SST anomaly (i-p). The 4 focus regions are distinguished by 4 columns. Furthermore, winter season (JJA, 1st and 3rd row) and summer season (DJF, 2nd and 4th row) are shown separately. IWV densities are shown for lowest quartile (red), the inter-quartile range (black) and the highest quartile (blue) of the respective indices. This way, the red curves represent conditions typically associated with La Niña like conditions and the blue curve represent conditions typically associated with El Niño like conditions. The analysis is based on monthly values of the indices.

different satellite observations (HOAPS, MODIS). On a climatological scale, spatial and seasonal patterns are well reproduced by ERA-20C. Furthermore, a dry bias between $2\,\mathrm{kg\,m^{-2}}$ to $3\,\mathrm{kg\,m^{-2}}$ for the ocean regions and between $4\,\mathrm{kg\,m^{-2}}$ to $6.5\,\mathrm{kg\,m^{-2}}$ for the inland regions is found, which can be attributed to an underestimation of the MBL and the coarse representation of the topography by ERA-20C. This is in agreement to Poli et al. (2016) who found a similar bias between HOAPS and ERA-20C for spatial averages over the tropical ocean (between 20°S and 20°N). Such biases are not uncommon among

reanalyses (Schröder et al., 2017) and do not hinder the representation of variability.

Thus, in order to study the temporal variability of the IWV, we consider anomalies. It is found that even specific synoptic features, i.e. cut-off lows, are represented very well in ERA-20C. Furthermore, a time series analysis for four different focus regions revealed that temporal variability of IWV in ERA-20C is overall in accordance with the satellite observations. In particular, the phase of the seasonal cycle is captured well with correlations reaching almost unity. The inter-annual variability is represented well with correlations between ERA-20C and HOAPS of about 0.82 (ocean north) and 0.89 (ocean south). However, for a short period between 2005 and 2008, the inter-annual variability of ERA-20C IWV deviates about $1\,\mathrm{kg\,m^{-2}}$ from HOAPS and MODIS observations. This period of disagreement becomes a dominant feature for the comparison of ERA-20C and MODIS IWV due to the short overlapping time period (2003–2010). A possible reason for the disagreement could be a change in the assimilated observations.However, the seasonal cycle and the month to month variability are still represented well even for this time period.

The validation is carried out for the time past 1988 for which HOAPS data, which is the longest record of satellite based IWV observations, is available. As no area wide IWV observations are available prior to the satellite era, we have to rely on reanalyses. A further assessment of the observation feedback archive for ERA-20C revealed an increased uncertainty for the beginning of the 20th century. However, no systematic biases were identified which would hamper a useful analysis. We could show that ERA-20C is capable of representing IWV variability realistically by only assimilating surface pressure and wind observations along with temporally varying forcing data for the Atacama region.

In the second part of this study, we investigated the IWV variability which is represented in ERA-20C over the course of 111 years (1900–2010). Inter-annual variability is high with deviations of the 12 months running mean from the overall mean rising above 30 % for example for the inland regions. Furthermore, varying trends on inter-decadal time scales are observed. For instance, an increase of IWV is detected after 1970 until the strong El Niño in 1997/1998 and a decline thereafter, which is in particular pronounced in the northern regions. This increase is consistent with global warming which increases the water vapor holding capacity. However, between 1979 and 2006 local cooling trends have been found for coastal stations in the Atacama (Falvey and Garreaud, 2009). Another explanation of the increased IWV during the last quarter of the century could be the PDO which shows a concurrent shift into a positive phase during the 1970s. The onset of this positive trend from the 1970s onward and its potential linkage to the PDO as well as various earlier trends could not be revealed using

reanalysis which only cover recent decades, such as ERA-Interim (Dee et al., 2011b).

In general, warm phases of PDO, ENSO and local SST are associated with increased IWV, except for the northern inland region. A more detailed picture is revealed when a seasonal distinction is applied. While El Niño typically features a dry austral summer, La Niña allows dry and wet summers. This phenomenon is identified for all regions but most pronounced in the northern regions. While the mechanisms behind dry summers under El Niño conditions have been described mainly for the Altiplano in the northeastern part of the Atacama (e.g. Vuille, 1999; Garreaud et al., 2003), it remains unclear if the same mechanisms, i.e. enhanced westerlies/ suppressed easterlies, explain the observed dry conditions for the southern regions during the warm phase of ENSO. For the austral winter season, El Niño conditions have an opposite effect compared to the summer. Then, increased IWV is associated with the warm ENSO phase, while drier conditions are found for La Niña. We found overall positive correlations between ENSO and regional IWV on inter-annual time scales (Fig. 5.12). This implies that the increase of IWV during winter is generally stronger than the overall lower IWV during the summer associated with El Niño. This seasonal opposite effects seem more balanced for the northern regions manifested in lower correlations. An even negative correlation for the northern inland region indicates the following: The drying effect due to enhanced westerlies during El Niño summers overcompensates the increased IWV signal during El Niño winters. This results in overall drier years associated with the warm ENSO phase. However, there might be potential tipping points for very strong El Niño events which might alter the previously described relations. An example could be the 1997/98 El Niño for which the northern ocean region yields the highest peak within the entire ERA-20C record (Fig. 5.9).

La Niña conditions can feature a wider range of IWV during the summer compared to El Niño. While dry summers are typical during El Niño conditions, they can also occur during La Niña conditions. However, wetter summers are limited to La Niña. This indicates, that while the ENSO state favors certain conditions, the synoptic variability is still an important factor to influence the IWV. This is also evident from the higher correlations between the indices and IWV for the long term moving averages (10-year) when synoptic variability is smoothed out compared to shorter time scales.

Generally, the local SST explains the most variability compared to ENSO and PDO. This might be due to the more complex relationship between ENSO (or PDO) and IWV. The impacts of ENSO depend on region and season. Additionally to an immediate impact on IWV via increased SST according to the ENSO phase, teleconnection patterns change the general circulation patterns over the study region complicating the relation to IWV.

Furthermore, the PDO as well as the Niño 3.4 index and the local SST can only partially explain the IWV variability within the four focus regions. Highest correlations are found between the local SST and the IWV of the ocean regions. Therefore, other features have to be identified which further influence the IWV. This could be different weather types which might in turn be controlled by ENSO to a certain degree. Future studies could further exploit ERA-20C. For instance, it would be interesting to investigate to which degree the circulation in different levels of the troposphere is influencing the regional IWV. Marín and Barrett (2017) found that during fall, winter and spring season, high IWV episodes in the Altiplano are concurrent with intensified troughing to the east of northern Chile. The resulting northwesterly flow advects anomalous humid air to the Altiplano. Their study is based on reanalysis data for more recent decades (1979–2010) which are dominated by PDO warm phase. The long record of ERA-20C would allow to test whether the PDO phase influences the weather patterns which yield lower or higher IWV. Moreover, the relationship between ENSO and IWV could be investigated in dependence on the PDO phase. Andreoli and Kayano (2005) already investigated this relationship with regard to precipitation for the period between 1948 to 1999 for the South-American continent. Their findings reveal more conspicuous El Niño signals during the warm PDO phase with more pronounced seasonal differences.

Clouds and precipitation are much more difficult to simulate compared to water vapor as they result from the interplay of many small scale processes. Only a few station records with partly missing data provide precipitation estimates within the Atacama. Available satellite estimates of precipitation, e.g. the Tropical Rainfall Measuring Mission (TRMM), suffer from uncertainties and biases which are greater than the actual signal for the extreme environment of the Atacama (Schween et al., 2020). Therefore, it would be desirable to potentially use IWV as a proxy for precipitation in order to estimate the amount of liquid water which is provided to the land surface and the biosphere. A regional study in the Spanish Mediterranean area demonstrated that IWV is closely related to extreme precipitation (Priego et al., 2017). To investigate whether IWV can serve as a proxy for precipitation, investigations with cloud resolving models are planned in the future.

5.6 REMARKS

SUPPLEMENT Supplement material is available in the Appendix A.2 or at: https://doi.org/10.1016/j.gloplacha.2020.103192

DATA AVAILABILITY ERA-20C data were downloaded from the ECMWF data server via Web-API. HOAPS data were ordered from the CM-SAF Web User Interface (https://wui.cmsaf.eu/, last ac-

cess: 16 October 2018). MODIS data were downloaded via MODIS Level 1 Atmosphere Archive and Distribution System (LAADS Web) from https://ladsweb.modaps.eosdis.nasa.gov/archive/allData/61/MOD05_L2/.

ACKNOWLEDGEMENTS We gratefully acknowledge financial support by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – project number 268236062 – SFB 1211.

6

FOG

TOWARDS A CLIMATOLOGY OF FOG FREQUENCY IN THE ATACAMA DESERT VIA MULTI-SPECTRAL SATELLITE DATA AND MACHINE LEARNING TECHNIQUES

Böhm, C., J. H. Schween, M. Reyers, B. Maier, U. Löhnert, and S. Crewell (2020). "Towards a climatology of fog frequency in the Atacama Desert via multi-spectral satellite data and machine learning techniques." Submitted to: *Journal of Applied Meteorology and Climatology*, date of submission: 10 Sep 2020.

ABSTRACT. In many hyper-arid ecosystems, such as the Atacama Desert, fog is the most important fresh water source. To study biological and geological processes in such water-limited regions, knowledge about the spatio-temporal distribution and variability of fog presence is necessary. In this study, in-situ measurements provided by a network of climate stations equipped, inter alia, with leaf wetness sensors are utilized to create a reference fog data set which enables the validation of satellite-based fog retrieval methods. Further, a new satellite-based fog detection approach is introduced which uses brightness temperatures measured by the Moderate Resolution Imaging Spectroradiometer (MODIS) as input for a neural network. Such a machine learning technique can exploit all spectral information of the satellite data and represent potential non-linear relationships. Compared to a second fog detection approach based on MODIS cloud top height retrievals, the neural network reaches a higher detection skill (Heidke skill score of 0.56 compared to 0.49). A suitable representation of temporal variability on subseasonal time scales is provided with correlations mostly greater than 0.7 between fog occurrence time series derived from the neural network and the reference data for individual climate stations, respectively. Furthermore, a suitable spatial representativity of the neural network approach to expand the application to the whole region is indicated. Three-year averages of fog frequencies reveal similar spatial patterns for the austral winter season for both approaches. However, differences are found for the summer and potential reasons are discussed.

6.1 INTRODUCTION

Fog is the major freshwater source for several plant species within the coastal parts of the Atacama Desert (Muñoz-Schick et al., 2001). The collection of fog water is also of major social and economic importance in this region (Schemenauer and Cereceda, 1994b; Osses et al., 2000; Larraín et al., 2002).

Fog is frequently created where the stratocumulus intercepts the coastal cliff and coastal cordillera. Its frequency decreases inland, which is consistent with a decrease in colonization rate of hypolithic cyanobacteria inland (Cáceres et al., 2007). Farther inland, favorable conditions for the formation of radiation fog, which is very difficult to detect, have been documented (Cereceda et al., 2002; Westbeld et al., 2009). With annual precipitation rates below 1 mm, fog water supply is the main driver for biological and geological processes for various regions within the Atacama Desert. However, a regionally resolved fog climatology is not available yet.

Highly adapted plant species, such as Tillandsia, rely almost solely on fog as a source of water and nutrition (Rundel et al., 1997; Koch et al., 2019). They typically grow in those regions of the coastal cordillera

where fog is frequently present (Pinto et al., 2006). The isotopic composition of nitrogen from Tillandsia plants is related to fog moisture (Latorre et al., 2011; Jaeschke et al., 2019). It is highly correlated to the height range of the maritime stratocumulus and the cloud cover frequency indicating its potential to serve as a proxy to reconstruct historic climate utilizing conserved Tillandsia remnants (Jaeschke et al., 2019). To calibrate such a proxy with higher accuracy, quantification of fog frequency for the current climate is required. Recently, a dieback of the Tillandsia populations has been observed which is believed to be related to decreasing fog occurrence at the respective locations (Rundel et al., 1997; Pinto et al., 2006; Schulz et al., 2012).

In addition to plants, soil is strongly affected by fog. The amount of organic traces (Mörchen et al., 2019) and the activity of microbial communities (Jones et al., 2018) within desert soil depend on water sources such as fog.

For individual locations, the near surface fog water supply has been quantified using different types of fog collectors (e.g. Cereceda et al., 2002; Cereceda et al., 2008b; Lobos Roco et al., 2018; Osses et al., 2005; Río et al., 2018). Based on a 17-year long time series of fog water collected at a research site at the coastal cliff (Alto Patache), the seasonal cycle has been investigated revealing that fog water yields are highest for the austral winter season (July, August, September) and lowest between December and April (Río et al., 2018). These observations, which are almost the only available in-situ source for fog, are consistent with satellite-based observations of the maritime stratocumulus deck which is most persistent with highest cloud cover during winter (Farías et al., 2005; Cereceda et al., 2008b; Lehnert et al., 2018b). Furthermore, the stratocumulus top and base height undergo a seasonal cycle with higher heights during summer and lower during winter (Muñoz et al., 2016; Böhm et al., 2019) with average nocturnal cloud bases below 800 m during winter and between 900 and 1100 m during summer for Arica and Iquique and even lower heights at Antofagasta (Muñoz et al., 2016). Due to the lower heights, the stratocumulus is more likely to intersect with the coastal cliff and mountain range, which prevents further inland advection during winter. Exceptions are regions at which the coastal cliff is intercepted by canyons or generally lower, allowing inland penetration. Such fog corridors have been identified and related to fog occurrence in the central depression (Farías et al., 2001; Farías et al., 2005).

The coastal fog related to the maritime stratocumulus shows a maximum occurrence during night time and a minimum during day time (Farías et al., 2005; Farías, 2007; Cereceda et al., 2008b; Muñoz et al., 2011). The nocturnal maximum is the result of the stratocumulus deck being advected towards the coast at this time while the circulation reverses during the day (Rutllant et al., 2003). Furthermore, entrainment of warm and dry free tropospheric air and warming forced by ab-

sorption of solar radiation lead to dissipation of coastal stratocumulus during day time (Garreaud and Muñoz, 2004). Further inland, e.g. in the Pampa de Tamarugal within the central depression, conditions are favorable for radiation fog (Cereceda et al., 2002; Farías, 2007), which is typically more likely to form at night time than at day time.

To date, multi-year averages based on satellite observations only exist for individual months and only regarding cloud cover as a proxy for fog (e.g. Cereceda et al., 2008b; Lehnert et al., 2018b). The goal of this study is to develop a satellite-based fog retrieval method for the region of the Atacama Desert. It needs to be suitable to derive a long-term climatology and to study seasonal and interannual variability of fog occurrence. As fog occurrence typically peaks at night, we focus on the detection of nocturnal fog. This way, all thermal infrared channels can be considered without additional complexity from the solar component which affects the middle infrared channels during day time.

Satellite-based detection techniques of clouds in general and low clouds or fog in particular rely on simultaneous measurements of radiances at various frequency bands. Varying relationships between emissivity and wavelength for different emitters, such as different soil types, water droplets, or ice particles, result in different radiative signatures at the top of the atmosphere. A distinction between different cloud scenes is thus possible by the evaluation of different channels, i.e. radiances or brightness temperatures at different wavelengths, or combinations of multiple channels. For instance, the absorption and emission increase stronger for ice compared to liquid water when the wavelength increases from 10 µm to 12 µm resulting in a distinct radiative signatures (Strabala et al., 1994). Typically, radiative transfer calculation are utilized to estimate channel dependent thresholds to make the desired distinction between different cloud scenes (e.g. Bendix et al., 2006; Cermak, 2012; Gaurav and Jindal, 2018). To carry out such calculations, assumptions are required regarding soil type, surface temperature, cloud droplet size distribution, cloud height and thickness, water vapor and temperature profile, etc. While it is comparatively easy to distinguish between low and high clouds using infrared wavelengths due to higher differences of the respective cloud top temperatures, the distinction between low clouds, fog and land surface is more difficult because the temperature at which thermal radiation is emitted is similar for these features (Güls and Bendix, 1996). A distinction between low clouds and land surfaces is, however, possible because the emissivity of liquid water clouds is significantly lower in the short wave infrared (e.g. 3.8 µm) compared to longwave infrared (e.g. 11 µm) (Hunt, 1973; Ellrod, 1995; Gaurav and Jindal, 2018). This results in greater brightness temperature differences observed for low level clouds compared to land surfaces (Ellrod, 1995). However, this is less pronounced for larger cloud droplets. Further

uncertainties arise from the water vapor dependence of the brightness temperature difference for clear sky cases (Ellrod, 1995). Andersen and Cermak (2018) mitigate the low cloud distinction problem by analyzing the spatial structure of the satellite image. Fog typically leaves a more homogeneous picture compared to the land surface. After an initial fog and low cloud identification based on thresholds for four different infrared channels, they carry out a structural similarity test to distinguish between surface and low stratus. However, their approach does not distinguish between fog and low cloud.

Present-day spaceborne sensors such as the Moderate Resolution Imaging Spectroradiometer (MODIS) on board the polar orbiting satellites Terra and Aqua, the Spinning Enhanced Visible and Infrared Imager (SEVIRI) on board the Meteosat Second Generation (MSG) platform or the Advanced Baseline Imager (ABI) on the Geostationary Operational Environmental Satellite (GOES) provide high spectral resolution. Instead of a threshold-based detection based on selected channels and channel combination, the aim of this study is to explore the entire spectral information available in the thermal emissive range. As mentioned above, multiple factors are influencing the radiative signature which is observed by spaceborne sensors. To capture these different, yet presumably distinct, radiative signatures for various fog, cloud or clear sky scenes, a machine learning technique is applied to recognize the relevant patterns. This way, all potential indicators within the spectral signature can be exploited to detect fog.

Machine learning techniques are becoming increasingly popular and have also been used to detect fog in previous studies. For instance, Egli et al. (2018) apply a random forest technique to process MSG data to predict cloud base height. Via a dynamical threshold for the derived cloud base height, a binary classification (fog or no fog) is achieved. To train the random forest, they apply cloud base height observations from Meteorological Aviation Routine Weather Reports (METAR) and synoptic weather observations (SYNOP) as ground-based reference. For the evaluation of their final binary classification (fog or no fog), they consider METAR and SYNOP visibility reports. These ground-based observation are widely available within their study region (Europe).

As there are only very few climate stations available within the Atacama Desert, an effort has been made by the Collaborative Research Center CRC1211 "Earth – Evolution at the dry limit" (sfb1211.uni-koeln.de; Dunai et al., 2020) to fill this observational gap by installing a network of climate stations which started in 2017 (Hoffmeister, 2017b; Schween et al., 2020). Now, sufficient in-situ reference measurements enable the development and validation of new satellite-based fog detection methods. Here, this unique opportunity is applied to train and examine a neural network approach to detect fog. To identify the benefit of the neural network retrieval method, we create an alternative

fog detection based on simple thresholds applied to a satellite-based cloud top height product.

The paper is structured as follows. In Section 6.2, the utilized satellite data products, and the climate station measurements, are described. In Section 6.3, the ground-based reference data set and the fog retrieval methods are introduced. In a twofold evaluation (Section 6.4), event based statistics according to a contingency table analysis are presented first followed by an investigation of the spatio-temporal representativity of the detection methods. The proposed fog retrieval methods are then utilized to derive a region wide distribution of fog occurrence frequency averaged for a 3-year period. Finally, the study is summarized and concluded (Section 6.5).

6.2 DATA

For the satellite-based fog retrievals methods introduced in this study, we utilize MODIS data products. Compared to the ABI on GOES, MODIS provides a few more channels and a longer data record enabling future climate studies. Operational data are provided for the time since February 2000 (Terra) and July 2002 (Aqua) for MODIS and December 2017 for the new GOES-16. After an overview on the MODIS instrument and the utilized data products, the climate station measurements building the reference data set are described in this Section.

6.2.1 *Moderate Resolution Imaging Spectroradiometer (MODIS)*

MODIS is an imaging sensor capturing data in 36 spectral bands at wavelengths ranging from visible (0.4 µm) to infrared (14.4 µm). The spatial resolution at nadir is generally 1 km, while a few channels in the visible range additionally provide data at 500 m and 250 m. The instrument is installed on both the Terra and the Aqua platform. Both satellites are in sun-synchronous orbits at a height of about 705 km and a swath width of approximately 2330 km. As mentioned in the introduction, we consider only the nocturnal overpasses to capture the diurnal maximum of fog occurrence reported in previous studies (e.g. Farías et al., 2005; Farías, 2007; Cereceda et al., 2008b; Muñoz et al., 2011) and potential radiation fog in the central depression (Farías, 2007; Cereceda et al., 2002). Due to orbit characteristics, the local time of the nocturnal overpasses over the Atacama Desert varies between 22:30 and 00:10 CLT (Terra) and 1:10 and 2:45 CLT (Aqua).

Various data products derived from MODIS targeted at different applications are available. In this study, we utilize the spectral radiances provided by the level 1B 1 km Calibrated Radiances Product (MOD021KM, MYD021KM; MODIS Characterization Support Team (MCST), 2017a; MODIS Characterization Support Team (MCST), 2017b)

which we convert to brightness temperatures (Sec. 6.2.1.1). Additionally, we utilize cloud top height provided by the level 2 Cloud Product (MOD06, MYD06; Platnick et al., 2017a; Platnick et al., 2017b) (Sec. 6.2.1.2) to derive an alternative fog retrieval. For geolocation of the acquired fields, longitude, latitude and elevation are taken from the Geolocation Fields Product (MOD03, MYD03; MODIS Characterization Support Team (MCST), 2017c; MODIS Characterization Support Team (MCST), 2017d). For all products, the collection 6.1 data are acquired for this study.

6.2.1.1 *MODIS brightness temperatures*

As only night time satellite overpasses are considered, only the thermal emissive bands (MODIS bands 20–25 and 27–36) are processed further. Central wavelengths and atmospheric features which are targeted by each band are listed in Table 6.1. Furthermore, band 36 (14.2 µm) is omitted because in about 13 % of all cases with collocated and coincidental station measurements, no valid spectral radiance retrieval is provided for this particular band. To predict fog using the neural network (Sec. 6.3.3), brightness temperatures T_B are used as explanatory variable. The spectral radiances are converted to brightness temperatures according to Planck's law.

6.2.1.2 *MODIS cloud product*

The MODIS Cloud Product provides the cloud top height which is utilized in this study. Here, a short summary of the cloud top height retrieval method is given mentioning ancillary data and potential sources of errors. The MODIS cloud top height is derived for pixels which are cloudy according to the MODIS cloud mask via the CO_2 slicing technique (Chahine, 1974; Smith and Platt, 1978) using four spectral bands near the CO_2 absorption region at 15 µm (Menzel et al., 2008; Baum et al., 2012). Differences between observed radiances and clear sky radiances are calculated for each utilized MODIS channel. The cloud top pressure can then be inferred by taking the ratios of the differences calculated for different channels. The determined cloud pressure is converted to height using atmospheric profiles derived from the National Centers for Environmental Prediction Global Data Assimilation System (GDAS; Derber et al., 1991). Clear sky radiances are determined via radiative transfer calculations using GDAS temperature, moisture and ozone profiles.

If the difference between observed and clear sky radiance is within the instrument noise level, which is typically the case for clouds below 3 km, the CO_2 slicing technique is not applied to infer cloud pressure. In such cases, the brightness temperature of the cloud is determined using the infrared window band at 11 µm. The cloud pressure and height are then inferred via brightness temperature profiles calculated

Table 6.1: MODIS thermal emissive bands overview. For each band, the center wavelength in µm and the primary use for feature detection is listed. Adapted from Xiong et al. (2008). Channel 21 and 22 differ in detection range: Band 22 saturates at 331 K, band 21 (the "fire channel") saturates at 500 K (Xiong et al., 2008).

band	center wavelength	primary use
20	3.75	
21	3.96	surface/cloud
22	3.96	temperature
23	4.05	
24	4.47	atmospheric
25	4.52	temperature
27	6.72	
28	7.33	water vapor
29	8.55	
30	9.73	ozone
31	11.03	surface/cloud
32	12.02	temperature
33	13.34	
34	13.64	cloud top
35	13.94	height
36	14.24	

from GDAS temperature, water vapor, and ozone profiles. In the presence of low level temperature inversions, the height of the matching temperature above the inversion is chosen introducing a positive bias into the MODIS cloud top height. Starting with MODIS collection 6, this problem is mitigated for retrievals over ocean (Baum et al., 2012). However, over land temperature inversions remain problematic.

6.2.2 *Climate stations*

The climate stations, which are part of the station network mentioned in the introduction, are deployed throughout the Atacama Desert in a southern (around 25°S, stations 31, 32, 33, 34), central (around 21.4°S, stations 11, 12, 13, 14, 15) and northern (20°S, stations 20, 21, 22, 23,24 , 25) latitudinal transect, making up a total of 15 stations (Fig. 6.1). The station network is assumed to be representative for the study region due to its spread across the Atacama Desert covering latitudes between 20°S and 25°S and topographic heights between 770 m and

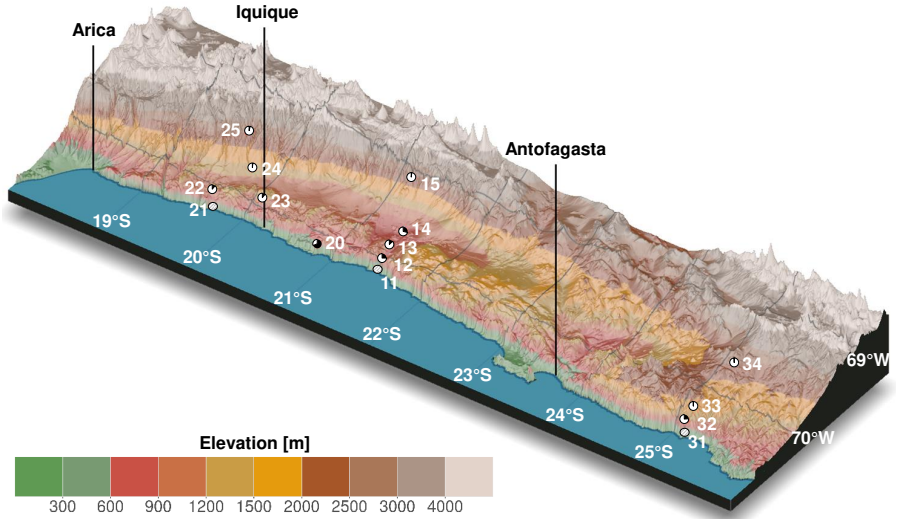

Figure 6.1: Topographic map of the study region. Color shading indicates the elevation above sea level according to Shuttle Rader Topography Mission (SRTM; Farr et al., 2007). Black vertical lines mark major cities (Arica, Iquique, Antofagasta) located at the coast. The climate stations are indicated by black and white pie charts along with their respective station ID. The pie charts indicate annual fog occurrence frequency (fog portion in black) determined from the stations as listed in Table 6.2. For coastal stations 11, 21, 31 these data are not available.

2630 m ranging between the coastal cliff and the slopes of the Andes. The installation was carried out between April 2017 and March 2018 and continuous measurements are provided since. The time intervals which are considered for this study along with the coordinates and elevation of each station are listed in Table 6.2.

Among other sensors, each station is equipped with a leaf wetness sensor which mimics the characteristics of a leaf and provides a voltage output (Campbell Scientific, 2018). Values above $U = 284\,\mathrm{mV}$ indicate that the sensor is wet according to the manufacturer. This threshold was validated for each station at the time of their installation (Schween et al., 2020). In the presence of fog, impacting water droplets wet the sensor's surface reducing the electrical resistance. At the end of a fog event, evaporation and drainage remove the water from the surface. Both processes at the beginning and end of a fog episode may take some time before the sensor is wet or dry, respectively. Thus, simply using the signal from the leaf wetness sensor would lead to temporal shifts of the fog episodes. Therefore, we include other measured variables to estimate fog conditions: relative humidity rH, air temperature at 2 m height $\vartheta_{2\mathrm{m}}$, surface temperature ϑ_{srf} measured by an IR thermometer, upwelling and downwelling longwave radiation P_{up} and P_{down}. The longwave radiation sensors are only installed at the so called master stations of each transect (stations 13, 23, 33, see Table 6.2) which are deployed about 20 km from the coast at heights

Table 6.2: Climate station metadata. Listed are station ID, longitude (lon), latitude (lat), altitude (alt), fog occurrence frequency (fof), start and end times of the considered measurements t_{start} and t_{end}. The fog occurrence frequency is given for the year between April 2018 and March 2019 for all stations except station 12 (December 2017 – November 2018) including only measurements which coincided with a nocturnal overpass by MODIS. Locations of the climate stations are indicated in Fig. 6.1.

station	lon [°]	lat [°]	alt [m]	fof	t_{start}	t_{end}
12	-69.96	-21.42	771	0.25	2017-04-03	2018-11-30
13	-69.84	-21.40	1152	0.13	2017-04-03	2019-12-31
14	-69.54	-21.36	795	0.29	2017-09-25	2019-12-31
15	-69.07	-21.11	2408	0.02	2017-09-26	2019-12-31
20	-70.16	-20.83	776	0.69	2018-03-07	2019-12-10
22	-70.10	-19.61	1179	0.11	2018-03-10	2019-10-09
23	-69.94	-20.07	1280	0.11	2018-03-10	2019-12-31
24	-69.65	-19.76	1392	0.02	2018-03-11	2019-10-09
25	-69.39	-19.53	2628	0.04	2018-03-10	2019-12-31
32	-70.40	-25.10	1026	0.24	2018-03-24	2019-12-31
33	-70.28	-25.09	1700	0.00	2018-03-24	2019-12-31
34	-69.65	-25.09	2535	0.00	2018-03-14	2019-07-23

between 1150 m and 1700 m above sea level. Measurements are taken every 10 s and averages are stored every 10 min.

Stations in close proximity to the Pacific had to be omitted from the analysis as they show an increase of fog frequency within two weeks after deployment and every cleaning which is inconsistent with measurements of relative humidity. This is probably due to salt deposition as they are exposed to sea spray (Schween et al., 2020). Furthermore, the leaf wetness sensor of station 12 faces technical issues since December 2018 so that measurements from this station are only considered until this break point.

6.3 FOG DETECTION METHODS

The ground-based measurements are utilized to create an unprecedented fog reference data set for the Atacama Desert based on the leaf wetness sensor and some additional constraints (Section 6.3.1). This reference data set is applied to develop a novel fog retrieval method via a deep learning technique based on brightness temperatures 6.3.3). To identify the benefit of the new retrieval, we introduce a simple

fog detection method based on thresholds for the MODIS cloud top height to compare with (Section 6.3.4). In the following subsections, the derivation of the ground-based reference fog data, and the two satellite-based approaches are presented.

6.3.1 *Ground-based reference*

The goal is to derive a binary classification, fog or dry, from the station measurements which can be applied to develop satellite-based retrieval methods. Using the leaf wetness sensor alone would be problematic at the beginning and end of fog episodes, as the sensor is expected to require some time to adjust to the change of the ambient conditions. Therefore, collocated and coincident measurements of U, rh, the temperature difference $\Delta\vartheta = \vartheta_{2\mathrm{m}} - \vartheta_{\mathrm{srf}}$, the longwave radiation budget $\Delta P = P_{\mathrm{up}} - P_{\mathrm{down}}$ (positive means energy loss at the surface) are taken into account as well. For each climate station s and each measurement time t_n, these variables are bundled in a state vector $\vec{v}_s = \vec{v}_s(t_n)$. To consider the response time of the leaf wetness sensor, two time differences are added to the state vector: (i) if the sensor is wet, the time from t_n until it is dry denoted as $\Delta t_{\mathrm{wet2dry}}$ and (ii) if the sensor is dry, the time from t_n until it is wet denoted as $\Delta t_{\mathrm{dry2wet}}$.

A priori, the response time of the leaf wetness sensor is not known. Furthermore, it varies most likely depending on the meteorological conditions. To determine a robust classification of the fog state, we investigate which configurations of $\vec{v}_s$ typically occur during fog or dry episodes, respectively. Typically, a high voltage of the leaf wetness sensor indicating foggy conditions coincides with a high relative humidity. At the same time, the air temperature is expected to be close to or below the surface temperature since fog cools from the top and hinders outgoing longwave radiation. Therefore, the surface does not cool as fast as during dry conditions. Fog conditions also lead to a reduced energy loss regarding the longwave radiation indicated by a lower ΔP. However, under certain conditions, these expectations may not be fulfilled. In particular at the beginning or end of a fog episode, U might not be consistent with the other quantities in terms of fog indication, because the leaf wetness sensor is expected to react slower than the other sensors to changing conditions. In order to identify such events for which the indication of the leaf wetness sensor needs to be rectified, a clustering method is needed which allows to quantify the configuration of U, rh, $\Delta\vartheta$, and ΔP typically associated with fog or dry conditions, respectively. By further considering the leaf wetness sensor response times $\Delta t_{\mathrm{wet2dry}}$ and $\Delta t_{\mathrm{dry2wet}}$, events with an inconsistent indication from the leaf wetness sensor can be extracted.

Such a clustering can be achieved in a straightforward way using a self organizing map (SOM). In order to investigate which configurations of the state vector typically cluster together, a SOM is cre-

ated according to the principle introduced by Kohonen (2001). In the following, a brief introduction to the concept is given. SOMs are a machine learning tool which reduces dimensionality of input data while mapping similar input vectors close to each other on a space of a lower dimension. In our case, the 6-dimensional state vectors ($\vec{v}_s\,(t_n) = \big(U(t_n), rh(t_n), \Delta\vartheta(t_n), \Delta P(t_n), \Delta t_{\text{wet2dry}}(t_n), \Delta t_{\text{dry2wet}}(t_n)\big)$) are mapped onto a 2-dimensional plane ($\vec{v}_s{}'\,(t_n) = (x(t_n), y(t_n))$). For a predefined 2-dimensional grid, each grid cell is assigned a random initial "codebook" vector at the beginning of the training interval. The first state vector is taken from the observations and assigned to the grid cell for which the euclidean distance between the state vector and the codebook vector is minimized. Then the codebook vector of that particular grid cell is adjusted according to the learning rate to become closer to the assigned state vector. Furthermore, to ensure that similar state vectors eventually will be assigned to neighboring grid cells, the codebook vectors of neighboring grid cells within a defined neighborhood radius are updated as well. Once each state vector from the observations has been presented to the map, one iteration is complete. The order of the state vectors is then resampled randomly and the process is started over. With each iteration the learning rate and the neighborhood radius are reduced so that eventually the map converges. Here, we applied the **kohonen** package for R (Wehrens and Buydens, 2007; Wehrens and Kruisselbrink, 2018) to create SOMs from the climate station data.

We filtered the data from the climate stations for times between 22:00 and 04:00 CLT since this study focuses on fog detection at night time. Prior to the training, the input data were scaled so that the input values for each quantity are centered (reduced by the respective means) and divided by the standard deviation of the centered values. After some testing, a 32×32 grid was chosen. The outcome of the map appeared to be not very sensitive to the grid size. However, if the size becomes very small, the variability of state vectors assigned to the same grid cell increases. On the other hand, if the grid size becomes too large, more and more grid cells are left blank without any assignment of a state vector. Further, we apply a hexagonal grid topology. The neighborhood radius decreases linearly after each iteration from an initial size that includes 2/3 of all unit distances to zero. Once the radius decreases to less than the distance between two units, only the codebook vector of the assigned grid cell is updated. The learning rate linearly decreases from 0.05 to 0.01 from the first to the last iteration. After 60 iterations the mean euclidean distance between input vectors and codebook vectors does not decrease anymore, indicating that the map converged. SOMs are trained for each station individually.

After each state vector is assigned to the grid cell for which the euclidean distance to the respective trained codebook vector is minimal, the grid cell mean can be calculated for each quantity. Then,

each quantity can be visualized separately on the two dimensional grid. By comparing corresponding grid cells, it is easy to see visually which values typically occur concurrently. An example is shown for master station 13 (Fig. 6.2). SOMs for the other stations are available in the supplement material (Appendix A.3). High values of U are usually accompanied by higher rh and lower values for $\Delta\vartheta$ and ΔP. However, there are some cases for which the leaf wetness sensor is indicating fog, even though $\Delta P >> 0$ reveals a strong energy loss at the surface indicating clear conditions (e.g. hexagons with a white "-" in Fig. 6.2). For this region within the SOM grid, $\Delta t_{\text{wet2dry}}$ is low, indicating that such configurations occur at the end of a fog episode. We assume that fog ended and the sensor is in the drying phase but still wet. From the values in the map, we can conclude that this drying period lasts between a few minutes up to 2.5 h. On the other hand, for another region on the SOM grid (e.g. bottom of upper right quarter in Fig. 6.2e), similar values for $\Delta t_{\text{wet2dry}}$ are shown while all other variables indicate that fog conditions are plausible. This indicates that the response time of the leaf wetness sensor varies greatly depending on concurrent meteorological conditions as has been expected.

In a similar way, grid cells of the SOM can be identified which represent the following situation: While the leaf is dry, a high rh, a low $\Delta\vartheta$, and a low ΔP indicate fog like conditions. For some of these cases, the time until the leaf wetness sensor switches to wet is less than 2.5 h (hexagons with white "+" in Fig. 6.2). Therefore, we assume that fog is already present but the sensor is not wet yet. Again, a varying response time is indicated by similar values of $\Delta t_{\text{dry2wet}}$ for other regions of the map for which other variables are consistent with dry conditions. While the leaf wetness sensor can serve as an initial classification, relative humidity, temperature difference, longwave radiation budget along with $\Delta t_{\text{wet2dry}}$ or $\Delta t_{\text{dry2wet}}$ should be considered to refine the initial classification.

By visual inspections of all the resulting SOMs, a fog definition is derived which consists of an initial classification according to U followed by additional tests taking the other variables into account. A flow chart summarizes the process (Fig. 6.3). If $U > 284\,\text{mV}$ the initial classification is set to fog otherwise set to dry. If fog was determined and any of the conditions $rh < 80\,\%$, $\Delta\vartheta > 1\,\text{K}$ or $\Delta P > 50\,\text{Wm}^{-2}$ is fulfilled and the leaf wetness sensor dries within the following 2.5 hours ($\Delta t_{\text{wet2dry}} < 185\,\text{min}$), the initial classification is revoked and set to dry. An initial dry classification is revoked in case all conditions $rh \geq 84\,\%$, $\Delta\vartheta \leq 0\,\text{K}$, $\Delta P \leq 40\,\text{Wm}^{-2}$ and $\Delta t_{\text{dry2wet}} < 155\,\text{min}$ are fulfilled simultaneously. This fog definition is applied equally to measurements from all considered climate stations. For a total of 278 events, the initial fog classification has been revoked which corresponds to 12.7 % of all initial fog events. This number is partly compensated by 187 events which have been switched to fog after initial dry classification. This

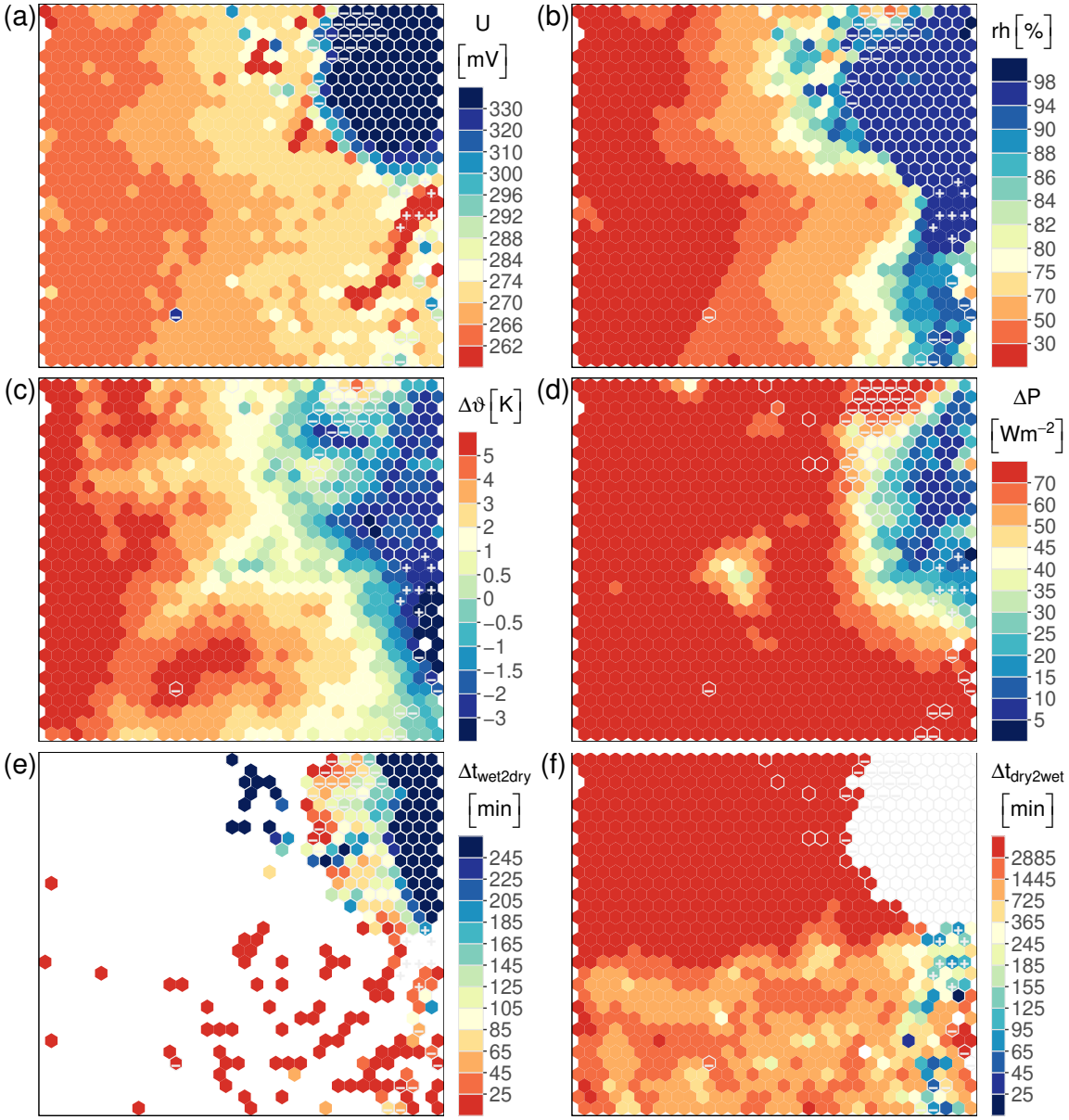

Figure 6.2: Self organizing map for climate station 13. Shown is the average for each grid cell for (a) leaf wetness sensor voltage, (b) relative humidity, (c) temperature difference between air and surface ($\Delta\vartheta = \vartheta_{2m} - \vartheta_{srf}$), (d) longwave radiation budget ($\Delta P = P_{up} - P_{down}$), (e) time until a wet leaf wetness sensor switches to dry, (f) time until a dry leaf wetness sensor switches to wet. Initially, grid cells are set to fog if the leaf wetness sensor is wet on average (white frames). Otherwise grid cells are set to dry. The initial classification is changed from fog to dry (white "-") or from dry to fog (white "+") according to additional tests (see text and Fig. 6.3). SOMs for other climate stations are available in the supplement (Appendix A.3).

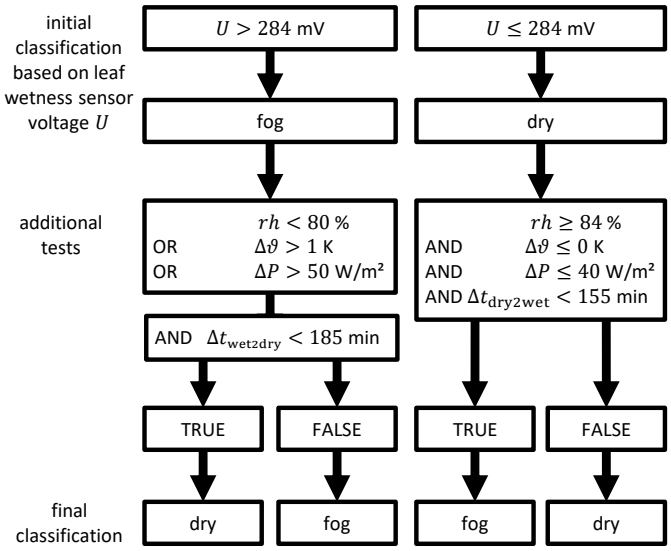

Figure 6.3: Flow chart of binary fog state classification (fog or dry) based on climate station measurements. After an initial classification according to the leaf wetness sensor voltage U, additional tests are applied for relative humidity rH, temperature difference $\Delta\vartheta$, radiation budget ΔP, and the time until U crosses the threshold of 284 mV $\Delta t_{\mathrm{wet2dry}}$ and $\Delta t_{\mathrm{dry2wet}}$. Note, that for the OR conjunction the whole box is "TRUE" if at least one of the listed conditions is fulfilled whereas the AND conjunction requires all individual conditions to fulfilled for the whole box to be "TRUE". Depending on the additional tests, the initial classification can be revoked or confirmed.

indicates that on average the time the leaf wetness sensor takes to dry after a fog episode is longer than the time it takes to turn wet at the beginning of a fog episode.

For regular stations which do not provide the longwave radiation budget, the condition for ΔP is omitted. Furthermore, for climate station 20, no valid humidity data are provided between September 2018 and February 2019. For station 14, the SOM analysis reveals a puzzling picture. When the leaf wetness sensor indicates fog, $\Delta\vartheta$ is mostly high, unlike what is seen for all other stations. Here, the air temperature is predominantly about 2 to 4 K higher than the surface temperature which is atypical for fog events and rarely observed at the other stations. Since the surface temperature is determined with an infrared thermometer with an assumed surface emissivity of 0.94 for all stations, the different temperature signature could be due to a local surface emissivity anomaly as the location is in close proximity to the Salar de Llamara, a salt flat in the central depression. Such an anomaly would also affect the radiative signature for the thermal emissive MODIS bands. Therefore, station 14 is left out for further analysis.

Following the presented fog definition, a ground-based reference data set is ready to be applied to derive and validate satellite-based fog retrieval methods. MODIS data and station data are assigned to each other by taking the nearest MODIS pixel and the station measurements which are closest in time. About 13800 valid MODIS and station measurement pairs are available within the considered time period including 11 stations.

6.3.2 *Classification assessment measures*

This study faces a binary classification problem (fog or dry conditions). Comparing the satellite-based fog detection introduced in the following sections with the reference data set can be done using a confusion matrix (2×2 contingency table for binary classification) (e.g. Cermak, 2012; Egli et al., 2017; Andersen and Cermak, 2018) which yields the number of true positives (correct fog prediction), true negatives (correct dry prediction), false positives (incorrect fog prediction or false alarm), and false negatives (missed fog event). These numbers are used to derive further evaluation measures such as the probability of detection (POD), also known as true positive rate (TPR), which gives the fraction of all ground truth fog events which are correctly detected, the accuracy (ACC), also know as portion or percent correct, which gives the fraction of all observations with correct classification, the false alarm rate (FAR), which gives the fraction of all fog predictions which are false alarms, the critical success index (CSI), which gives the portion of fog hits out of all false classifications and fog hits combined and the bias score (BS) which gives the bias of the classification with an overestimation of fog for $\mathrm{BS} > 1$ and an underestimation for $\mathrm{BS} < 1$. Furthermore, the Heidke skill score (HSS; Heidke, 1926; Hyvärinen, 2014) is applied as a measure for prediction skill. The HSS gives the fractional improvement compared to a random classification. A perfect forecast would result in $\mathrm{HSS} = 1$, a random forecast would result in $\mathrm{HSS} = 0$.

Definition of these measures are provided in the Appendix. Further insights into classification assessment methods is given by Fawcett (2006) or Tharwat (2018), for example.

6.3.3 *Neural network*

As indicated in the introduction, conventional satellite-based approaches to detect fog and low clouds are typically based on radiative transfer simulations which approximate the underlying physical laws and are typically carried out for selected wavelengths. To exploit the available spectral information and represent the interactions of various factors and processes involved in the radiative transfer, we employ a neural network to detect fog. Neural networks and other machine

learning techniques have been used in remote sensing and earth observation for various applications. Gardner and Dorling (1998) and Lary et al. (2016) provide overviews.

In general, neural networks map output variables to input variables by propagating the input (signal) through a net of nodes. Next to the input layer with the input nodes, several hidden layers with various numbers of nodes and an output layer can be set up. At each node, an activation function is applied to modify the incoming signal. Along each path between two nodes a weight factor is applied to the signal. These weights are modified during the training process in a way to minimize a defined loss function. The loss function provides a measure of error by comparing the final output of the network and the target values. This error is propagated backwards through the net and each weight is updated according to the gradient of the loss function weighted by the learning rate. The complexity of a network can be modified by varying the number of hidden layers and the number of nodes per layer. Furthermore, regularization options such as a drop out of randomly selected connection between nodes are available to prevent the network from specializing for the training data set (overfitting). Different software packages are available to build neural networks.

In this study, the Keras software package, a deep learning application programming interface (Chollet et al., 2015) written in Python, is utilized via the **keras** package for R (Chollet, Allaire, et al., 2017) with the TensorFlow (Abadi et al., 2015) machine learning platform selected as backend. Brightness temperatures from 15 unique emissive MODIS bands (Section 6.2.1.1) and the corresponding fog state (fog or dry) from the ground-based reference data set (Section 6.3.1) are used as input and target variables, respectively. The neural network architecture consists of an input layer with 15 nodes, one for each selected MODIS channel, several hidden layers with varying numbers of nodes and an output layer with one node (cf. Fig. 6.4). Hyper-parameters which have been chosen to maximize the accuracy according to some initial testing are listed in Table 6.3.

The neural network is trained in two different modes. For the first mode, all observations from all considered stations are randomly split into a training (75 %) and a test sample (ALL mode). For the second mode, observations from all considered stations except one are used for the training, and the station left out is used for evaluation (leave one out; LOO mode). An additional 20 % of the training data are set aside by the network itself so that the loss and accuracy of the model can be calculated for training and validation data separately during the training process (Fig. 6.5 b, c). This allows to evaluate whether the model has converged during the training process.

The output of neural network is not a binary classification immediately. Instead, the sigmoid activation function of the output node

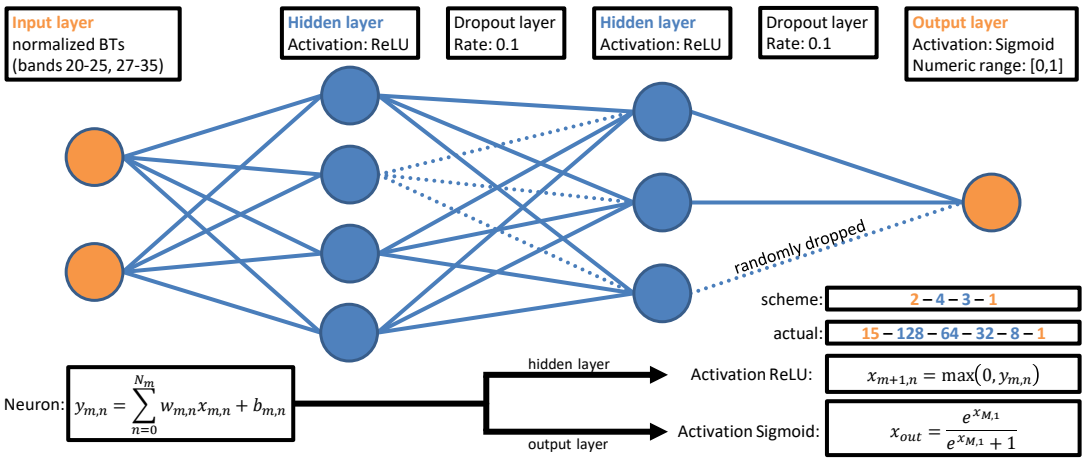

Figure 6.4: Schematic illustration of the setup of the neural network. Normalized MODIS brightness temperatures (BTs) are inserted into the network via the input layer (left two orange circles), which consists of 15 nodes (neurons, only two shown). At each node of a hidden layers (blue circles), the output $y_{m,n}$ from each node n of the previous layer m is used as input $x_{m,n}$. By applying the respective weights $w_{m,n}$ and biases $b_{m,n}$ to each input and summation over all these terms, the output for a hidden layer node is created. After the Rectified Linear Unit (ReLU) is applied as activation function, it serves as input for the nodes of the next layer. For the output of the final hidden layer, the Sigmoid activation function is applied, so that the final output (orange circle on the right) of the network is a value ranging between 0 and 1. To regularize the network (avoid overfitting), 10 % of the connections are randomly dropped after each hidden layer (dotted lines). For illustration purpose, the schematic only shows 2 hidden layers with 4 and 3 nodes, respectively. The setup of the actual chosen model consists of 4 hidden layers, with 128, 64, 32 and 8 nodes, respectively.

returns a value x_{out} in the range $0 \leq x_{\text{out}} \leq 1$ which can be seen as a probability of fog occurrence. To obtain a binary classification, a threshold for fog prediction has to be applied at which the output is divided into dry (below threshold) and fog (above threshold). Once the binary classification is determined according to the chosen fog prediction threshold, the true positive rate (TPR) and the false positive rate (FPR) can be calculated along with other skill scores (for definition see Appendix). By varying the fog prediction threshold, TPR and FPR vary accordingly. In a TPR vs. FPR diagram, also known as receiver operating characteristic (ROC) curve, the model performance can be visualized. Ideally, a low FPR coincides with a high TPR. However, in reality, a higher TPR usually is accomplished at the cost of a higher FPR.

To determine the optimal number of hidden layers and respective numbers of nodes, multiple neural networks were trained applying several setups. To evaluate the performance of each model, the area

Table 6.3: Parameters and schemes implemented in the neural network used for this study.

feature	value
activation fct. hidden layer	rectified linear unit (ReLU)
activation fct. output layer	sigmoid
dropout rate	10 %
iterations (epochs)	100
batch size	1000
weight update scheme	Adam optimization algorithm (Kingma and Ba, 2014)
learning rate	0.001
loss function	binary cross entropy

under the ROC curve (AUC) is determined (Fig. 6.5a) which serves as a measure of separability of the two classes (fog, no fog) by the trained neural network. A perfect separation would result in an AUC of 1, while no separation skill would result in an AUC of 0.5. Increasing the number of hidden layers from 1 to 4 steadily increases the AUC, indicating a better performance for deeper models. Adding a 5th hidden layer does not yield further improvement. Increasing the number of nodes for each of the 4 hidden layers slightly increases the AUC, indicating a better performance for wider models. After these considerations, a model is chosen with 4 hidden layers, consisting of 128, 64, 32, and 8 nodes, respectively in that order. For this model, the evolution of the loss and the accuracy (ACC) indicate that the model is sufficiently trained with marginal overfitting (Fig. 6.5 b, c). Its capability to separate between the two classes is manifested in the two separate peaks revealed by the distributions of the output of the trained neural network for fog and dry conditions according to the ground-based reference, respectively (Fig. 6.5d).

6.3.4 *MODIS cloud top height*

The MODIS cloud product provides the CTH above sea level. Since we are studying fog over land, we need to apply the heights above ground level (a.g.l.). The conversion is done by subtracting the elevation of the climate stations from the corresponding MODIS CTHs. In case of very low cloud top heights and complex topography with the station being higher than the satellite-based CTH, this can result in negative CTHs a.g.l.

The MODIS cloud top heights which are collocated with the climate stations reveal a bimodal distribution (Fig. 6.6a). While fog situations yield a pronounced peak between 2 and 4 km and a less pronounced peak around 11 km (high clouds, cirrus), dry situations result in a

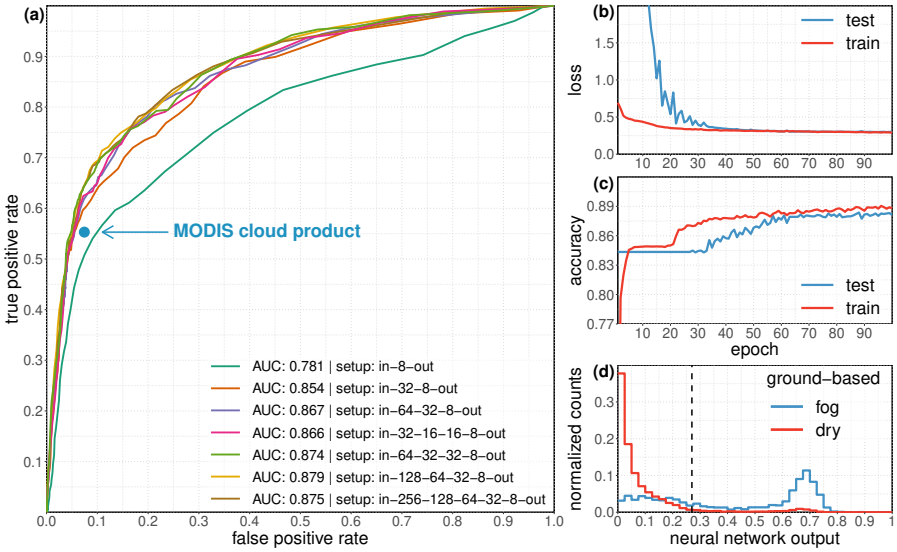

Figure 6.5: (a) Receiver operating characteristic curve of the neural network fog prediction calculated from the test data set (not used for training). Different curves (color) represent different architectures (numbers of layers and nodes) which are indicated by the legend along with the respective area under curve (AUC). True positive rate and false positive rate have been additionally determined for the MODIS cloud top height approach to detect fog (blue circle). (b) Evolution of the loss which represents the deviation between reference and neural network predicted values after each training iteration (epoch) calculated using the binary cross entropy function. It is shown for the training data (red) and test data (blue) separately. (c) Evolution of the accuracy of the neural network over the number of training iterations (epochs) for training data (red) and validation data (blue). To determine the accuracy (portion of correct classification, see Appendix for definition), the binary classification is made by rounding the network output at 0.5 which results in 0 (dry conditions) or 1 (fog conditions). (d) Histogram of the neural network output for fog (blue) and dry (red) conditions according to the ground-based reference classification. The y-axis represents the number of counts for each bin normalized by the total number of observations for each condition, respectively.

more uniform distribution with the low cloud peak being less distinct from the high cloud peak than for fog situations. This indicates, that fog occurrence is less likely alongside high cloud presence. However, fog events with simultaneous presence of on optically thick enough high cloud are missed because the view is obscured so that the CTH of a possible lower cloud cannot be provided. Furthermore, the low cloud peak is shifted to lower cloud top heights for dry situations, indicating that the lowest observed clouds are typically not associated with fog occurrence. An explanation might be, that nocturnal fog typically coincides with a ground inversion. This would lead to ambiguous cloud

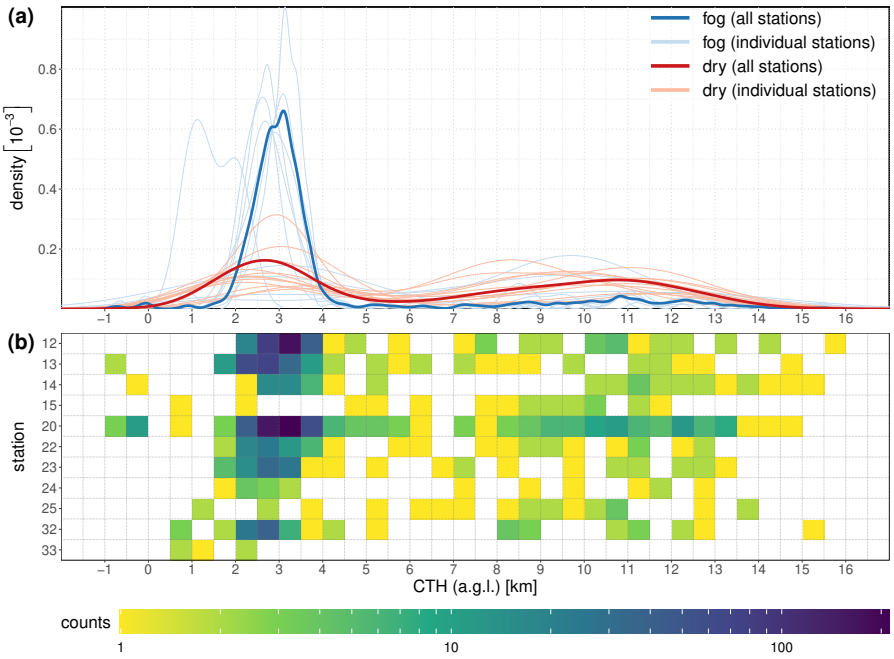

Figure 6.6: Distribution of MODIS cloud top height (CTH) above ground level (a.g.l.) for the nocturnal overpasses over the considered climate stations. (a) Normalized density of CTH distinguished by concurrent fog (blue shadings) or dry (red shadings) conditions at collocated climate stations. Densities are shown considering all stations together (thick lines) and for the individual stations (thin lines). (b) Counts of events for each MODIS CTH bin and each station under fog conditions according to the collocated station measurement. MODIS CTH is originally given above sea level. CTH a.g.l. is obtained by subtracting the elevations of the respective climate stations. In case of complex topography, this can lead to negative CTHs a.g.l.

top height retrievals for which the MODIS CTH retrieval algorithm chooses the highest possible option (Menzel et al., 2008; Baum et al., 2012, cf. Section 6.2.1.2). The distributions for fog situations for the individual stations peak similarly around 3 km except station 33 for which the peak is seen around 1 km. However, station 33 has detected almost no fog events (Table 6.2) which also manifests in very low counts within the histogram (Fig. 6.6b).

The goal is to refine a CTH range which gives the best prediction of fog occurrence. Therefore, a combination of a lower and an upper threshold height is estimated which maximizes the Heidke Skill Score (HSS). The best model is obtained if cloud top heights between 2000 and 3750 m are declared fog. This approach yields about 87 % correct classifications (accuracy) with a probability of detection of 55 % and a false alarm rate of 43 %. More parameters are listed in Table 6.4. MODIS CTH below the lower threshold, seem to indicate the absence of a ground inversion and thereby unlikely fog conditions. Therefore,

Table 6.4: Statistical evaluation measures based on a 2×2 contingency table for an event based comparison of binary classification (fog or dry) by the fog detection methods and the ground-based stations. Measures are listed for fog detection via MODIS cloud top height (CTH) and via neural network (NNet) including only the test samples which were not used to train the network and including all considered observations, i.e. both training and test samples. The prediction threshold (pred. thresh.) refers to the threshold to discriminate the output of the neural network between fog and dry conditions. The given threshold maximizes the Heidke Skill Score (HSS) and results in a bias score (BS) closest to unity. Further measures are the true positive rate (TPR), which is also known as probability of detection, false positive rate (FPR), accuracy (ACC), false alarm rate (FAR) and critical success index (CSI). Definitions of these measures are given in the appendix.

model	pred. thresh.	TPR	FPR	ACC	FAR	CSI	BS	HSS
MODIS CTH		0.55	0.07	0.87	0.43	0.39	0.96	0.49
NNet (test sample)	0.27	0.63	0.07	0.89	0.37	0.46	1.01	0.56
NNet (all obs.)	0.27	0.63	0.07	0.89	0.37	0.46	0.99	0.56

the predictive skill of this approach appears to stem from the detection of a possible ground inversion which serves as a proxy of fog presence.

6.4 EVALUATION

The evaluation of the fog detection approaches is carried out under two different aspects. First, the performance is evaluated on the basis of detection of individual events via a contingency table analysis. This way, various statistical measures can be derived which enable a comparison of both approaches (Section 6.4.2). To derive an uncertainty for these statistical measures, ensembles with multiple training runs of the neural network are created first (Section 6.4.1).

The second aspect is the suitability to derive (i) the variability of fog occurrence frequency on different time scales and (ii) a climatology of fog occurrence frequency for the Atacama region. Therefore, the output of the neural network and the classification according to the MODIS CTH and the ground-based stations are viewed as time series. A comparative analysis of the time series for the different locations of the climate stations allows an assessment of the spatio-temporal representativeness of the two introduced fog detection methods (Section 6.4.3). Following this assessment, both methods are applied to derive a 3-year climatology (Section 6.4.4).

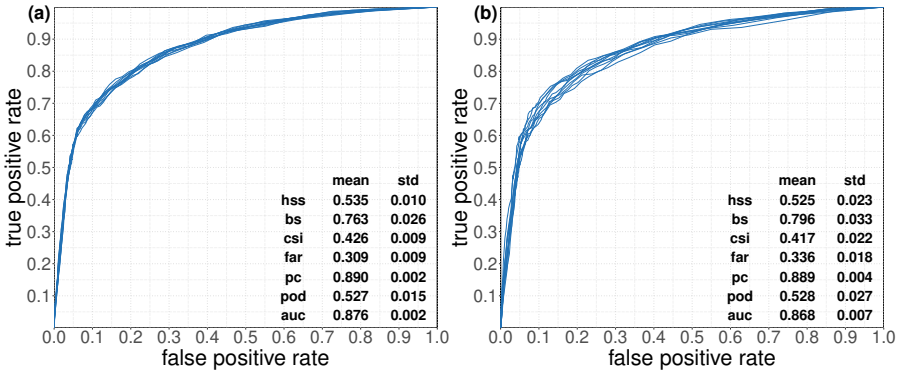

Figure 6.7: Receiver operating characteristic curves for the neural network. (a) 10-member ensemble trained with fixed training data sample. (b) 10-member ensemble trained with randomly drawn training data sample. Ensemble mean and standard deviation are given for the area under curve (AUC) and the following statistical measures (definitions are listed in the appendix): Heidke skill score (hss), bias score (bs), critical success index (csi), false alarm rate (far), percent correct (pc), probability of detection (pod).

6.4.1 *Neural network model sensitivity*

During the training of the neural network, random selection processes influence the values of the final model weights leading to different realizations even with exactly the same training data sample. To quantify the introduced variability, a 10-member ensemble is created which results in an ensemble mean of 0.876 for the AUC with a standard deviation of 0.002 (Fig. 6.7a). This indicates that the improvement, i.e. higher AUC, with increasing depths and widths of the models is mostly beyond two standard deviations (Fig. 6.5) indicating statistical significance since they are all trained with the very same training data sample. However, the improvement due to increased model complexity saturates at the finally chosen setup (Fig. 6.5, in-128-64-32-8-out) as only a slight AUC gain is observed compared to the model with slightly fewer nodes (in-64-32-32-8-out, AUC= 0.874) and a slight decrease in the AUC is observed when a 5th layer is added (AUC= 0.875).

Furthermore, the sensitivity to the training data sample is investigated. In a similar fashion, another 10-member ensemble is created by providing each member with a new randomly drawn training data sample, i.e. a new random selection of 75 % of all observations in a random order without replacement. The random processes which resulted in the standard deviation for the first ensemble stay in effect for the second ensemble. With the additional variation of the training data sample, a higher standard deviation can be expected. This time, the AUC is found to be 0.868 with a standard deviation of 0.007 (Fig. 6.7b). For a binary classification obtained by simply rounding the model

output, i.e. a fog prediction threshold of 0.5, a mean Heidke Skill Score (HSS) of 0.525 ± 0.023 is derived. The uncertainties for further statistical measures are given in Fig. 6.7b. The standard deviations of these measures derived here are useful to assess whether any of the different fog detection methods result in significantly different statistical measures.

6.4.2 *Event-based algorithm performance*

As introduced in Section 6.3.3, there is a functional relationship between the statistical measures, such as the Heidke skill score, and the fog prediction threshold (Fig. 6.8). Based on all observations, a maximum Heidke skill score of HSS $= 0.56$ is retrieved for the neural network with a fog prediction threshold of 0.27 (Table 6.4) which means the neural network has a detection skill much better than a determination by chance (HSS ≈ 0). The same fog prediction threshold also results in the smallest bias. The bias score of almost unity (BS $= 0.99$) indicates that the model estimates the total number of fog events almost correct. Based on the independent test data sample (comprising 25 % of all observations), the HSS is maximized for the same fog prediction threshold (0.27). Except for the bias score, all considered statistical measures are basically identical to the results based on all observations (Table 6.4). This further supports that the neural network is valid beyond the training data set at least for the locations of the climate stations included in the study.

Utilizing the MODIS CTH to detect fog results in a lower Heidke skill score (HSS $= 0.49$). While almost the same number of fog events are predicted by the MODIS CTH approach (BS $= 0.96$) compared to the neural network, the true positive rate is lower and the false alarm rate is higher (Table 6.4). This proves, that exploiting the spectral information of the MODIS emissive bands via a neural network outperforms a fog detection algorithm based on the MODIS CTH.

While for the MODIS CTH only 6 MODIS bands between 11 µm and 14.2 µm are considered, the neural network approach takes 15 MODIS emissive bands into account. The additional information from the 10 emissive MODIS between channels 3.8 µm and 9.7 µm or the way the neural network processes the data obviously outweighs the information added into the CTH retrieval by GDAS data.

6.4.3 *Spatio-temporal representativeness*

In order to asses spatial and temporal consistency of the proposed fog detection algorithms, time series of fog occurrence frequency are derived for the locations of the climate stations. To derive the time series, the binary classification from the station measurements and from the MODIS CTH detection approach (CTH inside or outside

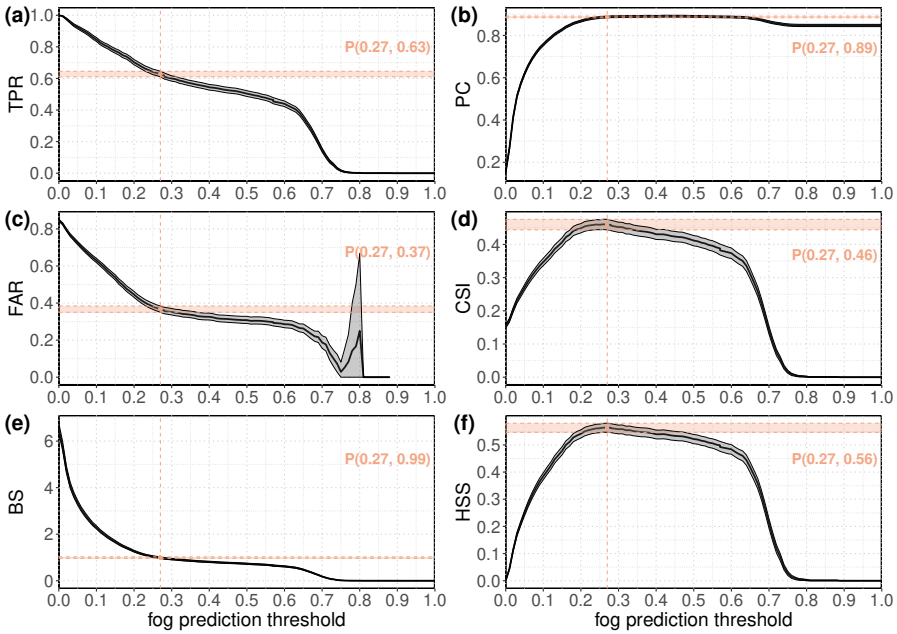

Figure 6.8: Statistical measures calculated for the neural network fog prediction in dependence on the fog prediction threshold which is applied to the network output to obtain a binary classification (fog or dry). The fog prediction threshold which maximizes the Heidke skill score (HSS) is highlighted as vertical orange dashed line in all panels. The point P of the interception with the curve is annotated in each panel (see also Table 6.4). Thick black line denotes the statistics calculated using all observations. Shaded area denotes the area between the 5th and 95th percentile of a distribution derived via a bootstrap resampling of all observations with 1000 iterations.

the designated height range) are converted to numeric values (fog$\widehat{=}$1, dry$\widehat{=}$0) to quantify the fog state numerically. The output of the neural network can be viewed as a fog probability for this analysis. The fog prediction threshold to make a decision for each event is not needed at this time because averages over varying time intervals are calculated so that a decision for each event is no longer necessary. Using the probabilistic output (NNprob) or the conversion to binary (NNbin) prior to deriving the time series may result in different mean fog frequencies depending on the sum of negative and positive residuals when the neural network output is converted to binary. Depending on whether the negative residuals outweigh the positive residuals, NNbin will result in lower or higher mean fog frequencies compared to NNprob. A priori, it is not clear which mode is more appropriate with better agreement to the ground-based observations. We decided to investigate NNprob first and discuss resulting differences for NNbin thereafter. To allow a fair comparison, the time series comprise only coincidental observations from MODIS and the climate stations. For every day, the mean fog state is calculated for each fog retrieval

approach. Then, centered moving averages using various interval lengths are applied to each of these daily resolved time series.

First, we assess the temporal consistency by comparing the time series from the neural network trained with the sample of observations including all stations (ALL training mode) to the ground-based reference. Here, station 13 is used as an example (Fig. 6.9a). On a synoptic scale (7-day moving average), many fog peaks are well in agreement with an overall Pearson correlation coefficient of $r = 0.75$. Extending the time interval of the moving average to a subseasonal scale (60-days) brings out seasonal variations which are represented well by the neural network ($r = 0.91$). Higher frequencies during late winter and early spring and lower frequencies during late summer and early fall with more pronounced seasonal cycles near the coast, which is reported in the literature for the study region (e.g. Farías et al., 2005) and in particular for sites at the coastal cliff and the coastal mountain range (e.g. Río et al., 2018), are consistently revealed by both the neural network and the ground-based fog retrievals and furthermore by the MODIS CTH based retrieval (Fig. 6.10a) with a slightly lower correlation to the ground-based reference ($r = 0.77$). Additional time series for each station can be found in the supplement (Appendix A.3).

Extending the analysis to all stations, correlations between time series derived from the neural network and the ground-based reference data set are studied in dependence on the moving average interval lengths (Fig. 6.9b). For increasing interval lengths, the correlation typically increases, which indicates that a better fog frequency representation is obtained on longer time scales. Overall high correlations, in particular on subseasonal scales prove the suitability of the neural network to represent seasonal and interannual variability of fog frequency. The same results hold true, although with overall lower correlation, for the MODIS CTH based fog detection (Fig. 6.10b).

The root mean square error (RMSE, Fig. 6.9d) decreases with increasing interval length for the moving average. It is highest for the stations which have the highest fog occurrence frequency (stations 12, 20 and 32, cf. Table 6.5). Therefore, their RMSE in relation to the mean fog occurrence frequency is comparably low (42 %, 20 %, 40 %, respectively). For the other stations, the absolute RMSE is below 4 % fog occurrence frequency for moving average intervals greater than 60 days. For six stations, the relative RMSE is below 45 %.

Next, the spatial consistency of the neural network is investigated by withholding one station during the training of the network and evaluating the time series for that station (LOO training mode). Compared to the neural network trained with all stations, the correlation remains almost the same for 4 stations (within ± 0.03 for stations 12, 13, 15, 20) and differs only slightly for 4 stations (within ± 0.08 for stations 22, 23, 24, 25) (Table 6.5). For station 32, the correlation drops from

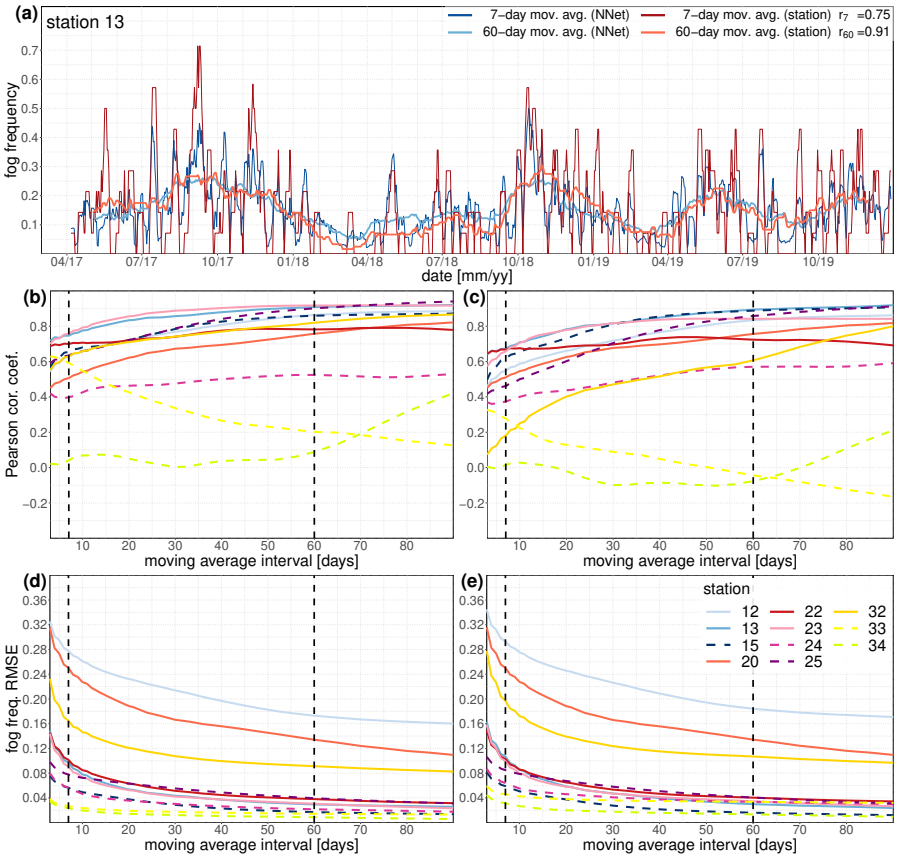

Figure 6.9: (a) Time series derived from the neural network probabilistic output (blue) and the station measurements (red) for a 7-day centered moving average (dark) and a 60-day centered moving average (light) at the location of station 13. Pearson correlation coefficient r between neural network and station derived time series is indicated in the upper right for both moving average intervals. (b) and (c) Pearson correlation coefficient and (d) and (e) Root mean square error (RMSE) in dependence on interval length of the moving average for each climate station. Two different training modes are distinguished: training on all stations (ALL mode: b, d) and leaving one station out from training and derive the statistics for this station (LOO mode: c, e). Dashed lines indicate stations with very low fog occurrence frequency ($f \leq 2\,\%$). Black dashed vertical lines highlight the moving average intervals for which the exemplary time series (a) is shown. The time series (a) is based on the ALL training mode. Time series for all stations based on the LOO training mode are given in the supplement (Appendix A.3).

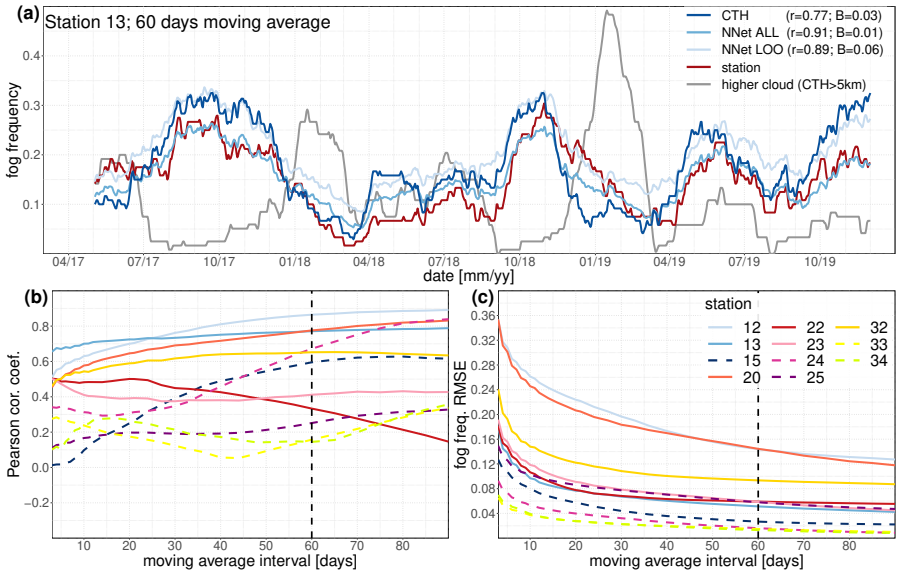

Figure 6.10: (a) Comparison of fog frequency time series derived from the fog detection based on the MODIS CTH (dark blue), based on the neural network probabilistic output in both ALL training mode (lighter blue) and LOO training mode (lightest blue), and based on climate station measurements (red) at the location of station 13. Time series is smoothed via a 60-day centered moving average. Pearson correlation coefficient r and bias B between the time series of the proposed detection algorithms and the reference time series (station) are indicated in the legend, respectively. (b) Pearson correlation coefficient and (c) root mean square error (RMSE) between time series derived from the detection via MODIS CTH and from the station measurements in dependence on the interval length of the moving average for each station. Dashed lines indicate stations with very low fog occurrence frequency ($f \leq 2\,\%$). Black dashed vertical lines highlight the moving average intervals for which the exemplary time series (a) is shown (60-day). Time series for other stations are provided in the supplement (Appendix A.3).

0.82 to 0.61 when it is left out from the training of the neural network. The comparably high drop might be due to a slightly different fog signature around the southern transect. Therefore, the neural network gains additional expertise when station 32 is included in the training. While for the other transects, more fog events from nearby stations can introduce the local fog structure, this is not the case for station 32 because the remaining stations of the southern transect do not provide substantial fog events which the network could learn from. For stations 33 and 34, which experience a fog occurrence frequency of almost 0 with very little variance so that correlation coefficients cannot be meaningful, the correlations decrease to insignificant values. The overall similar correlations compared for the two training modes (include all stations and LOO) indicate that the neural network ap-

Table 6.5: Statistical metrics such as the Pearson correlation coefficient r, bias, and root mean square error (RMSE) derived from the comparison of fog occurrence frequency time series based on the neural network and the MODIS cloud top height (CTH) classifications. For each station, the fog occurrence frequency (fof) over the entire available time is listed. Results are shown for the neural network trained on all climate stations excl. station 14 (subscript "ALL"), for the neural network trained on all stations excl. station 14 and leaving out the station for which the statistical metrics are displayed (subscript "LOO"), and for the fog detection algorithm based on the MODIS CTH (subscript "CTH"). Metrics are calculated for time series which were smoothed applying a 60-day moving average. Values in parenthesis denote the bias and RMSE in relation to the overall fof. The * denotes correlations which are not significant according to the p-value ($p > 0.05$)

stat	fof	r_{ALL}	r_{LOO}	r_{CTH}	$bias_{ALL}$	$bias_{LOO}$	$bias_{CTH}$	$RMSE_{ALL}$	$RMSE_{LOO}$	$RMSE_{CTH}$
12	0.41	0.86	0.83	0.86	-0.04 (-10%)	0.03 (7%)	0.05 (13%)	0.17 (42%)	0.18 (45%)	0.14 (35%)
13	0.15	0.91	0.89	0.77	0.01 (4%)	0.06 (37%)	0.03 (19%)	0.03 (20%)	0.03 (20%)	0.05 (35%)
14	0.30	-0.00*		-0.60	-0.21 (-70%)		-0.23 (-76%)	0.16 (52%)		0.19 (62%)
15	0.01	0.86	0.89	0.60	0.03 (184%)	0.05 (365%)	0.02 (114%)	0.02 (118%)	0.02 (111%)	0.03 (185%)
20	0.67	0.76	0.76	0.78	-0.27 (-41%)	-0.32 (-48%)	-0.17 (-26%)	0.13 (20%)	0.13 (20%)	0.14 (22%)
22	0.11	0.78	0.72	0.33	0.02 (18%)	0.04 (34%)	-0.00 (-2%)	0.04 (34%)	0.04 (36%)	0.06 (52%)
23	0.11	0.92	0.84	0.41	0.06 (50%)	0.10 (91%)	0.02 (18%)	0.03 (27%)	0.03 (31%)	0.06 (52%)
24	0.02	0.52	0.57	0.67	0.03 (164%)	0.08 (360%)	0.01 (28%)	0.02 (102%)	0.03 (158%)	0.02 (76%)
25	0.02	0.90	0.86	0.25	0.03 (112%)	0.05 (202%)	0.02 (108%)	0.04 (173%)	0.04 (178%)	0.06 (258%)
32	0.23	0.82	0.61	0.65	-0.07 (-28%)	0.01 (6%)	-0.13 (-57%)	0.09 (40%)	0.11 (46%)	0.09 (40%)
33	0.01	0.20	-0.04*	0.16	0.03 (487%)	0.07 (1311%)	0.01 (190%)	0.01 (278%)	0.03 (652%)	0.01 (287%)
34	0.00	0.09*	-0.08*	0.14	0.02 (999%)	0.04 (1854%)	0.02 (1018%)	0.01 (382%)	0.01 (586%)	0.01 (572%)

proach has potential to be generalized region wide even for locations it has not seen observations from during its training. This is further supported by the RMSE which remains within 2 % fog occurrence frequency for all stations when the two training modes are compared (Fig. 6.9 d, e and Table 6.5).

When the binary classification is applied to the neural network output (NNbin), basically identical correlations with the ground-based reference result for the time series for most climate stations with significant correlations. Only for station 22 (increase of 0.10) and stations 23 (decrease of 0.06) and 25 (decrease of 0.09) greater differences of the correlations are found for NNbin. This indicates, that the temporal representation of the fog frequency by the neural network does not depend much on the choice between NNprob and NNbin.

To further assess the spatial representativeness, the biases which result for different stations are investigated. Applying the LOO training mode of the neural network and a 60-day moving average for the derivation of the time series, biases for the locations of the climate stations range between 0.01 and 0.10 of fog frequency (Table 6.5, Fig.

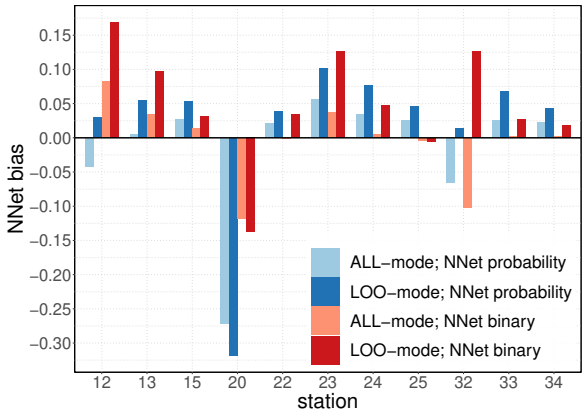

Figure 6.11: Bias between neural network derived fog frequency and reference data for each climate station. Distinguished are two different processing modes of the neural network output: using the fog probability without (blue shades) or with (red shades) the application of the binary classification prior to deriving the temporal average. Two different training modes of the neural network are considered: training with samples from all stations (ALL-mode, lighter shades), leave out one station and derive evaluation measures for that station (LOO-mode, darker shades).

6.11). This means the neural network overestimates the fog presence compared to the reference data set. An exception is station 20 for which a fog frequency bias of -0.32 is determined. An explanation for this strong dry bias could be the characteristics of the location of station 20. A strong topographic gradient at the coastal cliff with strong variations at scales smaller than the MODIS resolution (1 km) might lead to classification mismatches between the neural network and the station derived fog status. However, even for this challenging scene the correlation reaches $r = 0.76$ for the 60-day moving average interval which indicates that the high bias might have a different origin which possibly lies in the methodology of the retrieval. If the output of the neural network is converted to a binary fog state (NNbin), the bias can be decreased (absolute value) to -0.14 (Fig. 6.11). This decrease is mainly due to an increase of the fog frequency for the peak fog season during winter compared to the probabilistic output of the neural network (NNprob) (time series for NNbin not shown). However, applying the binary classification does not generally improve the bias. For instance, for station 12, the bias increases from 0.03 to 0.17 (Fig. 6.11).

Comparing the biases between ALL and LOO training mode reveals mostly higher wet biases with an increase ranging between 2 and 8 % (Table 6.5, Fig. 6.11). For station 13, for example, the fog frequency lies systematically higher throughout the considered period if the neural network has not seen any observations from this station (Fig. 6.10a). This illustrates how the neural networks is learning from the observations. While the observations from other locations generally

suffice to detect fog with great temporal representativeness ($r = 0.89$), it can learn more scene specific details once it is presented with local observations leading to further improvement ($r = 0.91$) and in this case bias reduction (from 0.06 to 0.01).

To investigate the neural network quality for a location that has not bee incorporated into the training base, we investigate the time series for station 13 based on the LOO training mode in more detail (Fig. 6.10a) For winter and spring 2017 (August–November), both the neural network and the MODIS CTH approach overestimate the fog frequency while agreeing very much among each other. The same is observed for the peak seasons in 2018 (September–October) and in 2019 (October–November). A possible reason could be, that very low clouds are present and classified as fog, even though a portion of them might not touch the ground and thus would be a false alarm with respect to the ground-based observation. The neural network appears to be able to distinguish these low cloud scenes from fog scene if it is provided with some of the local observations ($\rightarrow$ ALL training mode). At the end of the spring season in 2017, there is a short period of agreement of satellite and ground-based derived fog frequencies in December. Thereafter, the MODIS CTH based fog frequency stays in agreement with the ground-based retrieval while the neural network develops an increasing positive bias throughout the subsequent summer (December 2017 – April 2018). A possible explanation for this bias can be found in the coincidental rise of the high cloud frequency which is derived from the MODIS cloud top height by considering any cloud above 5 km a high cloud for this particular investigation. Such high clouds may introduce a positive bias for the neural network retrieval.

Investigating the distribution of the fog probability given by the neural network in dependence on the MODIS CTH for station 13 reveals the highest fog probability as expected for CTHs between 2 and 4 km (Fig. 6.12). This is consistent with the definition of the fog retrieval based on the MODIS CTH. Furthermore, for clear sky events according to the MODIS cloud product, the median fog probability amounts to about 0.05, which is similar to the overall bias for station 13 (0.06, Table 6.5). For high clouds (CTH $>$ 5 km), the distributions typically exceed the clear sky benchmark with median fog probabilities as high as 0.2. While it is possible that advection fog occurs while a high cloud is present, radiation fog is less likely. A nocturnal high cloud would prevent the surface from efficient cooling leading to less favorable conditions for fog formation. Thus the higher fog probability in the presence of high clouds might indicate that the network makes notice of the cloud and distinguishes the scene from a clear sky scene by slightly increasing the probability output. Thus, the arising positive bias of the neural network fog frequency for summer 2018 appears to be caused by the enhanced high cloud frequency. A similar

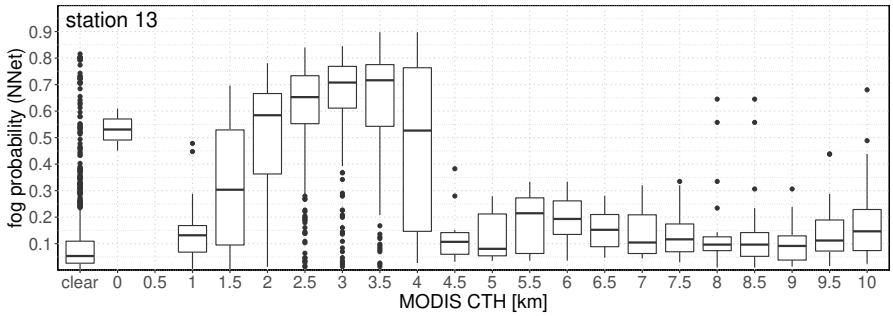

Figure 6.12: Fog probability predicted by the neural network in dependence on MODIS CTH. Boxplots at a specific MODIS CTH represent the distribution of fog probability for a range within ±500 m of the respective MODIS CTH. Thick horizontal lines denote the median, boxes the 25th and 75th percentile, upper and lower whisker the maximum and minimum, respectively, but not farther than 1.5 · IQR (inter-quartile range) away from the box with the remaining values plotted as dots (outliers). The first boxplot on the left represents clear sky situations.

phenomenon appears for the following summer (December 2018–March 2019). This time the high cloud frequency peaks even higher reaching almost 50 % in January 2019. For this summer, the neural network yields a higher fog frequency compared to the MODIS CTH retrieval similar to the previous summer. However, the ground-based fog frequency is higher this time, thus, in better agreement with the neural network. The negative bias for the MODIS CTH based fog frequency might be due to the enhanced high cloud frequency causing overall more dry classifications for this period. Even though, the high cloud frequency shows a strong seasonality, no clear indications for a seasonal dependence of the neural network bias exist neither for the probabilistic (NNprob) nor the binary (NNbin) consideration (Fig. 6.13 a, b). The fog frequency based on the MODIS CTH does not reveal a seasonal dependence of the bias either (Fig. 6.13c) which indicates that the scenes hidden by high clouds are typically dry scenes.

6.4.4 *Climatology*

To exploit the indicated potential of the neural network to represent fog within the entire region, a 3-year climatology (2017–2019) is derived. Both, the probabilistic output (NNprob) and the binary classification according to the fog prediction threshold (NNbin) are applied. For comparison, climatologies are also derived based on fog detection via the MODIS CTH.

For austral winter (July, August, September), the neural network and the MODIS CTH approach both reveal that the fog occurrence frequency is very high in the coastal regions ($> 50\,\%$) and low for most land areas away from the coast ($< 5\,\%$) (Fig. 6.14 a, b, c). For

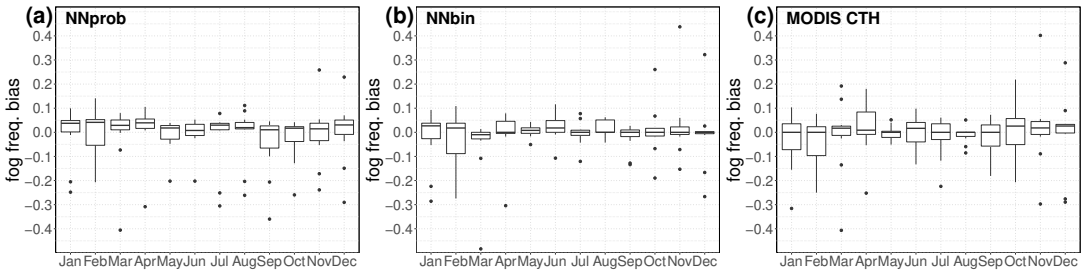

Figure 6.13: Monthly bias assessment from a comparison of the ground-based reference to the fog frequency derived from the neural network using the probabilistic output (NNprob) (a) and applying the fog prediction threshold for the binary classification (NNbin) (b) and derived from the MODIS CTH (c). Bias is defined as difference of the mean fog frequency (neural network or MODIS CTH minus ground-based reference). The neural network trained with samples from all stations is considered (ALL mode).

NNbin, even higher coastal frequencies ($> 70\,\%$) and lower inland frequencies are found increasing the West–East gradient. This pattern is expected because the near coast maritime stratocumulus is most persistent yielding the highest cloud cover (Farías et al., 2005; Cereceda et al., 2008b; Muñoz et al., 2016; Lehnert et al., 2018b) during austral winter. Furthermore, due to the lower cloud heights during winter (Muñoz et al., 2016; Böhm et al., 2019), the stratocumulus is more prone to intersect with the coastal cliff and mountain range which prevents further inland advection. Inland penetration is visible for both fog retrieval approaches at corridors were the coastal cliff is intercepted by canyons or generally lower, e.g. at the northern end around the Peruvian border. Such fog corridors have been identified and related to fog occurrence in the central depression (Farías et al., 2001; Farías et al., 2005). Individual areas within the central depression are standing out with enhanced fog frequencies up to 50 % for NNprob and MODIS CTH and even higher frequencies for NNbin, in particular between 20°S and 21.5°S where no validation data are available. Across the study region, the neural network estimates slightly higher fog occurrence frequencies compared to the MODIS CTH approach which is consistent with the higher bias shown for the LOO training mode (Table 6.5). However, higher fog frequencies are expected at the coast as daily mean low stratus frequencies exceeding 50 % for August 2001 based on retrievals from GOES have been reported (Farías et al., 2005; Cereceda et al., 2008b). For the nocturnal MODIS overpasses, even higher frequencies are expected considering the aforementioned diurnal cycle. Therefore, the NNbin based coastal frequencies seem most plausible. The dry bias found for NNbin for station 20 (Table 6.5, Fig. 6.11), which is the closest station to the coast, indicates that even for NNbin the coastal frequencies might still be underestimated. This means, the differences between coastal region and inland could

be even more amplified and the CTH approach might underestimate the coastal fog frequency. However, the overall patterns are mostly in agreement and appear plausible.

Consistent with the previous discussion, both approaches agree that fog occurrence is reduced at the coast (Fig. 6.14 e, f, g) for austral summer (January, February, March). A distinctly enhanced coastal fog frequency is only apparent south of 24°S for both approaches. Besides these agreements, the patterns derived from the neural network and the MODIS CTH differ overall. While the neural network reveals a North–South gradient with values up to 15 % in the North and below 5 % in the South based on the probabilistic output (NNprob) and with slightly lower frequencies for NNbin, the MODIS CTH approach results in a West–East gradient with values up to 35 % in the coastal cordillera decreasing to values below 5 % towards the Andes. For the neural network, the higher fog frequency in the north could be related to higher cirrus cloud frequencies potentially increasing the fog probability. These cirrus clouds are a recurring feature visible in the fog frequency time series for stations of the center and more so of the northern transect (Fig. 6.10 and supplement (Appendix A.3)) consistent with what is expected from the typical summer circulation with upper level moist easterlies for the northern Atacama (Garreaud et al., 2003). Further investigation on the influence of high clouds are necessary to draw a more definitive conclusion.

As for the winter season, both approaches show enhanced fog occurrence frequencies for some parts of the central depression for the summer season. However, for the neural network, this is hardly pronounced. On the contrary, the MODIS CTH approach reveals values exceeding 50 % for these regions. Cereceda et al. (2008) report that the region between 20°S and 22°S is mostly cloud free inland with some patches revealing mean low stratus frequencies of up to 5 % based on GOES retrievals for January 2002. Even though in their calculations all diurnal observations are included and the fog peak is expected at night, their revealed stratus frequency agrees more with the lower inland fog frequencies based on the neural network. It is possible that the height range applied to infer fog occurrence from the MODIS CTH has a seasonal dependence and is not appropriate for the summer season. As the predictive skill of the method presumably originates in the detection of a surface inversion, the designated height range for optimal fog detection can be expected to depend on the inversion strength and height. Therefore, the seasonal cycle of the inversion with higher base heights and weaker temperature jumps during the summer season (Muñoz et al., 2011) might lead to misclassifications. For a robust estimate of the potential seasonality of the optimal height range, a longer reference data record would be necessary.

Another feature from the summer climatology is that the MODIS CTH approach shows a much higher spatial variability on a scale of

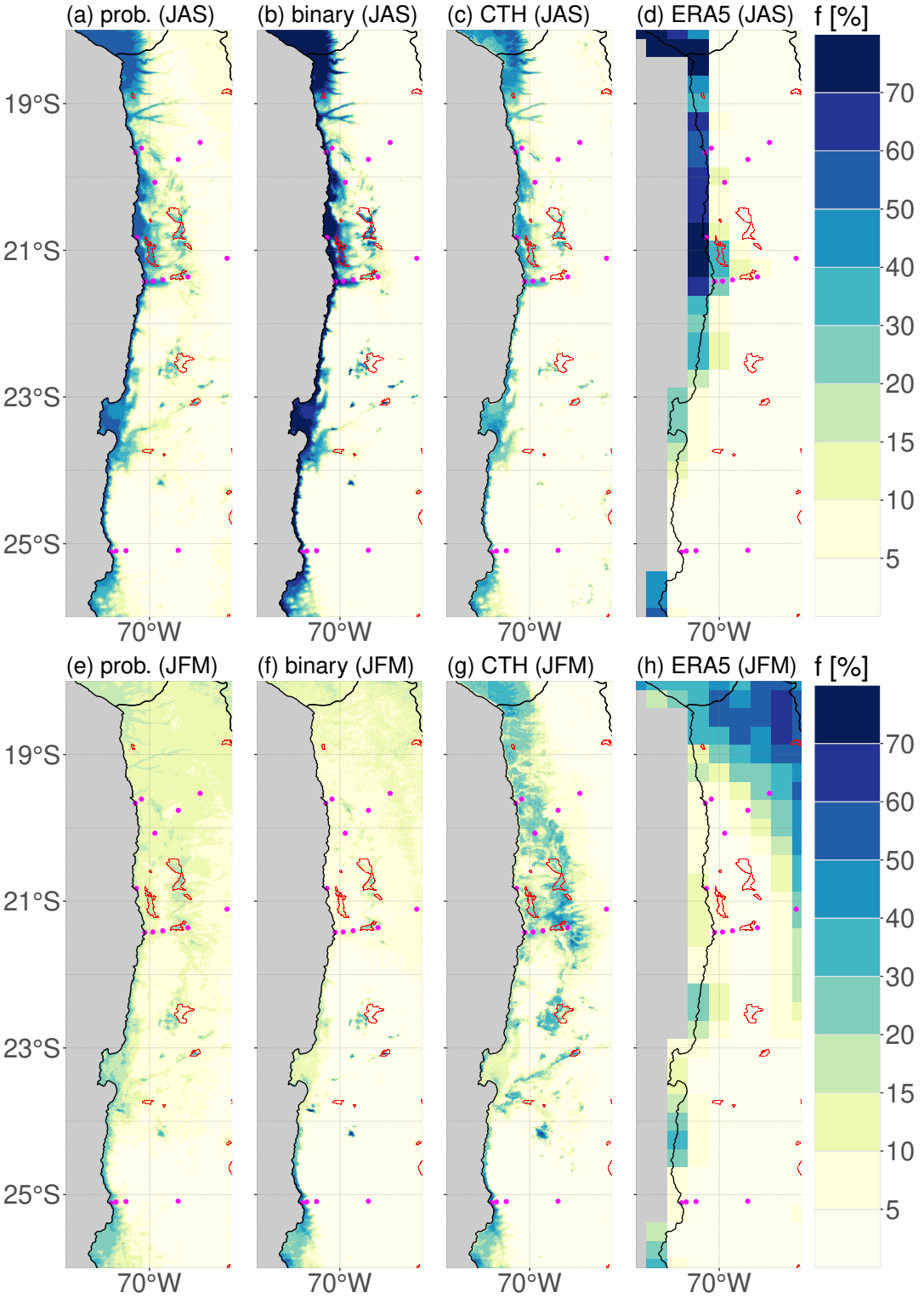

Figure 6.14: Seasonal climatology of fog occurrence frequency f for austral winter (JAS; a, b, c, d) and summer (JFM; e, f, g, h) based on the neural network probability output (prob.; a, e), neural network binary classification applying to the threshold of 0.27 derived in Section 6.4.2 (binary; b, f), MODIS cloud top height (CTH; c, g) and ERA5 low cloud cover (ERA5; d, h). Climate stations are indicated by magenta circles. Salt flats (salars) according to Albers (2013) are outlined in red.

a view kilometers compared to the neural network. Furthermore, for the neural network probabilistic based summer climatology (NNprob), canyons mouthing into the Pacific protrude with a higher fog frequency in particular north of 21°S. For NNbin, these local structures are not visible. For the MODIS CTH based summer climatology, an inverse pattern is observed: fog frequencies appear lower at the canyon compared to the surroundings. While it seems more plausible that these canyons allow more frequent inland penetration as it is observed for both approaches for the winter season, we cannot validate this any further at this point. Overall, further ground-based reference measurements are necessary to disentangle the observed discrepancies between the neural network and the MODIS CTH based summer climatologies.

6.5 CONCLUSION

This study introduces a new satellite-based fog retrieval approach for the Atacama region which utilizes a neural network trained with a novel ground-based reference data set. An attempt is made to derive a region wide climatology of fog frequency. The development of this approach benefits from a new network of climate stations deployed at various locations throughout the Atacama. Based on a leaf wetness sensor, and additional thresholds for relative humidity, difference of surface and air temperature and the longwave radiation budget, which are derived using self-organizing maps, a ground-based reference fog data set is derived which is utilized to train and validate the retrieval method. An uncertainty assessment for the reference data is difficult. For future validation of the ground-based fog retrievals, an additional installation of visibility sensors would be beneficial. Furthermore, MODIS CTHs are included in this study to develop an alternative fog retrieval method for comparison based on simple height thresholds.

A 2×2 contingency table analysis based on the binary classification of individual events results in an overall accuracy of 0.89, a POD of 0.63, a FAR of 0.37 and a HSS of 0.56 for the neural network when an optimal fog prediction threshold of 0.27 is applied to convert the probabilistic neural network output into a binary classification. For another recently developed satellite-based algorithm for fog and low cloud detection for the Namib region, a similar subtropical west coast desert environment, an accuracy of 0.97 and a HSS of 0.89 (Andersen and Cermak, 2018). However, their approach does not make a distinction between fog and low clouds which is attempted here. A pure fog detection approach is presented by Egli et al. (2018) who report a HSS of 0.58 for a satellite-based fog detection method which is validated against visibility provided by METAR and SYNOP reports for various regions across Europe. However, due to the completely

different reference data sets and different environmental conditions for their study region, a comparison to our study is difficult.

To further assess the suitability of the neural network approach to derive a climatology and to study the variability of fog frequency on different time scales, we derived time series for each climate station. For the ground-based reference and the MODIS CTH based fog retrieval, the time series are derived from the binary classification of individual events by applying moving averages of various interval lengths. For the neural network, both the probabilistic output (NNprob) and the binary classification according to the fog prediction threshold (NNbin) are applied to derive the time series. On a subseasonal scale (60-day moving average), Pearson correlation coefficients between the time series based on the neural network and the climate stations range between 0.76 and 0.92 for stations with overall fog frequencies greater than 2 % indicating a suitable representation of the temporal variability of the fog frequency. Slightly lower correlations are determined for the MODIS CTH approach.

To investigate whether the neural network is representative for locations aside from the climate stations which are included in the training process, a second training mode is introduced. By leaving out one station from the training and then use it to evaluate the performance (LOO mode), we simulate the situation which is faced once the network is applied to regions which do not host climate stations and hence cannot be trained for. The correlations for the individual climate stations remain similar or drop only slightly for most stations compared to the neural network trained with samples from all stations. Therefore, we conclude that the performance of the network does not depend greatly on the stations used in the training which means the network can be generalized and applied region wide across the Atacama Desert.

While not shown explicitly, we also investigated the suitability of additional input data such as (1) the 10m wind from reanalysis, (2) climatologies of the brightness temperatures for each channel, or (3) percentiles of brightness temperatures within a certain time window around each measurement. While These variables can improve the predictive skill of the model and increase the resulting correlations with the ground-based reference at an individual station, these improvements do not hold anymore when the training is carried out in LOO mode. A similar behavior is found when the station identification number is provided to the network. This indicates that any additional variable used in the training process which has unique signature at each station and thus serves as a proxy for the station identification number will result in an overfitted model which is specialized for the locations used in the training. Therefore, such a model would not be suitable to apply region wide. It may, however, be better suited to detect fog at the same stations. Moreover, incorporating further

neighboring MODIS pixels in addition to the nearest neighbor (3×3 pixels around the station) did not improve the detection skill.

The derived 3-year climatologies based on the MODIS CTH and the neural network reveal very similar patterns for the austral winter (JAS) with slightly higher fog frequencies derived for the neural network. Fog is mainly present at coastal regions and penetrates further inland through fog corridors which is consistent with reports from literature. Both methods reveal fog hot spots within the central depression which might be an indication of radiation fog which forms at night when the near surface layer cools. The required moisture could be advected from the Pacific with the westerly winds which develop typically during the day and reverse later at night (Schween et al., 2020).

The climatology reveals a more puzzling picture for the summer season (JFM). For the neural network a North–South gradient is revealed with higher fog frequencies in the north. This might be due to enhanced high cloud presence in the northern Atacama during the summer season which may introduce a positive bias for the neural network. Further reference measurements are needed. Furthermore, both methods reveal enhanced fog frequencies for some regions within the central depression. However, while this is hardly pronounced for the neural network, the MODIS CTH approach results in fog frequencies even exceeding 50 % for some of these regions. For January 2002, GOES-based low stratus detection revealed lower fog frequencies for this region (Cereceda et al., 2008b), suggesting that MODIS CTH overestimates the fog frequency. However, several issues complicate this comparison (coarser resolution of GOES, all times of the day included in the average, only a single month is considered, etc.). More observations in particular in the central depression would be beneficial to verify which method is more realistic.

An interesting issue could be identified for a station located close to the Salar de Llamara (station 14) which might be representative for a wider range of such salt flats. Unusual infrared temperatures measurements, reported by this station, might indicate a distinct surface emissivity anomaly. If the surface emissivity is significantly different at these salt flats, the ensemble of brightness temperatures for the various MODIS channels may differ for these regions. Therefore different relationships between individual channels would be expected compared to the other regions. Within the Atacama Desert, multiple salt flat regions have been identified (Albers, 2013, marked in Fig. 6.14). Therefore, it would be beneficial to have more ground-based measurements for validation in particular for the salt flat regions.

As we have generated the first satellite-based climatology for the Atacama, there are only reanalysis data available for comparison. Reanalyses provide atmospheric quantities with high spatial and temporal coverage. As an example, we include the low cloud cover derived from the European Centre of Medium-Range Weather Forecasts (ECMWF)

5th generation reanalysis ERA5 (Hersbach et al., 2020) to illustrate the capabilities of contemporary reanalyses. With a horizontal resolution of 31 km, which is comparably high for a reanalysis, it does not provide a realistic representation of the orography in particular of the coastal cliff and cordillera. Therefore, the advection of the stratocumulus deck is not represented correctly. Small corridors which are visible for the other fog detection approaches are not resolved (Fig. 6.14 d, h). This demonstrates that observations with much higher spatial resolution, such as the satellite observations which are utilized in this study, are required in order to study region-wide fog frequencies.

Aside from the salt flats which pose uncertain terrain due to the lack of in-situ reference data for these regions, the neural network fog detection approach reveals a robust representation of the study region and suitable skill to represent temporal variability on subseasonal and to some degree on a synoptic scale. In future studies it can be applied to derive a long term climatology including the entire MODIS data record, which dates back to the year 2000 (Terra) and 2002 (Aqua). This would enable to derive seasonal cycles, study interannual variability and the potential relationship to large scale climate variations such as the El Niño Southern Oscillation (ENSO), and investigate local trends of fog frequency. Furthermore, applying the methodology presented here to GOES-16 measurements, which are available at temporal resolution of 15 minutes, would enable to study the whole diurnal cycle and thus be complementary to the MODIS-based study.

6.6 APPENDIX: DEFINITIONS OF STATISTICAL MEASURES

True positive rate (TPR), probability of detection (POD), false positive rate (FPR), accuracy (ACC), false alarm rate (FAR) and critical success index (CSI), bias score (BS) and Heidke skill score (HSS) are defined as follows:

$$TPR = POD = \frac{a}{a+c} \tag{6.1}$$

$$FPR = \frac{b}{b+d} \tag{6.2}$$

$$ACC = \frac{a+d}{a+b+c+d} \tag{6.3}$$

$$FAR = \frac{b}{a+b} \tag{6.4}$$

$$CSI = \frac{a}{a+b+c} \tag{6.5}$$

$$BS = \frac{a+b}{a+c} \tag{6.6}$$

$$HSS = \frac{2(ad-bc)}{(a+c)(c+d)+(a+b)(b+d)} \tag{6.7}$$

Here, a is the number of true positives (fog hits), b is the number of false positives (false alarms), c is the number of false negatives (missed fog) and d is the number of correct negatives (correct dry).

6.7 REMARKS

SUPPLEMENT Supplement material is available in the Appendix A.3. This material is part of the submission to the Journal of Applied Meteorology and Climatology (Böhm et al., 2020b).

DATA AVAILABILITY Measurement data from the climate stations are available at the Collaborative Research Center 1211 Database (https://www.crc1211db.uni-koeln.de/wd/index.php). MODIS Geolocation Fields Product, level 1B Calibrated Radiances Product, and level 2 Cloud Product were downloaded from the NASA Level-1 and Atmosphere Archive & Distribution System Distributed Active Archive Center (LAADS DAAC) (https://ladsweb.modaps.eosdis.nasa.gov/archive/allData/). ERA5 data were downloaded from the Copernicus Climate Data Store (CDS) via Web-API.

ACKNOWLEDGEMENTS We gratefully acknowledge financial support by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – project number 268236062 – SFB 1211.

7

PRECIPITATION

THE ROLE OF ATMOSPHERIC RIVERS FOR PRECIPITATION IN THE ATACAMA DESERT

Böhm, C., M. Reyers, L. Knarr, and S. Crewell (2020). "The role of atmospheric rivers for precipitation in the Atacama Desert." Unpublished.

ABSTRACT. In the hyperarid Atacama Desert in northern Chile, very rarely occurring precipitation events can leave long-lasting geomorphological traces and have a strong impact on biota. While for central and southern Chile, ARs account for 40–60 % of the total precipitation, their role for the Atacama Desert has not been investigated previously. The presented study reveals that about four ARs per year make landfall in the Atacama Desert. According to simulated precipitation which has been validated against gauge measurements, more than 40 % of the total precipitation is associated with ARs for most parts of the study region. For some locations this fraction even exceeds 70 %. Furthermore, a case study reveals an elevated vertical structure, which is confirmed by a composite analysis. The elevated moisture transport (peak around 500–600 hPa) associated with a mid-tropospheric trough is typically decoupled from the MBL. Near surface moisture transport is guided by the prevailing southeast Pacific anticyclone even during AR passages. The example case for an AR landfall on 7 June 2017 is presented in more detail. Back trajectories reveal the Amazon Basin as a source of moisture. Implications of the results on paleoclimate reconstructions are discussed.

7.1 INTRODUCTION

Enduring extreme hyperaridity together with extraordinary ultraviolet irradiation (Rondanelli et al., 2015) and low soil carbon content (Crits-Christoph et al., 2013; Knief et al., 2020) make the soils of the Atacama Desert one of the most hostile environments of the world for microbial life (Jones et al., 2018). Nevertheless, microorganisms can potentially metabolize if liquid water is present for a sufficient duration (Crits-Christoph et al., 2013). Soil microbial communities are activated by precipitation (Jones et al., 2018), so that more humid locations in the Atacama Desert are characterized by higher microbial diversity and abundance (Knief et al., 2020). On the contrary, sudden surplus of liquid water associated with extreme precipitation can lead to degradation of microbial species which are highly adapted to the prevailing aridity (Azua-Bustos et al., 2018).

Moreover, precipitation events can activate germination of many species (Pliscoff et al., 2017) and even trigger rapid flowering, causing the spectacular "blooming desert" (Chávez et al., 2019; Astorga-Eló et al., 2020). The latter phenomenon attracts many tourists and scientists to the Atacama Desert, which has a socioeconomic impact and generates the demand for a better prediction of such events (Chávez et al., 2019; Astorga-Eló et al., 2020). Extreme precipitation can cause vegetation outburst across areas which may have been barren for decades (Jordan et al., 2019). Vegetation development of such a dimension leaves traces through storage of plant residue in rodent middens which are conserved over geological time scale (e. g. Latorre et al.,

2002; Díaz et al., 2012; Gayo et al., 2012). Therefore, a reconstruction of pluvial episodes is possible by investigation of such geological archives.

Depending on the magnitude of the precipitation and terrain characteristics, precipitation can also trigger debris flow (Vargas et al., 2006). The debris flows are associated with dominant morphodynamic processes which shape coastal alluvial fans (Haug et al., 2010; Walk et al., 2020). Therefore, such precipitation episodes and their atmospheric driver are strongly related to landscape evolution. By dating alluvial fan depositions, climate history can be reconstructed (Bartz et al., 2020).

Another important aspect of precipitation is the origin of the water. The source and pathway define the isotopic fractionation of the water, which, in turn, modifies the isotopic composition of the rain water, run-off and vegetation (Jordan et al., 2019). The isotopic composition is conserved, for example, in rodent middens and rodent teeth, which constitute climate archives and can be used to reconstruct prevailing moisture sources for past climatic episodes (Jordan et al., 2019).

The fingerprints of the present climate are manifested in representative quantities, such as grain size and stable oxygen and hydrogen isotopes. Extensive field work has been carried out to unleash these precipitation archives and infer climate history from it (Ritter et al., 2019; Diederich et al., 2020). For a better interpretation of these archives, it is necessary to understand the drivers and processes which guide precipitation variability.

So far, the precipitation regimes in the Atacama Desert have mostly been divided into two main areas which are affected by different drivers. The southern and coastal part of the desert are characterized by winter rain which is associated with cut-off lows and troughs emanating from the midlatitudinal storm track (Houston and Hartley, 2003; Houston, 2006; Reyers et al., 2020). Under El Niño conditions, winter precipitation in the southern and coastal desert regions can be enhanced (Houston, 2006; Vargas et al., 2006; Valdés-Pineda et al., 2018). This can be attributed to a typical blocking situation in the midlatitudes during the developing phase of an El Niño, which forces the midlatitudinal low pressure systems northward (Montecinos and Aceituno, 2003; Vargas et al., 2006).

A second precipitation regime is established for the northeastern part of the Atacama. Predominant summer precipitation is controlled by the position of the Bolivian High (Vuille, 1999; Garreaud et al., 2009). A southward displacement of this upper tropospheric high pressure system results in enhanced moist easterlies over the Altiplano leading to convective storms (Vuille, 1999; Garreaud et al., 2003). Associated precipitation rates are strongly related to the topography with rapidly decreasing rainfall amounts towards lower heights at the western slopes of the Andes (Houston and Hartley, 2003). However, in case

of a southwestward displaced Bolivian High, these storms can be carried across the Andes and cause precipitation over the central valley farther west (Reyers et al., 2020). La Niña conditions favor these moist easterlies and, thus, can result in wet Altiplanic summers, whereas El Niño conditions result in dry summers (Vuille, 1999; Garreaud and Aceituno, 2001; Garreaud et al., 2003; Canedo-Rosso et al., 2019).

Besides cut-off lows, troughs and variations of the Bolivian High, another synoptic feature called Atmospheric Rivers (ARs) is at play. So far it has received little to no attention within the Atacama Desert. The term AR was first introduced by Zhu and Newell (1998) after Newell et al. (1992) identified these synoptic features as filamentary structures of enhanced water vapor in the troposphere. These structures are associated with major moisture and heat transport from the tropics to higher latitudes. Efforts to derive a more elaborative definition are described by Ralph et al. (2018).

For the west coast of South America, the impact of ARs has mainly been studied for central and southern Chile (Viale and Nuñez, 2011; Garreaud, 2013; Viale et al., 2018; Saavedra et al., 2020). Most frequent landfalls are observed between 38°S and 50°S (Viale et al., 2018). While north of 43°S, AR landfall frequency peaks during winter and spring, south of 43°S the peak seasons for ARs are summer and fall. For central Chile (32°S–37°S), about 45 %–60% of the total precipitation are related to land-falling ARs. Furthermore, warm winter storms in central Chile have been associated with AR (Garreaud, 2013).

For the Andean mountain range between 26.5°S and 36.5°S, about 50 % of the accumulated annual snow fall is related to ARs (Saavedra et al., 2020). Individual snowfall events are about 2.5 times higher under AR conditions compared to non-AR conditions. La Niña is associated with a reduction of AR frequency and a lower snow contribution, whereas weak to moderate El Niño conditions result in higher AR frequency and, consequently, more snowfall.

Intriguingly, some few AR landfalls have been detected even for northern Chile (Guan and Waliser, 2015, Fig. 8). While Reyers et al. (2020) pointed out that troughs and cut-off lows which bring precipitation to northern Chile are often associated with ARs, this study investigates the role of these synoptic features for the Atacama Desert in more detail. The study is guided by three overarching hypotheses:

1. A substantial amount of precipitation within the Atacama Desert is due to ARs.
2. The main origin of the AR related moisture is the Amazon Basin.
3. AR related moisture transport and precipitation formation mostly takes place in mid- and upper tropospheric layers and is decoupled from the MBL.

For the analysis, the AR catalog by Guan and Waliser (2015) providing spatial shapes and landfall times of identified ARs is utilized.

This catalog and other utilized data are described in more detail in Chapter 7.2. Thereafter, a case study is presented which exemplarily highlights the potential moisture source and the vertical structure of an AR (Chapter 7.3). Furthermore, a composite analysis of synoptic conditions and the vertical structure of all ARs identified for this region is conducted in Chapter 7.4, followed by a conclusion and outlook (Chapter 7.5).

7.2 DATA

7.2.1 *Atmospheric river catalog*

The AR catalog by Guan and Waliser (2015) is used to identify times with AR landfalls at the coast of the Atacama Desert. This catalog is maintained with frequent updates and improvements of the identification algorithm (https://ucla.box.com/ARcatalog; Guan et al., 2018). Here, Version 2 available for different reanalysis data sets is applied. By choosing the AR data set based on ERA-Interim (Dee et al., 2011b), consistency with an ERA-Interim-driven WRF simulation, which is applied in this study and introduced later in this Chapter, is ensured.

The AR identification algorithm considers a percentile-based Integrated Water Vapor Transport (IVT) threshold to identify objects of interest, which are further tested against orientation and geometry criteria (Guan and Waliser, 2015). Here, essential identification steps are briefly outlined:

1. The 85th percentile is calculated at each grid cell for each month using a 5-month rolling window centered on that month for the period between 1997–2014. For the Atacama Desert and the bordering southeast Pacific, this results in an IVT threshold between 50–100 $\mathrm{kg\,m^{-1}\,s^{-1}}$ for both summer and winter (c. f. Fig. 1 in Guan and Waliser, 2015). Over oceans, such low values are otherwise only found in polar regions.

2. The IVT direction of each grid cell within an identified object needs to be coherent with the overall mean IVT direction. Furthermore, the poleward oriented meridional IVT component needs to exceed 50 $\mathrm{kg\,m^{-1}\,s^{-1}}$.

3. A minimum length of 2000 km and a length-to-width ratio greater than 2 is required.

4. If 2. and 3. fail, then 1.–3. are repeated, whereby the IVT threshold in step 1. is increased by raising the percentile by 2.5 for each iteration (applied in Version 2; Guan et al., 2018).

The data set based on ERA-Interim is provided at a 6-hour temporal resolution on a 1.5° by 1.5° grid. For this study, identified ARs (Guan,

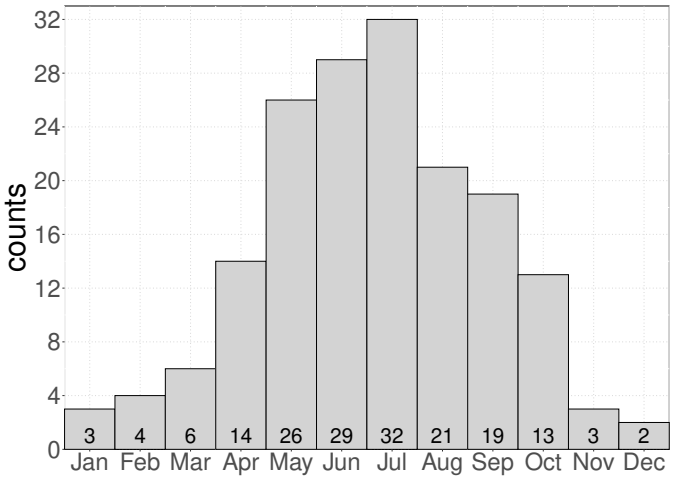

Figure 7.1: Total monthly counts of AR landfalls between 17°S and 30°S between 1979 and 2019 according to the AR catalog by Guan and Waliser (2015) version 2 (Guan et al., 2018). Landfall times have to be at least 48 h apart to count as separate ARs. The total of 172 AR landfalls correspond to a mean of 4.2 landfalls per year.

2020) are filtered for objects which made landfall at the coast of the Atacama Desert between 17°S and 31°S. If two consecutive landfalls are at least 48 hours apart, they are counted as two separate ARs. This way, a total of 172 AR landfalls are detected for the period between 1979 and 2019, which corresponds to about 4.2 landfalls per year on average. Monthly counts reveal a distinct seasonal cycle with the most active season between May and September and only a total of 12 landfalls between November and February within the 41-year period (Fig. 7.1).

7.2.2 *Regional climate model precipitation observations*

Precipitation data utilized in this study are taken from a simulation with the regional climate model WRF (Reyers, 2018). The data set comprises daily (separated at 0 Coordinated Universal Time (UTC)) accumulated precipitation on a 10 km resolution for a 36-year period (1982–2017). For the simulation, a double one-way nesting is applied using ERA-Interim as boundary conditions. The inner domain roughly spans the region between 16°S–26°S and 74°W–67°W. More details of the model setup are described in Reyers et al. (2020). Furthermore, hourly model output is used to illustrate the temporal evolution of relative humidity, cloud liquid water and wind within a case study.

A validation of the precipitation data set against gauge measurements provided by the Center for Climate and Resilience Research (CR2) (Center for Climate and Resilience Research, 2018) reveals suitable accuracy of the model (Reyers et al., 2020). Significant biases were only detected for steep slopes of the Andean Precordillera and can be

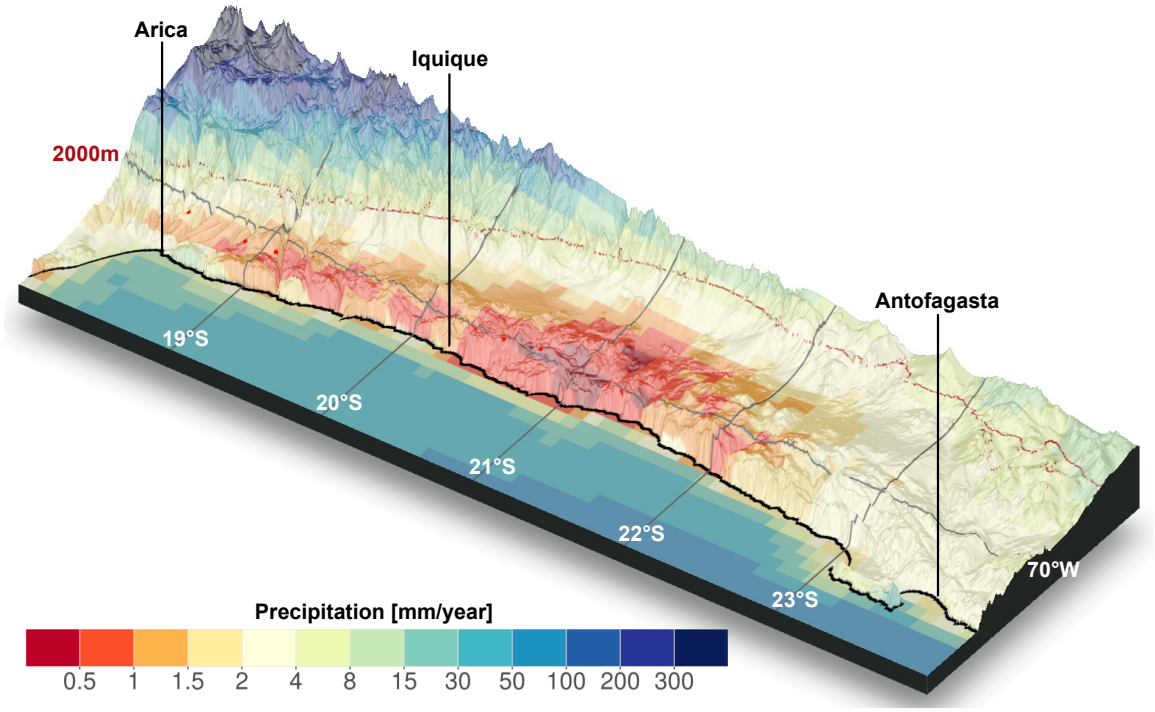

Figure 7.2: Precipitation climatology for the Atacama Desert derived from a WRF-simulation (1982–2017; Reyers et al., 2020). Topography is derived from Shuttle Radar Topographic Mission (SRTM) (Farr et al., 2007). The red lines denotes the 2000 m elevation line. The locations of three coastal cities (Arica, Iquique, Antofagasta) are annotated.

attributed to either an underestimation of summer convection in the convective scheme, the lack of moisture transport from the eastern side of the Altiplano or an underrepresentation of orographic gradients in the WRF model domain (Reyers et al., 2020).

A climatology of the simulated precipitation reveals a hyperarid core region roughly between 20°S and 22°S, including areas with annual precipitation rates below 1 mm encircled by a larger area with rates between 2 mm and 4 mm along with the coastal city Antofagasta (Fig. 7.2). Furthermore, a steep gradient is apparent with rapidly increasing mean annual precipitation rates towards the Altiplano in the northeast, which is consistent with results from previous studies (e. g. Houston and Hartley, 2003; Houston, 2006).

The gauge observations (Center for Climate and Resilience Research, 2018) which have been used by Reyers et al. (2020) are also included in this study to further assess the validity of the modeled precipitation. Furthermore, more recently updated gauge observations which are utilized for the presented case study (Chapter 7.3) are obtained from a dedicated website hosted by the CR2 (`http://explorador.cr2.cl`; Center for Climate and Resilience Research, 2019).

7.2.3 *Reanalysis and radiosounding data*

ECMWF's 5th generation reanalysis (ERA5) data (Hersbach et al., 2020) are utilized for a composite analysis. Hourly fields with a spatial resolution of 0.25° of longitude and latitude are obtained for mean sea level pressure, geopotential heights for 500 hPa and 700 hPa pressure levels, as well as specific humidity q, temperature and the eastward and northward wind components u and v for all available pressure levels and IWV. The total water vapor transport $\vec{Q}$ between two pressure levels p_{bot} and p_{top} is calculated manually using the horizontal wind vector $\vec{U} = (u, v)$ and the gravitational acceleration g of the Earth according to Equation 7.1:

$$\vec{Q} = \frac{1}{g} \int_{p_{\text{bot}}}^{p_{top}} q(p)\vec{U} \cdot \mathrm{d}p \tag{7.1}$$

ERA5 data are further applied to derive back trajectories of air parcels arriving at the Atacama Desert during an AR (case study). Therefore, the Hybrid Single Particle Lagrangian Integrated Trajectory Model (HYSPLIT) (`https://www.arl.noaa.gov/hysplit/`, National Oceanic and Atmospheric Administration (NOAA) Air Resources Laboratory; Stein et al., 2016) is run using hourly 3-D fields of meteorological data. Back trajectories are calculated for different target heights and locations starting 7 days prior to the target times, respectively.

For the presented case study, radiosonde data are obtained from the Integrated Global Radiosonde Archive (Integrated Global Radiosonde Archive (IGRA), Durre et al., 2016) for Antofagasta (23.5°S, 70.4°W). Data are available from daily launches at 12 UTC.

7.3 CASE STUDY

A heavy precipitation event occurred in June 2017 affecting a region within the hyperarid core of the Atacama Desert. Accumulated precipitation between 6-7 June 2017 exceeded 30 mm in a band of roughly 50 km width which stretched from the Mejillones Peninsula southeastward towards the Andean Precordillera in both the model and the gauge observations (Fig. 7.3b). Some regions received rain amounts even higher than 50 mm, which surpasses the ten-fold mean annual precipitation rate (c. f. Fig. 7.2). In the course of the event, various microbial species vanished due to the ponding of a lagoon near Yungay, located within this rain band (Azua-Bustos et al., 2018).

The event was associated with a trough visible in 500 hPa and 700 hPa which occurred anomalously far north and approached the Atacama Desert from the west (Fig. 7.3a). Ahead of the trough, a massive amount of water vapor was transported along the Peruvian coast towards the Atacama Desert. This transport has been identified

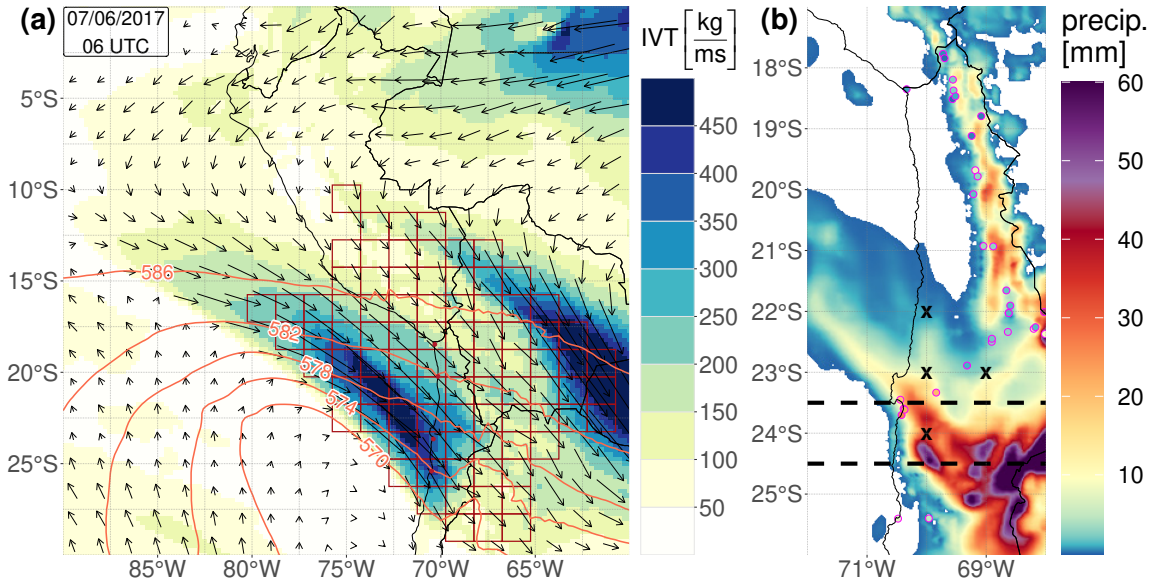

Figure 7.3: (a) Integrated Water Vapor Transport (IVT) integrated between 850 hPa and 200 hPa (color shading and arrows) and geopotential height in 500 hPa (light red contours). Dark red boxes denote the identified shape of the AR according to the applied AR catalog. All quantities are displayed for 7 June 2017 at 6 UTC. (b) WRF and CR2 station (magenta circles) 2-day precipitation sum for 6-7 June 2017. Crosses indicate target coordinates for back trajectories shown in Fig. 7.4. Dashed lines denote locations of vertical cross sections shown in Fig. 7.5.

as an AR according to the detection algorithm by Guan and Waliser (2015) (Section 7.2.1).

The disturbance only affected tropospheric levels above the MBL, whereas the prevailing southeast Pacific anticyclone remained intact at the surface (not shown). In fact, the IVT integrated between the surface and 850 hPa has a northward component (not shown). This indicates that the relevant moisture transport is limited to the free troposphere which is reflected in Fig. 7.3a by only considering pressure levels between 850 hPa and 200 hPa.

The anticyclonic nature of near surface levels is visible from the back trajectories which have been calculated for this event (Fig. 7.4). For the lowest target heights (500 m and 100 m), the 7-day path prior to the arrival at the central Atacama Desert denotes an east- to northward curving inflow over the Pacific. For higher target heights, the trajectories mainly originate over the Amazon Basin. They cross the Andes between the equator and 10°S, then turn southeastward over the tropical eastern Pacific and pass by the Peruvian coast before they reach the Atacama Desert. This means, the free tropospheric air which dominates the moisture transport towards the Atacama Desert (Fig. 7.3a) appears to be originating from the Amazon Basin. Such moisture would essentially be of Atlantic origin (e. g. Garreaud et al., 2009).

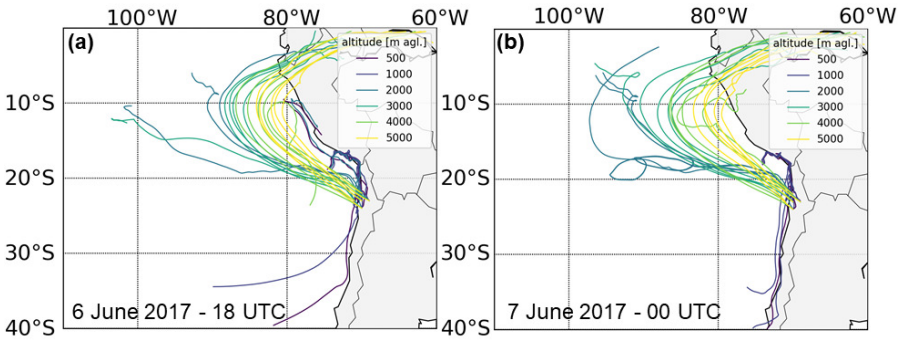

Figure 7.4: HYSPLIT back trajectories based on ERA5 for different target locations (crosses in Fig. 7.3) and target heights (see legend). Trajectories start 7 days prior to the target time on 6 June 2017 at 18 UTC (a) and on 7 June 2017 at 0 UTC (b). Figures were created by Leon Knarr and rearranged for this thesis with consent of Leon Knarr.

Variations of the upper level trajectories indicate that some air from the tropical Pacific is mixed in as well.

The fact that the moisture transport is mostly constrained to higher levels in the troposphere is further highlighted by longitude-height cross sections of relative humidity (Fig. 7.5) across the area of the main precipitation band (horizontal lines in Fig. 7.3b). The MBL and a humid layer at the verge of saturation above 5 km are separated by a layer of very low relative humidity (Fig. 7.5 a, c). This indicates that there is no significant exchange between the upper moist layer and the MBL. Furthermore, the strongest winds are collocated with the elevated moisture peak. Thus, the main moisture transport towards the Atacama Desert appears to be decoupled from the MBL and is realized at midtropospheric heights so that it can cross the coastal mountain range.

These mountains appear to introduce a vertical wind component which is transferred to higher altitudes. The resulting lifting of the moist air leads to cloud formation manifested in enhanced cloud liquid water (Fig. 7.5 a, c). At a later stage during the event, the dynamics of the accompanying trough result in a more complex circulation which appears to allow downward mixing of the humidity or cooling of the lower layer. This is indicated by higher relative humidity at lower altitudes over land and further cloud formation (Fig. 7.5 b, d).

The vertical structure of this AR is further investigated by analyzing radiosounding profiles which are available for Antofagasta with daily launches at 12 UTC (Fig. 7.6). Between 2–6 June 2017, the temperature inversion monotonically weakened (Fig. 7.6a), which can be attributed to the approach of the mid-tropospheric trough. At the same time, specific humidity increased slightly in the MBL and considerably between 450 hPa and 750 hPa, whereas a layer right above the temperature inversion at around 850 hPa remained dry (Fig. 7.6b). This is consistent with the dry layer revealed by the WRF simulation and confirms that

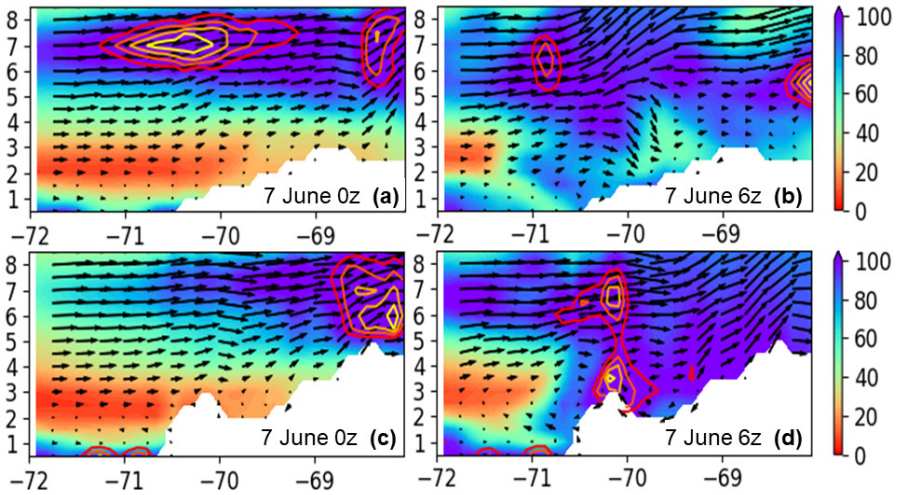

Figure 7.5: WRF vertical cross section (height in km, longitude in °E) of relative humidity in % (shading), u and w wind (arrows, w multiplied by a factor of 30) and cloud liquid water (contours, higher values for lighter color) along 23.5°S (a, b) and 24.5°S (c, d) corresponding to dashed lines in Fig. 7.3. Figures were created by Mark Reyers and rearranged for this thesis with consent of Mark Reyers.

the MBL is decoupled from the moisture accumulation within higher layers. The latter is explained by the increasing inland water vapor flux leading to the precipitation event (Fig. 7.6 c, d).

The precipitation in the central valley southeast of Antofagasta most likely occurred during the night from the 6–7 June 2017. This cannot be deduced directly from the CR2 and the WRF precipitation data sets because they only provide daily accumulated values. However, operating CRC 1211 station 13 (Cerros de Calate), which is located further north (21.4°S, 69,8°W), reveals the onset of the rainfall period at 1 UTC and the end at 6:30 UTC on 7 June 2017 (Hoffmeister, 2017a). Therefore, the radiosounding profile for 7 June 2017 at 12 UTC (8 a. m. local time) unveils the situation a few hours after the precipitation event. While there is still a strong southeastward water vapor flux peaking at 700 hPa, the temperature inversion dissipates. This could be due to evaporative cooling upon descending water droplets, which is consistent with the observed moistening of the dry layer around 850 hPa for this particular day (Fig. 7.6). An onshore homogenizing of the relative humidity is also indicated by the simulation (Fig. 7.5 b, d).

7.4 COMPOSITE ANALYSIS

In order to generalize the results from the case study, a composite analysis is carried out. For a 36-year period (1982-2017), 521 days which are related to a land-falling AR are determined, including the day before and after. For these days, the total AR-related sum of

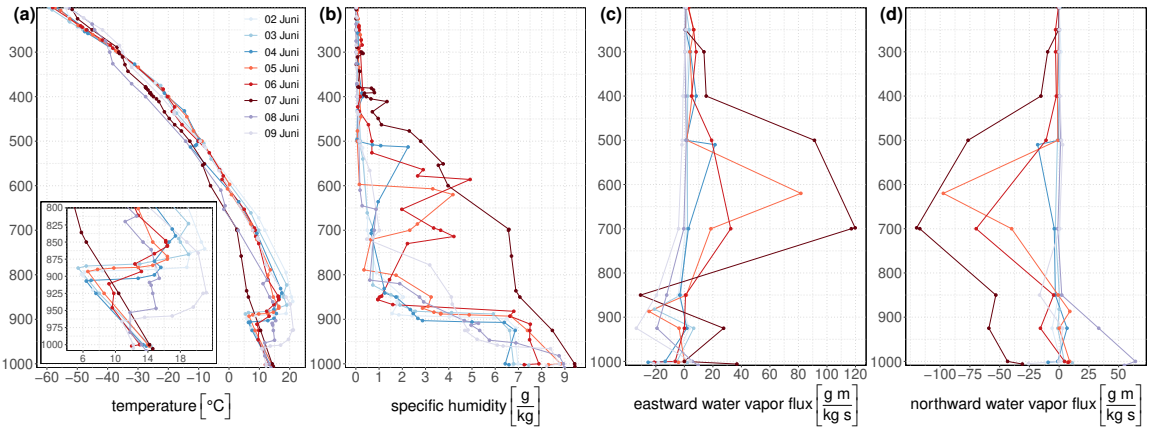

Figure 7.6: Vertical profiles derived from radiosoundings at Antofagasta of (a) temperature, (b) specific humidity, (c) eastward water vapor flux and (d) northward water vapor flux for the days between 2–9 June 2017 (see legend in a).

precipitation and composites are derived from the WRF precipitation data set and ERA5 data, respectively.

The AR-related fraction of total precipitation reveals an overall gradient between higher fractions in the southwest and lower fractions in the northeast (Fig. 7.7a). This pattern is consistent with the general understanding that AR frequency declines from a maximum in mid-latitudes towards the subtropics (Viale et al., 2018). In particular in the northeast, the pattern resembles the precipitation gradient which is dominated by summer time precipitation in association with moist easterlies (Vuille, 1999; Garreaud et al., 2003; Houston and Hartley, 2003). However, further south, maximum fractional precipitation appears for a distinct region between 23°S and 24°S, bounded by the city of Antofagasta in the west and stretching into the central valley to the east. For this region, more than 70 % and for some spots even more than 80 % of the total precipitation is associated with an AR making landfall. Generally, values exceed 40 % for most of the Atacama Desert south of 20°S. The magnitudes derived from the simulation are in reasonable agreement with magnitudes derived from the CR2 station data set. These findings demonstrate the importance of ARs for precipitation in the Atacama.

Precipitation events are rare in the Atacama Desert, which means that some locations do not receive any rain within several years (Cereceda et al., 2008a). In particular in the coastal desert, sparse precipitation can occur in form of drizzling fog. Such small precipitation amounts are very important for coastal loma vegetation (e. g. Pinto et al., 2006; Jaeschke et al., 2019) but are not sufficient to trigger mudslides or debris flow which causes landscape evolution detectable in paleoclimate archives (Vargas et al., 2006). Therefore, the fraction of days with precipitation is another quantity which characterizes the

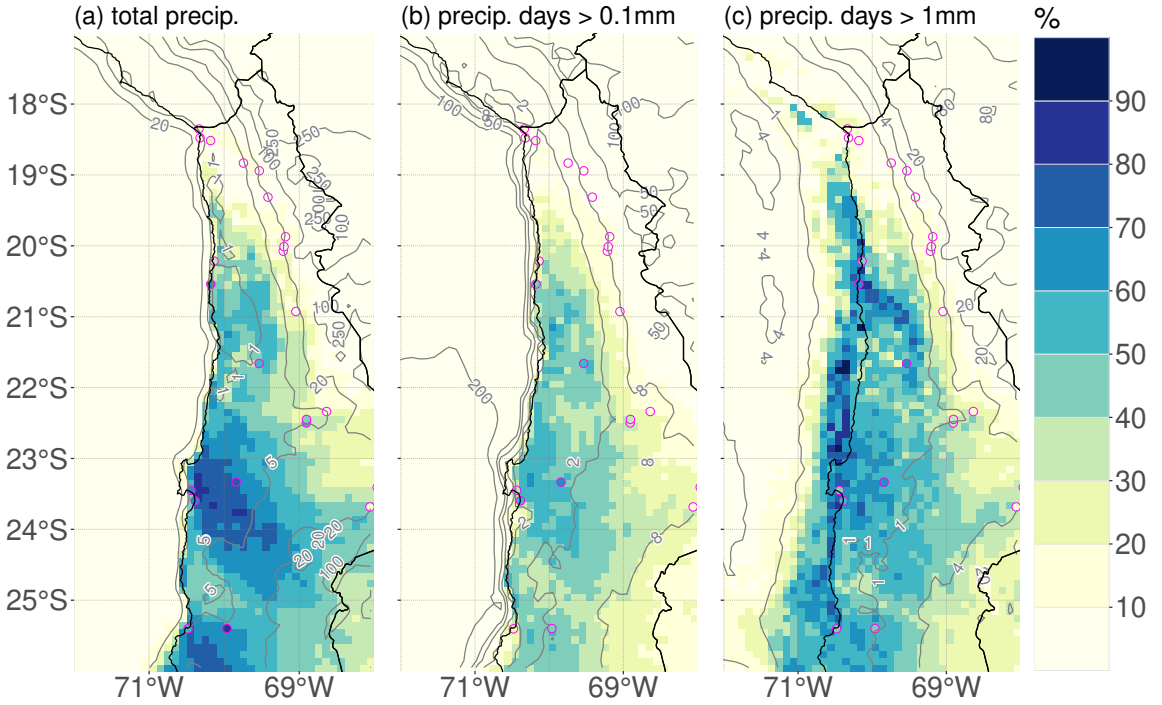

Figure 7.7: AR-related fraction of (a) total precipitation amount, (b) days with precipitation greater than 0.1 mm, (c) and days with precipitation greater than 1 mm based on a WRF simulation (Reyers et al., 2020). Contour lines represent the annual mean for each variable. Circles denote the respective analysis based on gauge measurement (source: CR2).

impact of ARs for the Atacama Desert. More than a third out of all days with detectable precipitation (greater than 0.1 mm) are associated with an AR landfall for most of the Atacama Desert south of 20°S (Fig. 7.7b). A regional maximum with the fraction exceeding 50 % is found to the east of the Mejillones Peninsula. If the threshold is increased to 1 mm, the AR-related fraction increases overall with values exceeding 50 % for most parts (Fig. 7.7c). In other words, days with higher precipitation amounts are more likely to be AR-related.

Days which are associated with ARs making landfall at the coast of the Atacama Desert typically feature an upper level trough ahead of the coast (Fig. 7.8). The southeast Pacific anticyclone is still visible according to the sea level pressure, which means that the prevailing southerlies within the MBL along the Chilean coast remain present in most AR cases. Regarding the synoptic overview, the previously described June 2017 event appears to be a representative case (c. f. Fig. 7.3a) for AR-related situations.

Additionally, AR-related days are associated with enhanced IWV across the southeast Pacific with a maximized positive anomaly a few hundred kilometers west of the coast of northern Chile. According to the composite mean geopotential height fields in 500 hPa and 700 hPa,

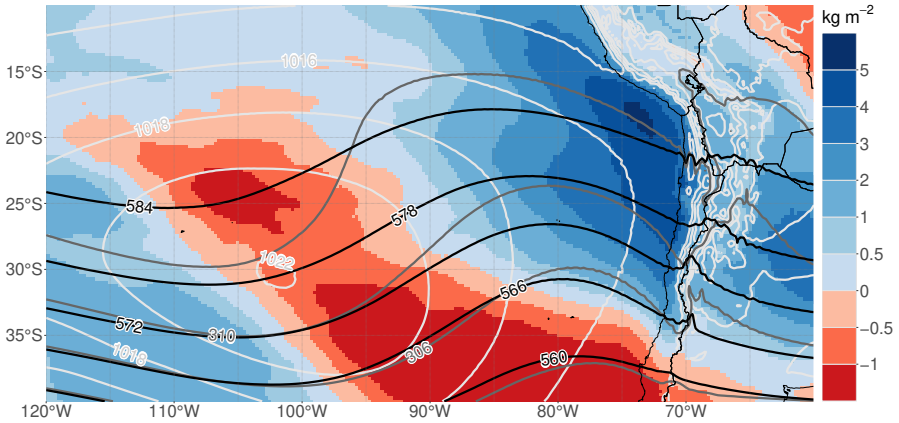

Figure 7.8: Composite IWV anomaly (shading), mean sea level pressure (white contours), geopotential height in 700 hPa (gray contours) and 500 hPa (black contours). IWV anomaly is derived as the mean deviation from the monthly climatology for each AR-related day. Data source: ERA5

a northwesterly airflow can potentially transport this moisture excess towards the Atacama Desert. To gain more insides regarding the moisture transport of regional ARs, the vertical structure of the atmosphere is studied for the AR-related days.

For this purpose, a specific vertical column is selected from the gridded ERA5 data set. The selection is carried out individually for each considered AR day according to the following procedure. Within a box bounded by 18°S–30°S and 73°W–71.5°W, the pixel with maximum absolute IVT is identified. Then, temperature, specific humidity and both horizontal components of the water vapor flux are stored for each available pressure level between 1000 hPa and 200 hPa. From this data set, the composite mean and the standard deviation are calculated for each pressure level (Fig. 7.9). Additionally, seasonal climatologies are derived for the entire bounding box for temperature and specific humidity (Fig. 7.9 a, b).

In the AR composite vertical profiles, winter and fall situations are overrepresented due to the seasonal distribution of AR landfalls (7.1). Even in this respect, the composite temperature is considerably colder within the free troposphere compared to the coldest season (JAS, Fig. 7.9a). The composite mean temperature profile within the boundary layer follows closely the seasonal average for spring and fall. Therefore, the inversion above the MBL is overall weakening during AR events, which can be attributed to reduced subsidence related to the presence of a midtropospheric trough.

Within the MBL, the composite mean specific humidity profile is again closely following the average spring and fall season profiles (Fig. 7.9b). However, above 950 hPa, the decrease with height appears at a lower rate. Even though winter time situations are overrepresented, the composite mean of the specific humidity exceeds the mean sum-

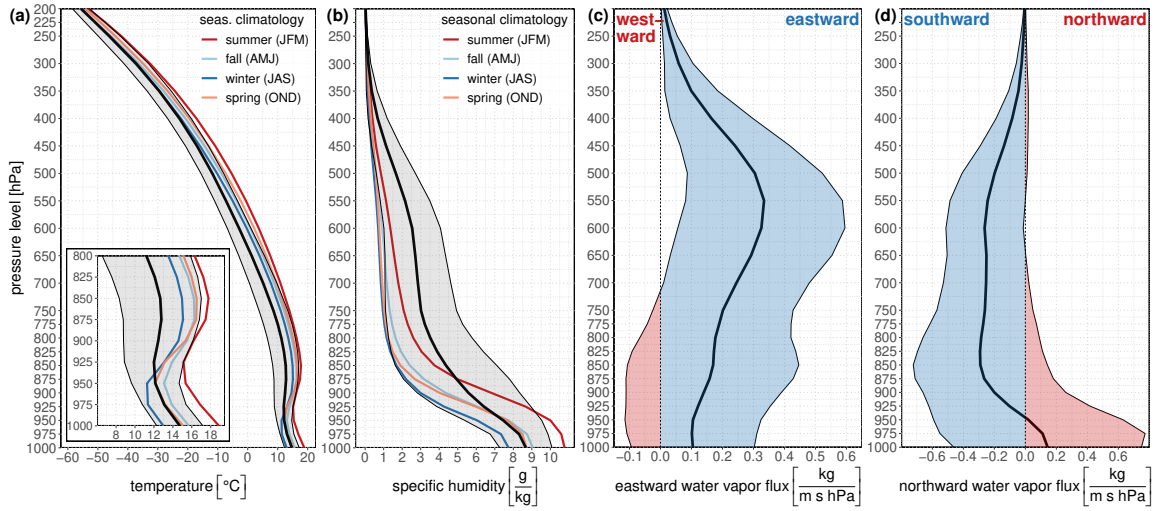

Figure 7.9: AR-composite vertical profiles of (a) temperature, (b) specific humidity, (c) eastward water vapor flux, (d) northward water vapor flux. AR-composite mean (black line) and standard deviation (shaded) are denoted for each pressure level. Colored lines in (a) and (b) denote respective seasonal climatologies. Data source: ERA5

mer time peak above 875 hPa. Overall, an AR-related enhancement of specific humidity is limited to the free troposphere.

Furthermore, the highest water vapor flux is observed between 500 hPa and 600 hPa on average (Fig. 7.9 c, d). At these pressure levels, a southeastward transport can be deduced. While the composite mean zonal component remains positive across all pressure levels, the meridional component changes the sign for the lowest pressure levels. This can be expected, as the southeast Pacific anticyclone remains intact near the surface, supporting the prevailing southerlies along the Chilean coast (Fig. 7.8). These results also confirm that the moisture transport mainly occurs at higher levels.

7.5 CONCLUSION

In this study, the role of ARs for the Atacama Desert is investigated. The results, which confirm all hypotheses raised in Chapter 7.1, are concluded and discussed in the following.

About 4 landfalls per year accumulate along the coast of northern Chile on average. Viale et al. (2018) report that in southern Chile more than 40 days per year are associated with ARs for coastal locations. This rate decreases towards the north to about 4 days per year at 27°S (Viale et al., 2018, their Fig. 7). While their detection algorithm differs slightly from the one applied in this study, this rate appears to be consistent with the number of landfalls per year found in this study.

Even though the AR landfall frequency is comparably low, these synoptic features can have substantial impact considering the overall low precipitation rates for this hyperarid environment. For example, an AR which made landfall on 7 June 2017 resulted in rainfall amounts which exceeded average decadal rates (Fig. 7.3b, Chapter 7.3). Events of such scale presumably dominate the traces which are identified in paleoclimate archives (Jordan et al., 2019; Ritter et al., 2019; Diederich et al., 2020).

For most regions of the Atacama Desert, more than 40 % of the total precipitation is related to AR landfalls. For a region east of Antofagasta, the fraction exceeds 70 %. As the case study for the June 2017 AR indicates, individual events can have a considerable impact on the spatial distribution of the fractional precipitation. Given that a single AR can be accounted for a decade's worth of rain, the 36-year long data record which is considered in this study appears quite short. Therefore, in future studies a longer data record would be desirable.

The investigation of the mean vertical profile of ARs related to the Atacama Desert reveals an elevated structure. Moisture transport maximizes typically between 500 hPa and 600 hPa and is decoupled from the MBL. While a mid-tropospheric trough drives a northwesterly flow towards the Atacama Desert, the southeast Pacific anticyclone continues to drive southerlies along the coast at the surface layer. This differs from the common understanding that these synoptic features are associated with a surface front (e. g. Gimeno et al., 2014; Payne et al., 2020). For higher latitudes, AR-related water vapor flux typically peaks below 750 hPa (Guan and Waliser, 2015). For example, for a location in southern Chile, the maximum water flux appears at 900 hPa (Guan and Waliser, 2015, their Fig. 8b). Therefore, the elevated AR structure identified here for the Atacama Desert can be considered a regional flavor.

This study reveals that the moisture is generally advected from the northwest, i. e. the tropical eastern Pacific. The case study explicitly demonstrates that a substantial part of the moisture is of Atlantic origin. The path leads from the Amazon Basin to the west, crosses the Andes between the equator and 10°S and turns southeastward over the tropical eastern Pacific to arrive at the Atacama Desert (Fig. 7.4). Similar pathways were identified for extreme precipitation events occurring during winter season in the southern, southeastern and northern Atacama (Reyers et al., 2020). An isotopic analysis of rainwater and stream flows associated with the Atacama March 2015 flood (Bozkurt et al., 2016; Wilcox et al., 2016; Barrett et al., 2016) indicated a tropical moisture source (Jordan et al., 2019). This is of interest because the March 2015 flood is also related to an AR (not shown) similar to the case study presented above (Fig. 7.3b), This further illustrates that the tropical origin might be a more common source region associated with AR-related moisture.

The hyperarid core constitutes a biotic barrier (Ruhm et al., 2020) but also a climatic divide (Houston, 2006). Precipitation in the northeast is usually related to continental interior origin, whereas precipitation in the southwest is usually of Pacific origin (Houston and Hartley, 2003; Houston, 2006; Jordan et al., 2019). This study reveals a third route which is further supported by Jordan et al. (2019) and Reyers et al. (2020). This additional mechanism has to be considered when climate archives are interpreted (Jordan et al., 2019).

ARs influence the hydroclimate of mid-latitude regions since the mid-Holocene (Skinner et al., 2020). An equatorward shift of AR trajectories between the mid-Holocene and the preindustrial period has been determined (Skinner et al., 2020). This shift is consistent with enhanced precipitation at the northern margin of the westerly wind belt during the late Holocene compared to the early Holocene (Lamy et al., 2010). For some mid-latitude regions, ARs account for nearly 100 % of the detected precipitation change between the mid-Holocene and the preindustrial period (Skinner et al., 2020).

To date, studies on paleoclimate reconstructions related to the Atacama Desert mostly consider SST, latitudinal shifts of winter midlatitude westerlies and cut-off lows in order to discuss implications from climate archives (Ritter et al., 2019; Bartz et al., 2020; Diederich et al., 2020). Wetter periods during the Pleistocene and Holocene are usually attributed to eastern or southwestern moisture sources (Jordan et al., 2019). An additional scenario which should be considered to interpret implications of climate archives is constituted by ARs as discussed in this study.

7.6 REMARKS

DATA AVAILABILITY The AR catalog created by Bin Guan is available via `https://ucla.box.com/ARcatalog`.

ACKNOWLEDGMENTS I would like to gratefully acknowledge the contributions from Leon Knarr, who carried out the HYSPLIT analysis and provided Fig. 7.4, and from Mark Reyers, who set up the WRF simulation and created the precipitation data set used in this study. Furthermore, Mark Reyers provided the longitude-height cross sections (Fig. 7.5) which are based on the WRF simulation and gathered the radiosounding data from IGRA. The AR data were provided by Bin Guan via `https://ucla.box.com/ARcatalog`. Development of the AR detection algorithm and databases was supported by NASA. I gratefully acknowledge financial support by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – project number 268236062 – SFB 1211.

Part III

DISCUSSION AND CONCLUSION

8 DISCUSSION

In the following, key findings of previous analyses are summarized and discussed. This includes additional scientific contributions, which built on these results and draw further conclusions.

8.1 CLOUD HEIGHTS

The maritime cloud heights ahead of the coast of the Atacama Desert control where the stratocumulus intersects with the coastal orography. Therefore, these cloud properties determine the location of fog-depending loma vegetation (Rundel et al., 1997; Pinto et al., 2006; Westbeld et al., 2009; González et al., 2011). Based on multiangular satellite data, a novel approach has been introduced which allows to retrieve both cloud top height and cloud base height with region-wide coverage (Böhm et al., 2019, Chapter 4). The so called MISR Cloud Base Height (MIBase) algorithm is applicable if some cloud gaps occur within a predefined detection area (e. g. a cell of a 0.25° longitude by latitude grid).

A thorough validation against more than 1500 ceilometer stations results in a Pearson correlation coefficient for the cloud base height of 0.66, a Root Mean Square Error (RMSE) of 385 m and a bias of −59 m, which proves the applicability of the algorithm to study seasonal changes and interannual variability. The spatial distribution of ground-based reference stations across the continental USA and the utilized sampling period covering a full seasonal cycle suggest that various cloud types were included in the validation.

An advantage of the MIBase algorithm is given by the employed Multi-angle Imaging SpectroRadiometer (MISR) cloud top height product, which uses a multiangular geometric approach for its cloud top height retrievals. Therefore, it is independent of any auxiliary data which would introduce additional uncertainties. A drawback is the MISR threshold height, which is applied to distinguish between cloud and surface retrievals. Therefore, the algorithm is blind to cloud heights below 560 m (ocean and flat terrain).

The application to the southeast Pacific region reveals the spatial distribution of the stratocumulus heights and their seasonal variations. Complex orography at scales below the resolution of the MIBase product complicate the interpretation of retrievals over the coastal mountain range of the Atacama Desert. By considering cloud heights over the nearby ocean and projecting them onto the coastal cliff, more reliable retrievals can be achieved (Fig. 8.1e). A slight North–South

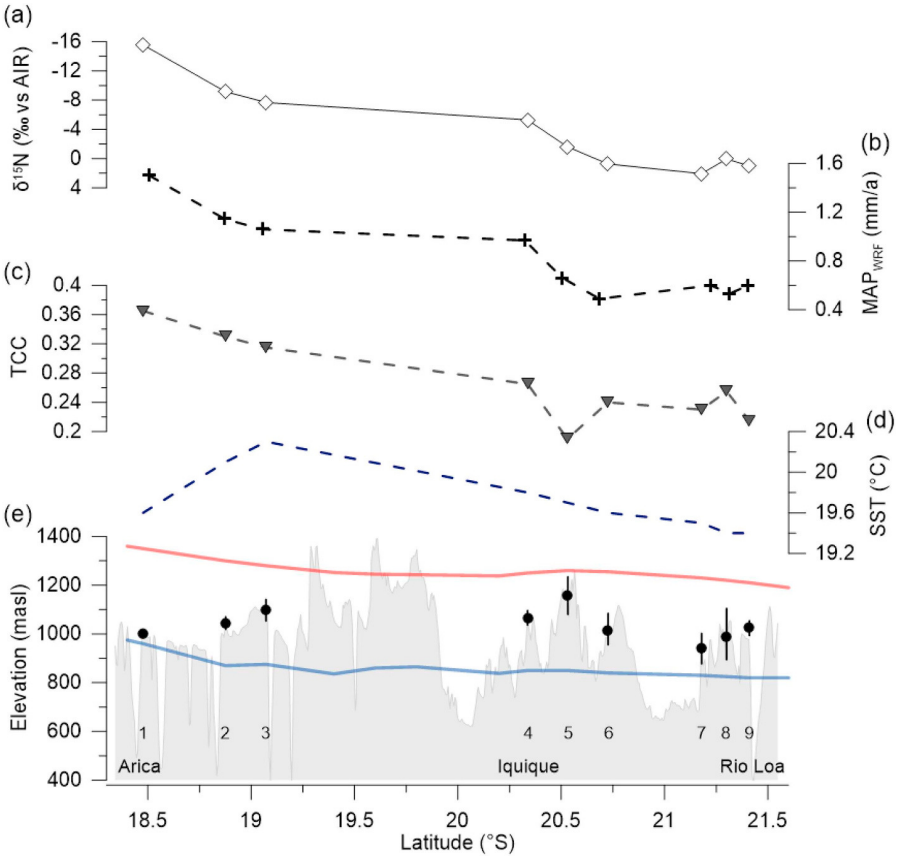

Figure 8.1: Latitudinal transect along the coast of northern Chile of (a) δ^{15}N (site-averaged) of *Tillandsia* populations, (b) mean annual precipitation (MAP) according to a WRF simulation (Reyers, 2018), (c) total cloud cover (TCC) from MODIS observations (Ackerman et al., 2015) for night time and morning overpasses, (d) mean annual SST (Locarnini et al., 2013), (e) median cloud top (red line) and base (blue line) height over the near coastal Pacific ocean according to MIBase (95th and 15th percentile, respectively), and positions of various *Tillandsia* populations (black circles with bars indicating respective height ranges) together with averaged topography between these populations (gray shading). This Figure is taken from (Jaeschke et al., 2019, their Fig. 9, used under the Creative Commons CC-BY-NC- ND license).

gradient is found which is consistent with long-term in-situ measurements at coastal stations (Muñoz et al., 2016). Furthermore, an East–West gradient is visible, which is consistent with a similar gradient determined from radiosonde profiles acquired during the Variability of the American Monsoon System (VAMOS) Ocean-Cloud-Atmosphere-Land Study – Regional Experiment (VOCALS-REx) campaign (Rahn and Garreaud, 2010, and Fig. A.1).

An application of the MIBase product in another study reveals further consistency of the detected cloud heights with the height range of fog-dependent coastal vegetation (Jaeschke et al., 2019, and Fig. 8.1e). This

study investigates the nitrogen isotopic composition (δ^{15}N) for various coastal *Tillandsia* populations and relates it to properties of the near coastal stratocumulus. On a global scale, foliar δ^{15}N can be used to approximate water supply along precipitation gradients according to a negative correlation across various terrestrial ecosystems (Handley et al., 1999; Amundson et al., 2003; Craine et al., 2009). However, these global observations do not include hyperarid regimes. For water-limited regions, such as the Namib and Atacama Desert, a reversal of this relationship has been noted (Soderberg, 2010; Díaz et al., 2016). Moreover, it has been indicated that for those regions δ^{15}N may be controlled by fog water supply (González et al., 2011; Latorre et al., 2011).

The $\delta^{15}N$ notation refers to the comparison of the ratio between nitrogen isotopes ^{15}N and ^{14}N for a sample (s) and a reference (r): $\delta^{15}\mathrm{N} = \left(\frac{\left(\frac{^{15}\mathrm{N}}{^{14}\mathrm{N}}\right)_\mathrm{s}}{\left(\frac{^{15}\mathrm{N}}{^{14}\mathrm{N}}\right)_\mathrm{r}} - 1 \right) \cdot 10^3\,‰$

Jaeschke et al. (2019) reveal a relationship between isotopic composition and altitude with decreasing δ^{15}N along increasing height transects within individual *Tillandsia* populations. This indicates a higher water availability at greater heights, which might reflect the higher cloud liquid water content near cloud top compared to cloud base within the stratocumulus (Wood, 2012). This explanation seems plausible because the explored heights range around the median cloud base and top heights determined using the MIBase.

Investigating δ^{15}N for various *Tillandsia* populations, a strong relationship between δ^{15}N and cloud properties including cloud heights and cloud cover was discovered (Fig. 8.2). This indicates that the isotopic composition is controlled by characteristics of the stratocumulus, which, in turn, determine fog frequency and intensity. Decreasing δ^{15}N reflects increasing fog water availability along a south to north gradient.

At study sites for which generally high δ^{15}N values have been found, collocated dead plants were available to investigate. A comparison shows that the dead plants have even more enriched δ^{15}N values, which indicates dehydration. This provides evidence that conditions at the coastal desert become increasingly water-limited, which is consistent with a coastal cooling (Falvey and Garreaud, 2009; Schulz et al., 2012) and decreasing cloud heights reported for recent decades (Schulz et al., 2012; Muñoz et al., 2016). Lower cloud heights are related to lower inland cloud cover in accordance with a lower chance for the clouds to overcome the coastal mountain barrier (Cereceda et al., 2008b) and, thus, limited fog water supply farther inland.

Jaeschke et al. (2019) also analyzed fossil *Tillandsia* excavated out of the characteristic *Tillandsia* dunes (Fig. 3.1). After establishing the relationship of δ^{15}N with various water supply related properties under current climatic conditions, fog water supply during past climatic episodes could be reconstructed. In this light, the study constitutes a first step towards a calibrated proxy applicable to reconstruct characteristics of the maritime stratocumulus.

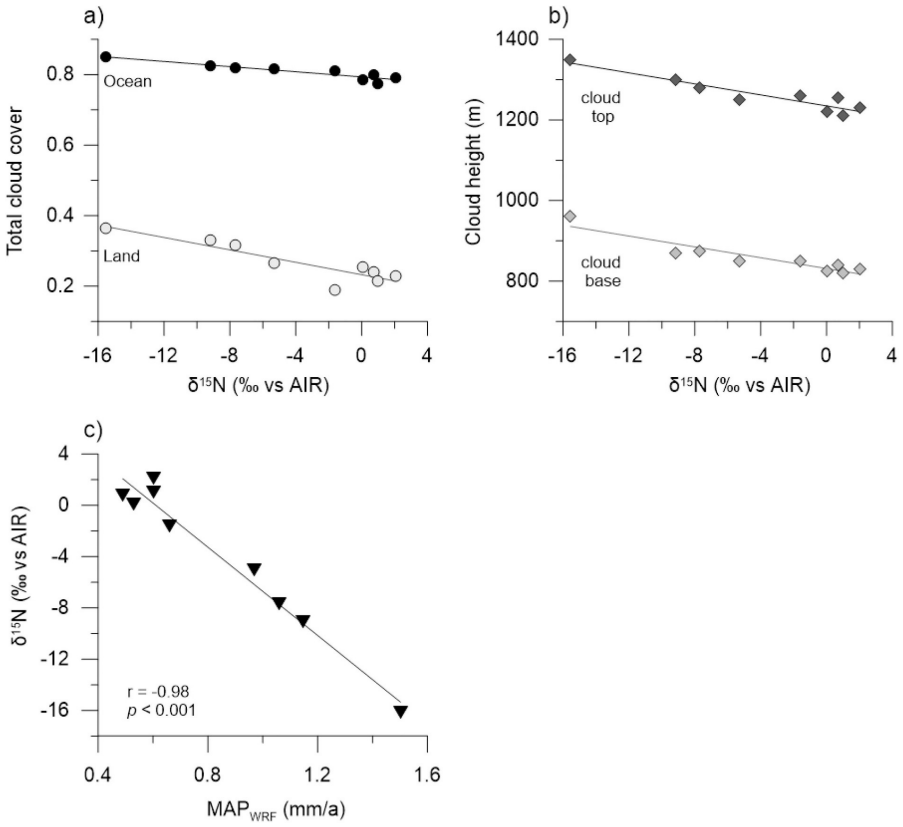

Figure 8.2: Scatter plots of site averaged δ^{15}N and (a) total night time and morning averaged cloud cover (MODIS; Ackerman et al., 2015), (b) cloud top and base heights (MIBase) (c) mean annual precipitation (WRF simulation; Reyers, 2018). Total cloud cover is determined at the Tillandsia locations (nearest pixel) indicated by "land" and over the near coastal ocean (using latitudinal extension to assign corresponding values) indicated by "ocean". Cloud heights represent median heights over ocean over a 17-year period. The Figure is taken from (Jaeschke et al., 2019, their Fig. 8, used under the Creative Commons CC-BY-NC- ND license).

In summary, the nitrogen isotopic composition of coastal *Tillandsia* depends on water supply by fog, which is directly related to properties of the near coastal maritime stratocumulus deck. This provides clear evidence that the southeast Pacific Ocean is the relevant source of moisture for these coastal ecosystems via its boundary layer capping stratocumulus. A method to monitor not only the stratocumulus top height but also the base height is provided by this thesis. In the future, MIBase can be applied for longer time periods to identify local trends of these properties on a region-wide scale and to relate them to changes in the vegetation state.

8.2 WATER VAPOR VARIABILITY

The confirmed dehydration and retreat of *Tillandsia* populations (Jaeschke et al., 2019) has been attributed to decreasing cloud heights (Schulz et al., 2012; Muñoz et al., 2016) and fog water availability. Lower cloud heights may be the result of near-surface cooling trends together with a warming of the free troposphere across the Atacama region within recent decades (Falvey and Garreaud, 2009). The resulting stronger inversion enhances stratocumulus formation and the colder MBL is associated with lower cloud heights (Wood, 2012). The near surface cooling trend is the result of a larger scale temperature trend across the Pacific Ocean, which resembles the PDO pattern (Falvey and Garreaud, 2009) and is consistent with a decreasing PDO index, i. e. a shift towards the negative phase (Newman et al., 2016). Therefore, the recent retreat of coastal vegetation (Rundel et al., 1997; Pinto et al., 2006; Schulz et al., 2012; Jaeschke et al., 2019) might be related to internal climate variability on interannual to decadal time scales associated with ENSO and the PDO.

To investigate the large-scale variability of water supply to the Atacama Desert, the long-term reanalysis ERA-20C has been utilized (Böhm et al., 2020a, Chapter 5). It provides a consistent record of the state of the atmosphere for a century-long period at the coast of a coarse resolution (130 km). While precipitation and clouds are subject to complex small-scale processes, which have to be parameterized at such coarse resolution, IWV proved to be represented with suitable accuracy. In the process of a thorough validation, no systematic biases in space and in time which would hamper a consistent analysis could be found. Instead, good agreement of IWV anomalies with state-of-the-art satellite-based retrievals prove the suitability for future studies of the long-term IWV variability across the Atacama Desert.

Over the course of the 20th century, an overall IWV variability around 30 % on a year-to-year basis as well as some varying trends were detected. On an interdecadal time scale, the PDO shows the highest correlation with IWV compared to ENSO and the local SST for the southern Atacama, whereas in the northern part the influence of the local SST is the most dominant factor (Fig. 5.12).

On shorter time scales, overall lower correlations between IWV and large-scale indices are found, which indicates the importance of synoptic noise on shorter time scales. This implies that climate modes do not directly cause IWV anomalies but favor certain synoptic conditions, as proposed by Houston (2006).

While synoptic noise affects correlations on shorter time scales, e. g. using annual averages, opposite seasonal influences of the large-scale modes further degrade the apparent correlations. An analysis of distinct seasonal ENSO impact reveals that El Niño summers are typically dry, whereas La Niña features both dry and wet summer

periods. This pattern is seen for all regions but most pronounced in the northern Atacama. An opposite ENSO relation is found for the winter season, with enhanced IWV for El Niño and drier conditions for La Niña across all regions including near-coastal ocean and inland.

For the summer season, similar seasonal ENSO modulations have been reported for the Altiplano regarding precipitation (Vuille, 1999; Garreaud and Aceituno, 2001; Garreaud et al., 2003; Houston, 2006). The consistency between the ENSO impact on IWV addressed in this thesis and on precipitation according to previous studies suggests that the underlying mechanisms are the same and that IWV can potentially serve as a proxy variable for precipitation. Further studies on the relationship between IWV and precipitation would be beneficial to confirm the latter point. Furthermore, this thesis revealed that the ENSO influence is not restricted to the Altiplano in the northeast but affects also the southern parts and the near-coastal ocean regions.

Considering annual averages, the correlations between IWV and large-scale indices (ENSO and PDO) are overall positive, which indicates that the positive relationship identified for the winter season dominates the integrated annual signal. For the winter season, wetter conditions associated with El Niño have been attributed to a northward shift of midlatitude disturbances (cut-off lows and troughs) (Vargas et al., 2006). These disturbances are generally related to moisture supply of Pacific origin (Vuille and Ammann, 1997; Houston and Hartley, 2003). The winter moisture excess during the positive ENSO phase overcompensates the enhanced moisture signal. As this signal is related to moist easterlies for the summer season, the dominating moisture source appears to be the Pacific Ocean. However, a new analysis reveals a more complex pathway from the Amazon Basin across the Pacific towards the Atacama Desert for extreme precipitation events occurring during winter (Reyers et al., 2020), indicating Atlantic moisture origin even for westerly dominated weather events.

An exception is the northern Atacama region, for which a weak negative correlation between IWV and the Niño 3,4 is found on an interannual basis. Therefore, the moisture from the interior continent might play a more dominant role for the northern Atacama Desert.

Regarding the PDO, the seasonal distinction did not reveal a clear shift of the IWV distribution. This could be explained by a possible interference with the ENSO signal. Previous studies suggest an amplification or damping of the ENSO signal according to positive or negative interference, respectively (Kayano and Andreoli, 2007; Valdés-Pineda et al., 2018). To further investigate the impact of the PDO, the seasonally varying ENSO impact could be distinguished by different phases of the PDO in a future study.

Furthermore, the validation of IWV variability in ERA-20C is essential for putting IWV trends which are determined from shorter data records into perspective. In another study, water vapor trends at the

Paranal Observatory are investigated in the context of climate change (Cantalloube et al., 2020). As the Atacama Desert is presumably the driest desert on Earth, it is home to various astronomical observatories. Moist episodes with enhanced water vapor degrade the quality of astronomical observations. The study shows a recent increase of favorable conditions with extremely low IWV. This recent trend is consistent with the IWV evolution according to ERA-20C. However, the reanalysis reveals that the decrease of IWV started after the 1997/98 El Niño and was preceded by a period of increasing IWV starting in the 1970s. These varying trends may be related to the PDO, which shows a similar evolution over this time period.

Cantalloube et al. (2020) also point out that ARs might frequently impact the quality of astronomical observations carried out at various sites within the Atacama Desert. The study on the role of ARs for the Atacama Desert (Chapter 7) is concluded in Chpater 8.4.

8.3 FOG

In the study by Böhm et al. (2020) (Chapter 6), a satellite-based fog retrieval method utilizing brightness temperatures from MODIS emissive bands has been developed. As prerequisite, a ground-based reference data set was created. It is mainly based on in-situ measurements of leaf wetness sensors but includes also standard meteorological data, such as relative humidity and temperature. These ground-based data were acquired from the newly installed CRC 1211 network of climate stations. By applying this novel ground-based fog reference to develop the retrieval algorithm, a distinction between fog conditions or dry conditions can be made at the ground level. Most satellite-based methods aim to only identify fog and low clouds as one common category (e. g. Cermak, 2012; Egli et al., 2017; Andersen and Cermak, 2018). The identification of actual fog is essential to relate the spatio-temporal variability of fog to biological and geological processes at the ground.

The newly developed algorithm has been applied to derive mean seasonal fog distributions for a 3-year period (2017–2019) as a preliminary step towards a long-term climatology. For the winter season (JAS), the pattern is consistent with an auxiliary approach developed in parallel from the MODIS cloud top height product and with previous studies regarding the coastal maximum. Furthermore, previously identified fog corridors, which allow inland penetration of the marine stratocumulus, a clearly visible for a region between 20°S and 22°S (Farías et al., 2005, c. f. their Fig. 13a). In fact, the higher resolution of this new retrieval introduced in this thesis allows a better estimation of the extents of fog corridors (Figure 6.14).

Mostly connected to these corridors, some areas within the central valley are identified with fog frequencies exceeding 30 % or for some few locations even 70 % during the winter months. Not all of these

inland locations can be traced back to the Pacific Ocean via an apparent band of enhanced fog frequency. Furthermore, most of these inland fog spots appear to be collocated with or in close proximity to salt flats. They also lie predominantly within the region which exhibits the most hyperarid conditions according to annual precipitation rates from the WRF simulation (Fig. 7.2 Reyers, 2018; Reyers et al., 2020) and according to an analysis of the chemical compounds of the desert soil (Voigt et al., 2020).

A part of this region within the central valley is the Salar de Llamara, where fog has been studied previously in the framework of a dedicated field campaign (Cereceda et al., 2002). During this campaign, frequent observations were conducted within a week in August 1998. The authors postulate different scenarios for the origin of the observed fog. One possibility is the advection of coastal fog which survived longer in the central valley compared to the fog within the corridor. The advected stratocumulus may also have dissipated, leaving humid air in the central valley which condensed during nocturnal cooling. A third possible moisture source could be underground water which evaporated during the day over the salt flats.

The fog detection algorithm introduced in this thesis allows to address this question of fog moisture origin. In a subsequent study, fog composites could be created and related to synoptic conditions to identify potential pathways of moisture. Furthermore, the influence of large-scale patterns on the fog occurrence at different regions can be assessed. At a coastal research site (Alto Patache), enhanced fog water yields are observed for the summer season when the SST anomaly within the Niño 3,4 region exceeds 1 K (Río et al., 2018). The question whether this relationship holds on a region-wide scale and in particular for the inland fog has not been investigated yet. In summary, the newly developed fog retrieval method introduced in this thesis enables to further test the hypotheses raised by this thesis.

8.4 ATMOSPHERIC RIVERS

In addition to cloud heights, water vapor and fog which are associated with moisture supply to the Atacama Desert, this thesis investigated the role of ARs for this region (Chapter 7). It was demonstrated that these features account for a substantial fraction of the total precipitation, exceeding 40 % for most regions and even 70 % for a region of Antofagasta. While an overall North–South gradient can be identified with greater impact in the south, local maxima of the spatial distribution of AR fractional precipitation might be traced back to individual record breaking events. Even though a long-term data set (36 years) was applied, the frequency and quantity of precipitation is so low in the Atacama,that one event may contribute rainfall on the order of the ten-fold annual rate.

The analyzed ARs typically advance towards the Atacama Desert at the foreside of a midtropospheric trough over the southeast Pacific. They can also be associated with a cut-off low (Reyers et al., 2020). The southeast Pacific anticyclone remains intact at the near-surface layer so that the moisture transport towards the Atacama is mostly confined to the free troposphere above the MBL. This result contrasts the lower height of maximum water vapor flux, which is typically observed for ARs at higher latitudes (Guan and Waliser, 2015). Furthermore, the temperature inversion weakens, which can be attributed to the presence of the trough but usually persists even at AR landfall time. Therefore, it is concluded that the elevated moisture transport is decoupled from the MBL with very limited exchange between these layers.

Due to this limited exchange between the local boundary layer and the elevated moisture transport band, the moisture is most likely of remote origin. Back trajectories calculated for one example case mark the tropical Pacific and predominantly the Amazon Basin as source regions. Reyers et al. (2020) identified similar origins of air parcels for selected extreme precipitation events which occurred during the winter season. These events were mostly associated with cut-off lows and ARs or a combination of both.

According to a clustering which identifies similar temporal characteristics in the seasonal cycle of precipitation, Reyers et al. (2020) divided the Atacama in a northeastern (summer rainfall peak), northern (winter and summer peaks), southern (winter peak) and southeastern part (winter peak pronounced in June). For each domain, the top ten most extreme precipitation events between 1982 and 2017 were selected for a composite analysis. For the southern Atacama, eight of them were associated with an AR, whereas for the northern Atacama, cut-of lows dominated with only two ARs. This indicates a stronger role of the precipitation in the southern part, which is consistent with the detected North–South gradient of AR fractional precipitation (Fig. 7.7).

Furthermore, the composite analysis by Reyers et al. (2020) revealed that the precipitation events in the northern Atacama are not associated with enhanced SST over the eastern tropical or southeastern Pacific. On the contrary, for the southern region, a positive composite anomaly of the eastern tropical Pacific SST is revealed. This indicates that the occurrence of ARs, which appear more predominant in the southern Atacama compared to the north, might be sensitive to Pacific SSTs, which, in turn, suggests a potential influence by ENSO. This points to the conclusion that precipitation in the southern Atacama is driven mainly by large-scale patterns, whereas the northern part is predominantly affected by independent synoptic variability. However, further studies are needed to determine the sensitivity of ARs to Pacific SSTs and ultimately to large-scale indices, such as ENSO and PDO. To study

relationships between frequency and intensity of regional ARs and large-scale patterns, a longer time series would be required. ERA-20C constitutes a suitable record to study water vapor related quantities (Chapter 5, Böhm et al., 2020a).

Establishing the role of large-scale climate modes would also be necessary to separate potential influence of global climate change. The impact of climate change on ARs is barely understood for South America (Payne et al., 2020). The multidisciplinary approach carried out by the CRC1211, which includes climate and paleoclimate research together with investigations of natural climate archives manifested in desert soils, alluvial fans or fossil plants, can help unravel some of the uncertainties.

9 CONCLUSION AND OUTLOOK

This thesis investigated key components of the atmospheric water cycle which are related to water supply to the Atacama Desert. Thereby, it provides a comprehensive meteorological perspective on this unique ecosystem. Two overarching hypotheses guided the conducted research. These hypotheses led to two key questions: Where does the atmospheric moisture come from and what are the dominant drivers controlling the water input? To address these key questions, new retrieval methods were developed to expand the observational background.

Essentially, the moisture source depends on the form of supply, which may be through clouds, water vapor, fog or precipitation, as well as on the season and region within the Atacama Desert. By means of a newly developed algorithm detecting cloud base heights using satellite-based observations (Böhm et al., 2019), the impact of the maritime stratocumulus on coastal fog-dependent ecosystems could be studied in detail. Together with further atmospheric characteristics, a clear link between the state of the coastal *Tillandsia* populations and the marine stratocumulus could be established (Jaeschke et al., 2019). This manifests the strong influence of the southeast Pacific on the coastal desert and the dependence on the near-coastal cloud properties.

The coastal cloud heights further determine where this maritime moisture can find its way inland through corridors within the coastal orography. These potential pathways may control the chemical compounds within the desert soil (Voigt et al., 2020). Therefore, the introduced MIBase algorithm could be exploited further in a future study to investigate the cloud height response to different climate modes and to determine long-term trends. Such a study might also allow to answer the second key question of this thesis regarding the influence of large-scale drivers on the stratocumulus. However, given that the required satellite observational records only began in the year 2000, more research should be conducted in the future to provide a definite answer.

To address the influence of large-scale drivers, such as ENSO and PDO, ERA-20C was utilized (Böhm et al., 2020a). This reanalysis provides a consistent century-long time series of the atmospheric state by only assimilating surface pressure and surface wind observations to constrain the simulation. While cloud liquid water and precipitation are expected to be challenging variables because they involve small-

scale processes far below the resolution of ERA-20C, IWV representation has proven to be suitable for further investigations.

For the northern Atacama, which includes the Altiplano at this coarse resolution, seasonal relationships of IWV and ENSO were identified. These relationships are consistent with the findings reported in previous studies regarding precipitation. For the summer season, El Niño is related to enhanced dryness, whereas La Niña allows wetter episodes. The similarity to the ENSO influence on precipitation which has been explained for the Altiplano strongly suggests that a similar mechanism is at play and that IWV has potential to serve as proxy for precipitation. This mechanism involves favored moist easterlies under La Niña conditions. In this thesis, it is shown that IWV enhancement extends beyond the Altiplano towards the Pacific Ocean and towards the southern Atacama.

While the previously described mechanisms imply a continental interior moisture source for the summer season, the opposite is found for the winter season. The latter is characterized by enhanced moisture for El Niño conditions and reduced moisture for La Niña conditions. This relationship is most pronounced over the near-coastal ocean, which indicates an apparent coupling of these regions to the Pacific Ocean. Together with previous studies which report enhanced precipitation in relation to a northward shift of the midlatitudinal storm track region under El Niño conditions, the described ENSO relationship strongly suggests the dominance of the Pacific Ocean as the main source of IWV for austral winter. However, in particular during the cold season, the study of a previously unaddressed synoptic feature, i. e. ARs, shows that not only Pacific moisture may enter the Atacama Desert from the west.

To further investigate whether the mechanisms controlling the IWV are also responsible for precipitation, ERA-20C could be further exploited to study the atmospheric flow in relation to water vapor availability and large-scale climate modes. The thorough validation of the IWV already laid the groundwork for such additional investigations. A following study already made use of the validated IWV record to put water vapor trends available from finer resolved short-term observation into perspective (Cantalloube et al., 2020).

While El Niño conditions prevent enhanced IWV, in particular for the northern offshore region ahead of the Atacama coast, increased amounts of fog water are collected at the coastal research site at Alto Patache (Río et al., 2018). To study the water supply by fog, a novel fog detection algorithm has been developed within this thesis (Böhm et al., 2020b). The spatial distribution of mean fog frequencies for a preliminary 3-year period demonstrates the potential of this new method. The fog frequency maximum at the coast which is related to the maritime stratocumulus can be observed region-wide along

with few corridors which allow inland penetration through the coastal mountain barrier.

Furthermore, isolated locations with enhanced fog frequencies within the central valley could be identified. These are mostly in the vicinity of salt flats and within the hyperarid core of the Atacama Desert. It remains unclear which mechanisms control fog formation in these regions. The presented fog detection algorithm provides a tool to relate these fog events to atmospheric circulation and to identify potential moisture sources. Additionally, a future study can investigate potential influences of the regional SST variability and ENSO by exploiting the full MODIS record starting in the year 2000 (Terra satellite).

Regarding precipitation, SST sensitivity appears to be higher in the southern Atacama Desert compared to the northern part (Reyers et al., 2020). This collocates with the region where ARs have the highest impact regarding fractional precipitation (over 50 %), which is revealed in the study presented in Chapter 7. A substantial fraction of precipitation is related to ARs for most regions within the Atacama Desert (over 30 %). The main moisture transport related to these atmospheric features occurs in the free troposphere decoupled from the MBL. This elevated structure suggest that remote sources are at play.

In a case study within this thesis the Amazon Basin was identified as source region by means of back trajectories. A previous study of the major March 2015 Atacama flood reported that the related precipitation was of tropical origin according to an analysis of the isotopic composition (Jordan et al., 2019). While it was not noted previously, this event was also associated with an AR according to the AR catalog applied in this thesis (Guan and Waliser, 2015). Furthermore, Reyers et al. (2020) investigated back trajectories for extreme winter precipitation events within the northern and southern Atacama Desert. These events reveal origins within the tropical eastern Pacific and the Amazon Basin. Therefore, aside from subtropical Pacific moisture associated with winter storm westerlies or continental interior moisture advected by summer time easterlies, a third scenario of moisture supply to the Atacama could be revealed. Starting from the Amazon Basin, the moisture crosses the Andes slightly south of the equator and turns southeast over the eastern tropical Pacific to approach the Atacama Desert. As pointed out by Jordan et al. (2019), this new scenario should be considered when climate archives from this study region are interpreted.

The results and newly developed methods presented by this thesis may foster further collaborative research to investigate mutual relationships between biological, geological and atmospheric processes. The interdisciplinary approach which has been initiated by the CRC 1211 opens great new opportunities with benefits for various scientific communities.

Part IV

APPENDIX

APPENDIX

A.1 CLOUD HEIGHTS

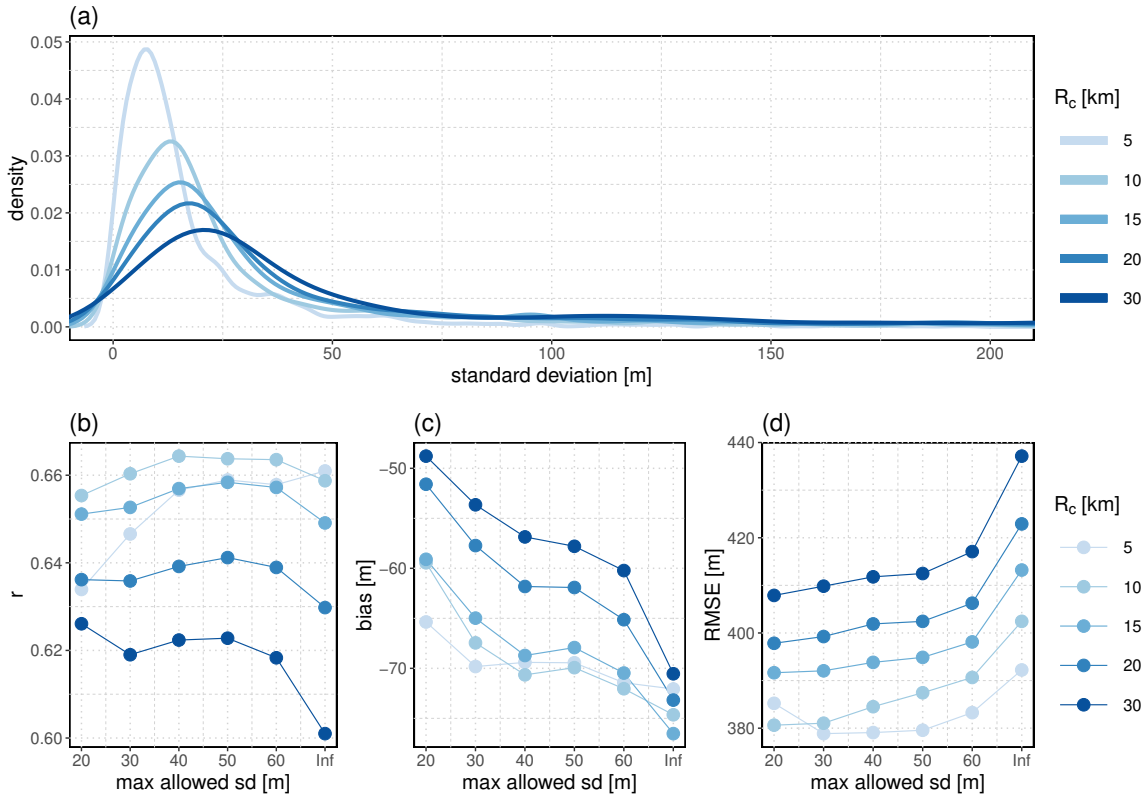

Figure A.1: a) Normalized frequency of occurrence of the occurring standard deviations (sd) of the average scene elevation (ASE) for various radii. The ASE is provided by the MISR ancillary product. Recalculated are the correlation coefficient *r*, bias, RMSE using only ceilometer stations below a threshold sd as denoted on the abscissas.

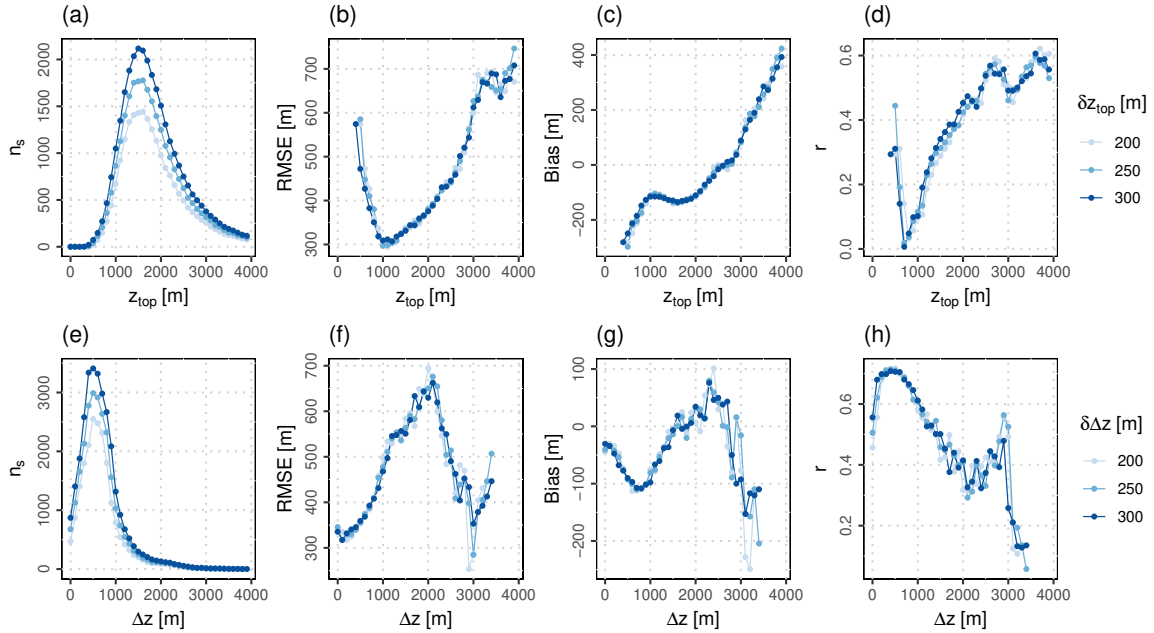

Figure A.2: Number of samples n_s, RMSE, bias, correlation coefficient r for the comparison of MIBase and ceilometer retrievals in dependence on z_{top} (top row) and cloud vertical extent Δz (bottom row). Each data point is calculated for a sub sample which includes only $z_{top} \pm \delta z_{top}$ or $\Delta z \pm \delta \Delta z$, respectively. The various widths of the considered z_{top} or Δz windows are indicated by the blue shading.

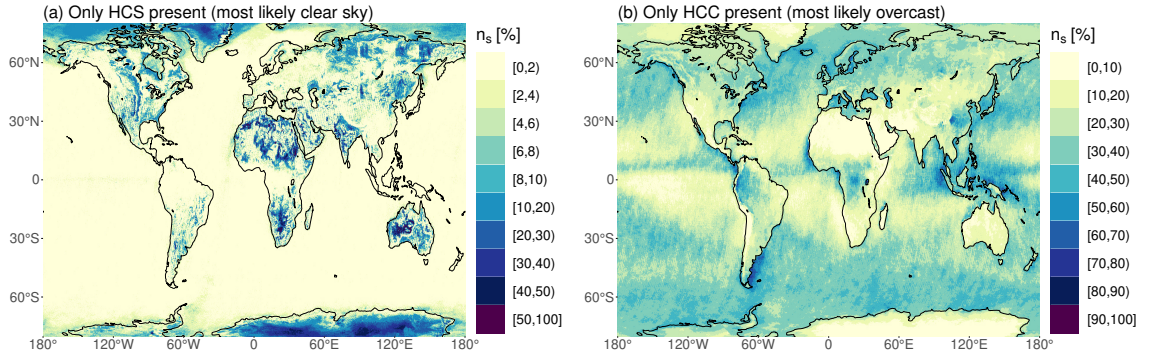

Figure A.3: Relative occurrences of different stereo-derived cloud mask (SDCM) configurations within the three-year period (2007–2009). The reference sample size n_s includes all overpasses per grid cell which contain valid z retrievals and corresponds to 100 %. These configurations are: (a) Only high confidence surface (HCS). These cases should be mainly clear sky cases. (b) Only high confidence cloud (HCC). These cases should be mainly cloud scenes with apparent overcast.

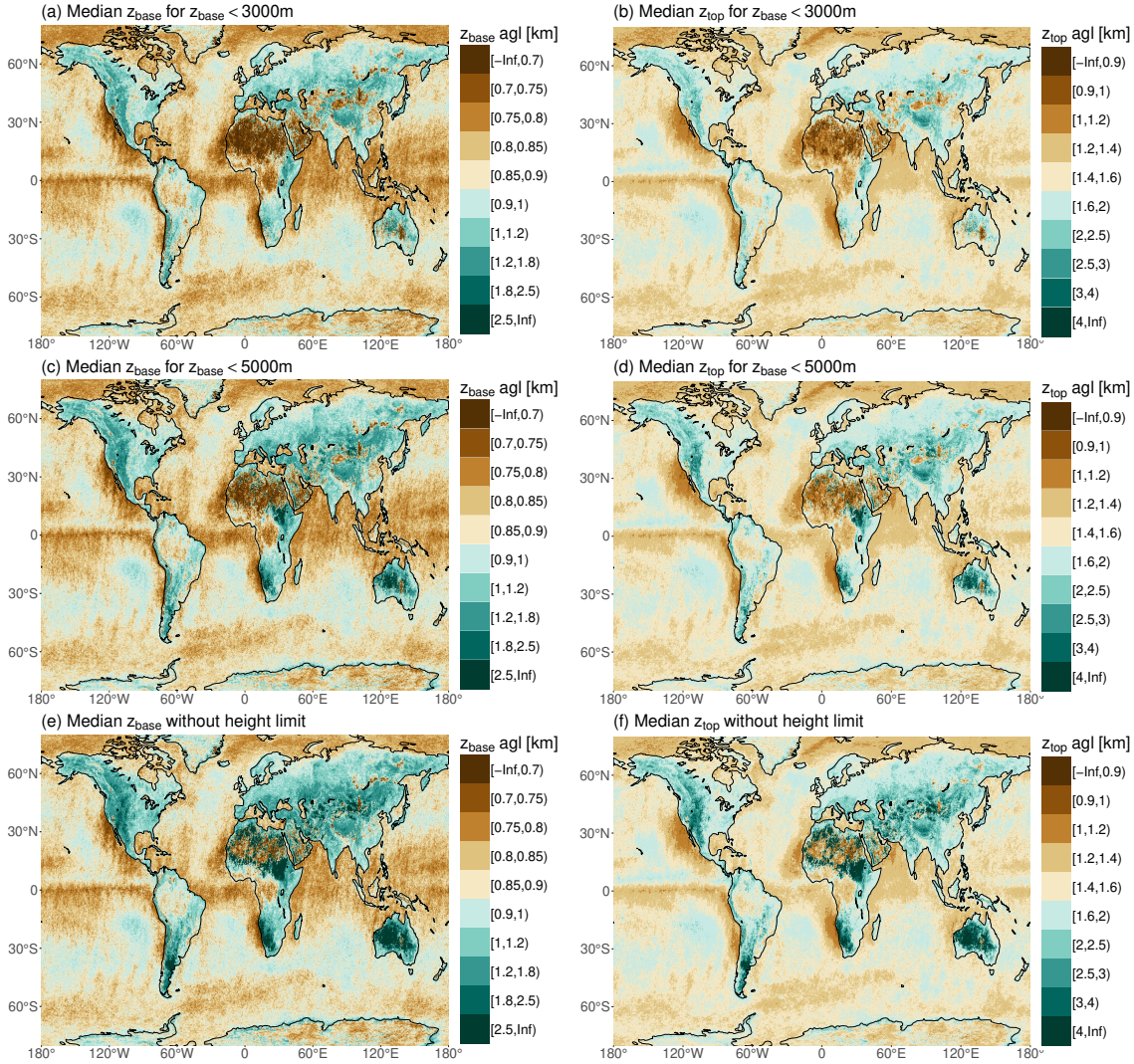

Figure A.4: Global distribution of median cloud heights for a 3-year period (2007–2009). Shown are z_{base} (left) and z_{top} (right) on a $0.25° \times 0.25°$ latitude–longitude grid. z_{base} and z_{top} are above ground level (agl). z_{base} and z_{top} retrievals are only included in the statistic if z_{base} is below 3000 m (a, b), 5000 m (c, d). For (e) and (f), all z_{base} and z_{top} retrievals are included without an upper height limit.

A.2 WATER VAPOR VARIABILITY

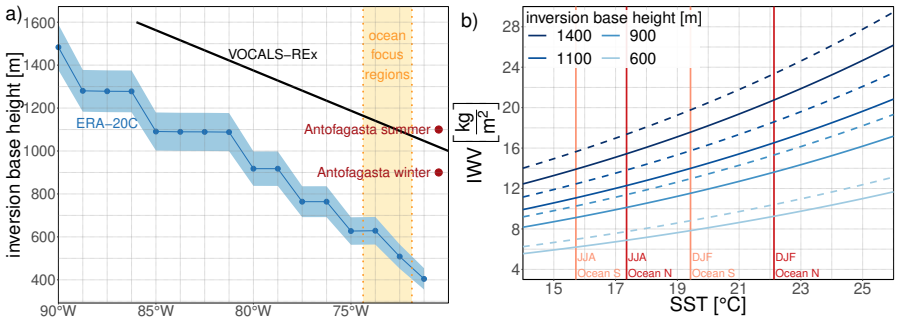

Figure A.5: a) Inversion base height derived from ERA-20C (blue) in a longitude–height cross section at 20°S over the southeast Pacific. The inversion base height was derived as follows: For the period from 15 October 2008 to 15 November 2008 the temporal mean vertical temperature profile was calculated for each horizontal grid point. The geometric height of the model level which showed the minimum temperature below 2500 m was taken to be the inversion height. The mean vertical extent of that level is indicated by blue shading. The two grid points which are covered by the northern ocean focus region are highlighted in orange. The location and time period for this cross section were chosen to match the observation campaign VOCALS-REx (Rahn et al., 2010). The black line indicates a rough linear fit for the campaign retrievals (from radio soundings) published by Rahn et al. (2010; their Fig. 4a). Red dots indicate the inversion base height at Antofagasta for summer and winter according to Muñoz et al (2011) who evaluated a multi-year record of radio sounding retrievals. b) Theoretical integrated water vapor (IWV) portion of the maritime boundary layer (MBL) in dependence on sea surface temperature (SST) for various inversion base heights and two different estimates of surface relative humidity ($rH = 0.8$, solid; $rH = 0.9$, dashed). The MBL component of IWV is derived as follows: $w_{\mathrm{MBL}} = \frac{1}{g} \cdot 0.622 \cdot \frac{rH \cdot e_s(\vartheta_{\mathrm{SST}})}{p_{\mathrm{surf}}} \cdot \Delta p$. The saturation vapor pressure $e_s(\vartheta_{\mathrm{SST}})$ is approximated using the Magnus Formula. The MBL thickness Δp is derived from the inversion base height via barometric formula. Mean SSTs from ERA-20C for the southern (orange) and northern (red) focus regions are indicated by vertical lines for austral summer (DJF) and winter (JJA) season.

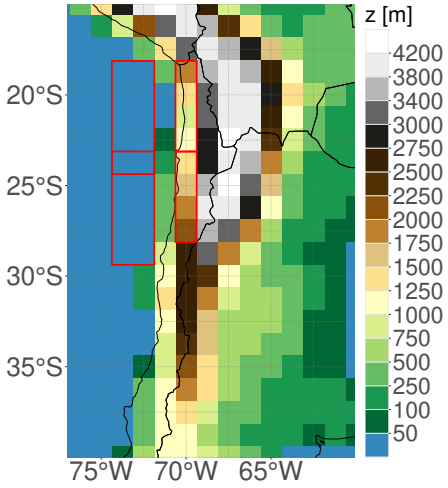

Figure A.6: Representation of topography in ERA-20C on a regular longitude by latitude grid with a horizontal resolution of 1.25°. Red rectangles illustrate the four focus regions of this study.

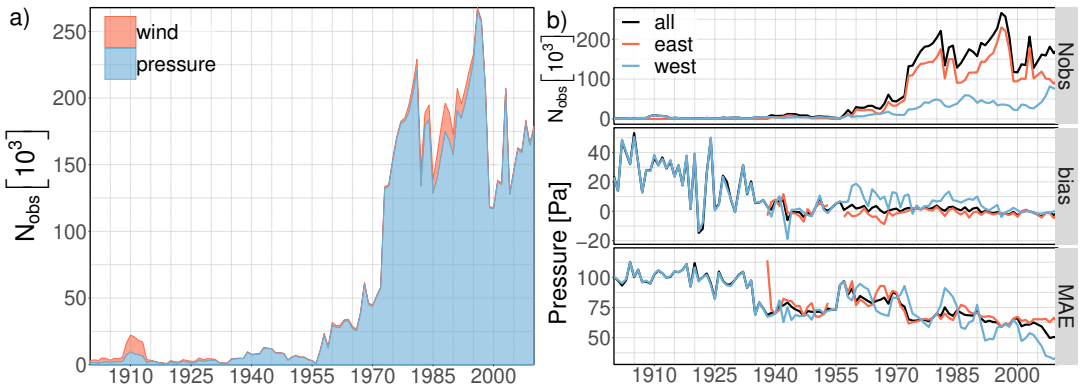

Figure A.7: a) Number of assimilated wind and pressure observations per year over time for ERA-20C. Lines are stacked. Only observations between 15°S–40°S and 90°W–60°W are considered (region displayed in Fig. A.8). b) Number of assimilated surface pressure observations (N_{obs}, top), annual mean bias (middle), bias corrected annual mean absolute error (bottom) between ERA-20C and assimilated observations of surface pressure. Results are displayed for the whole region (black) and for west (blue) and east (orange) of 70°W, roughly splitting the region into an ocean and a land part. For the eastern region, the statistics are only shown from 1938 onward because of very low numbers of observations for previous years which can result in large biases due to the random nature of model/observation biases. The bias correction is carried out monthly. Therefore, for each month, the mean departure is subtracted from each individual observation departure within the respective month. N_{obs} is given in thousands.

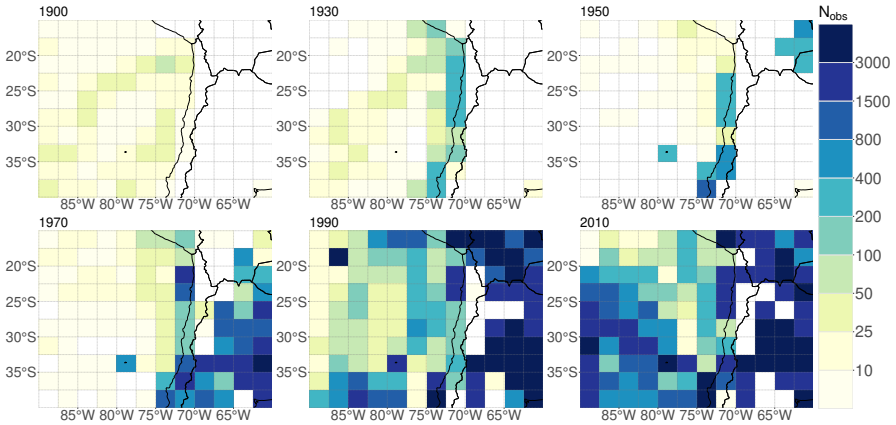

Figure A.8: Spatial distribution of the number of assimilated surface pressure observations per year for ERA-20C for selected years.

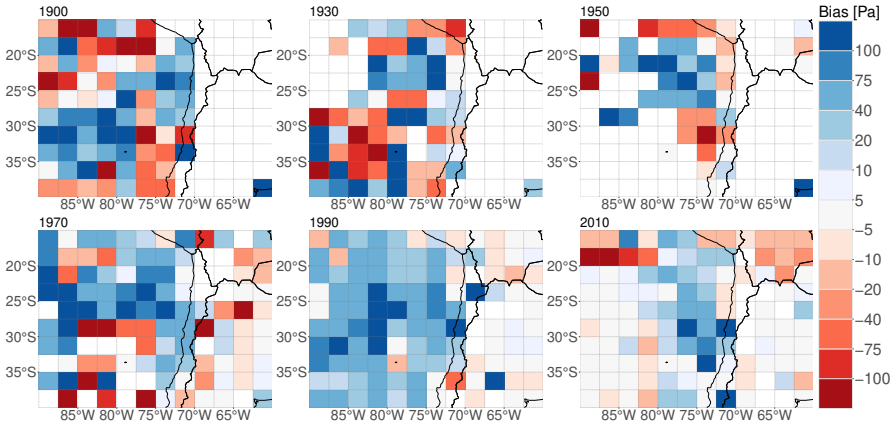

Figure A.9: Spatial distribution of annual mean departure of assimilated surface pressure observations from ERA-20C model analysis (bias) for selected years.

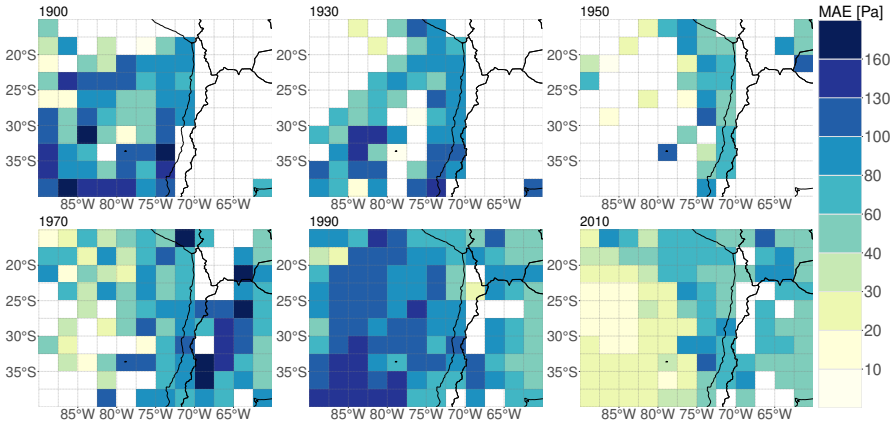

Figure A.10: Spatial distribution of annual mean absolute departure of assimilated surface pressure observations from ERA-20C model analysis (MAE) for selected years.

A.3 FOG

FIGURES A.11–A.22:

Self organizing maps (SOMs) for all stations analog to Fig. 6.2 of the manuscript. Additionally, the number of events N assigned to each grid cell is provided.

FIGURES A.23–A.26:

Time series of the fog frequency derived from the MODIS CTH, the neural network probabilistic output (NNprob) in ALL- and LOO-training mode, and the ground-based reference measurements and the high cloud frequency according to the MODIS CTH for all climate stations analog to Fig. 6.11a of the manuscript.

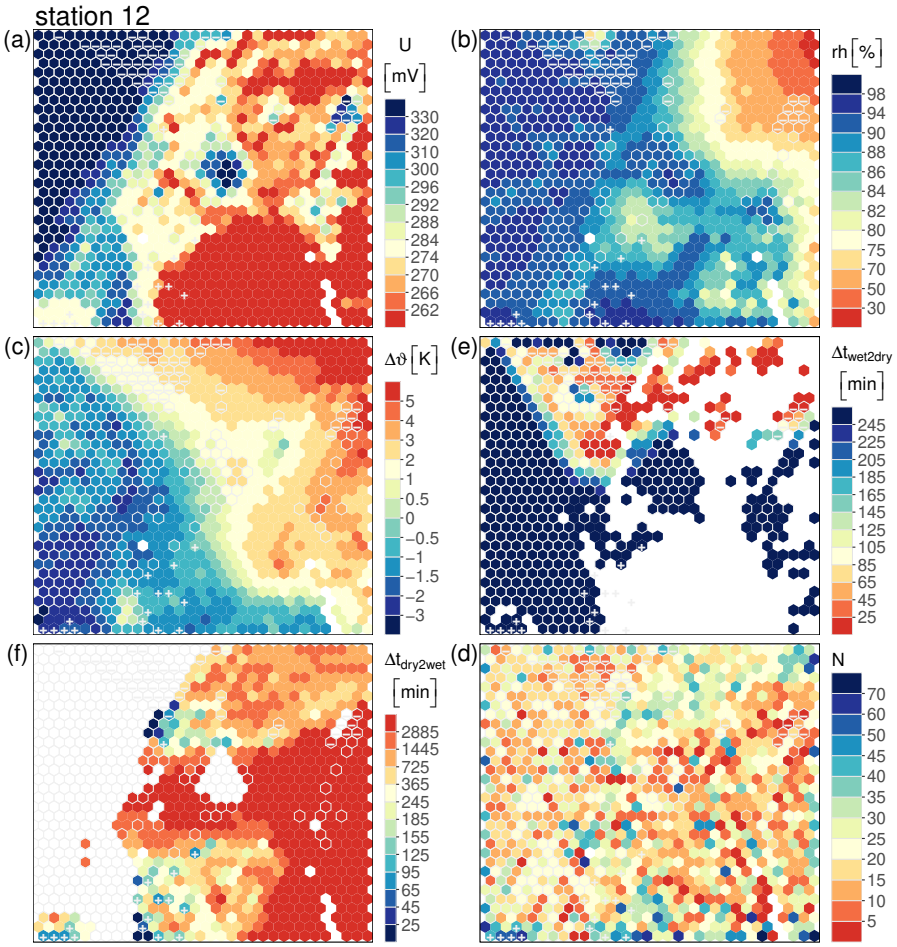
station 12
(a)
U
[mV]
330
320
310
300
296
292
288
284
274
270
266
262
(b)
rh [%]
98
94
90
88
86
84
82
80
75
70
50
30
(c)
Δϑ [K]
5
4
3
2
1
0.5
0
−0.5
−1
−1.5
−2
−3
(e)
Δt wet2dry
[min]
245
225
205
185
165
145
125
105
85
65
45
25
(f)
Δt dry2wet
[min]
2885
1445
725
365
245
185
155
125
95
65
45
25
(d)
N
70
60
50
45
40
35
30
25
20
15
10
5

Figure A.11

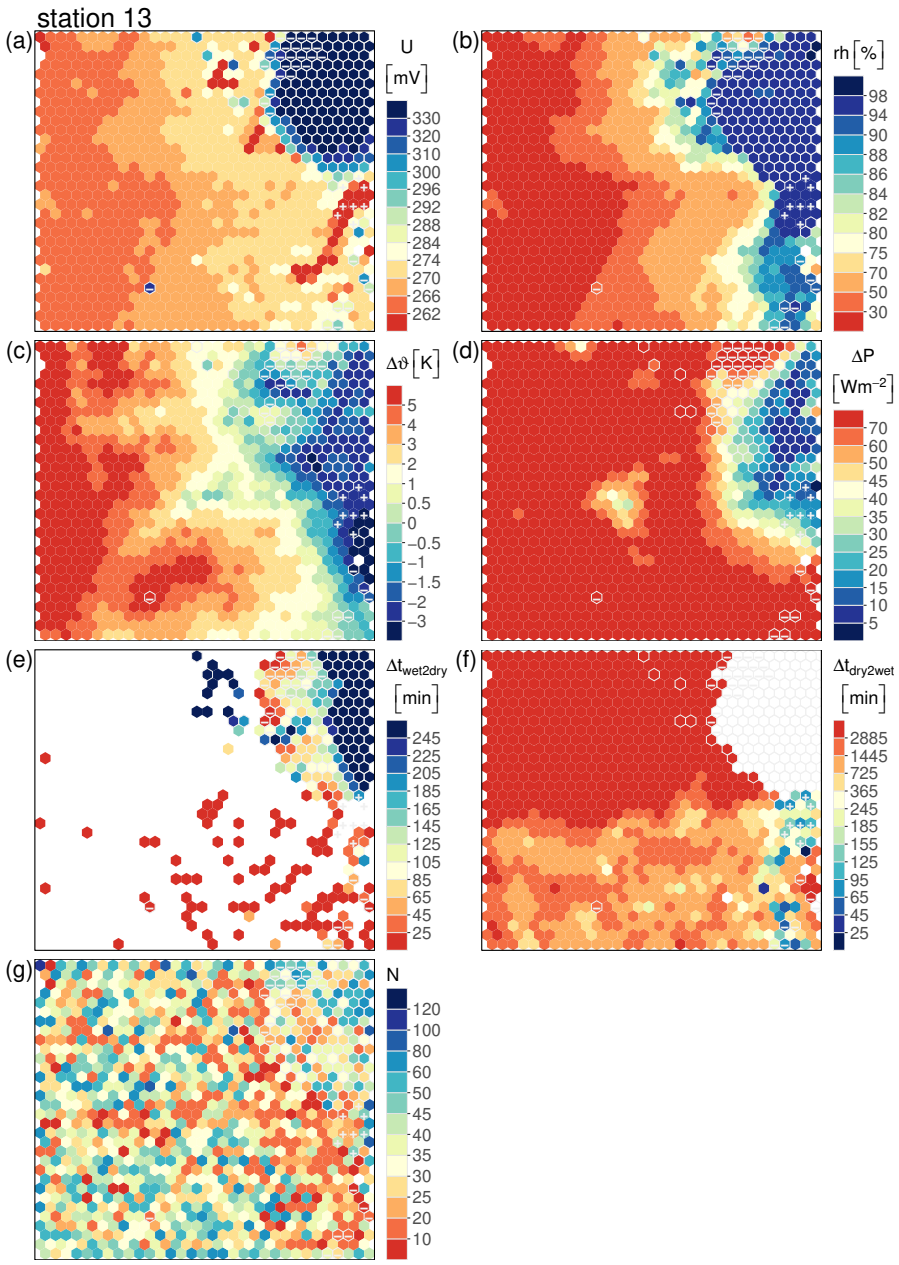
station 13
(a)
U
[mV]
330
320
310
300
296
292
288
284
274
270
266
262
(b)
rh [%]
98
94
90
88
86
84
82
80
75
70
50
30
(c)
Δϑ [K]
5
4
3
2
1
0.5
0
−0.5
−1
−1.5
−2
−3
(d)
ΔP
[Wm−2]
70
60
50
45
40
35
30
25
20
15
10
5
(e)
Δtwet2dry
[min]
245
225
205
185
165
145
125
105
85
65
45
25
(f)
Δtdry2wet
[min]
2885
1445
725
365
245
185
155
125
95
65
45
25
(g)
N
120
100
80
60
50
45
40
35
30
25
20
10

Figure A.12

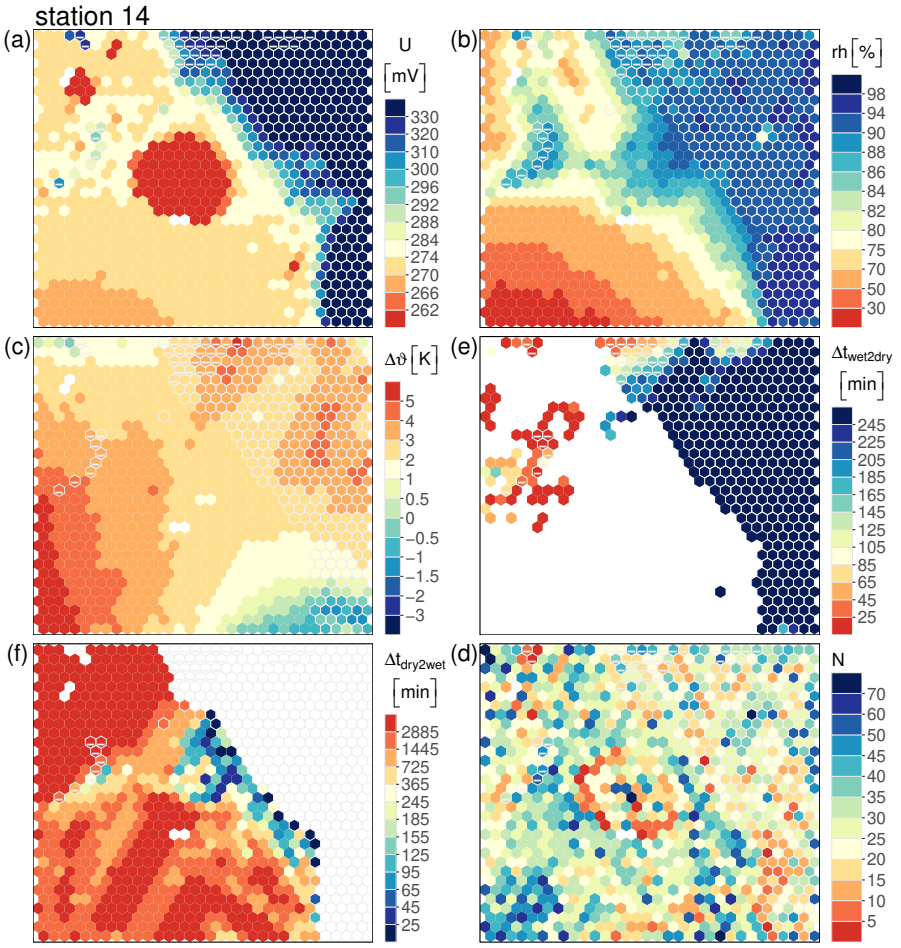
station 14
(a)
U
[mV]
330
320
310
300
296
292
288
284
274
270
266
262
(b)
rh [%]
98
94
90
88
86
84
82
80
75
70
50
30
(c)
Δϑ [K]
5
4
3
2
1
0.5
0
−0.5
−1
−1.5
−2
−3
(e)
Δt wet2dry
[min]
245
225
205
185
165
145
125
105
85
65
45
25
(f)
Δt dry2wet
[min]
2885
1445
725
365
245
185
155
125
95
65
45
25
(d)
N
70
60
50
45
40
35
30
25
20
15
10
5

Figure A.13

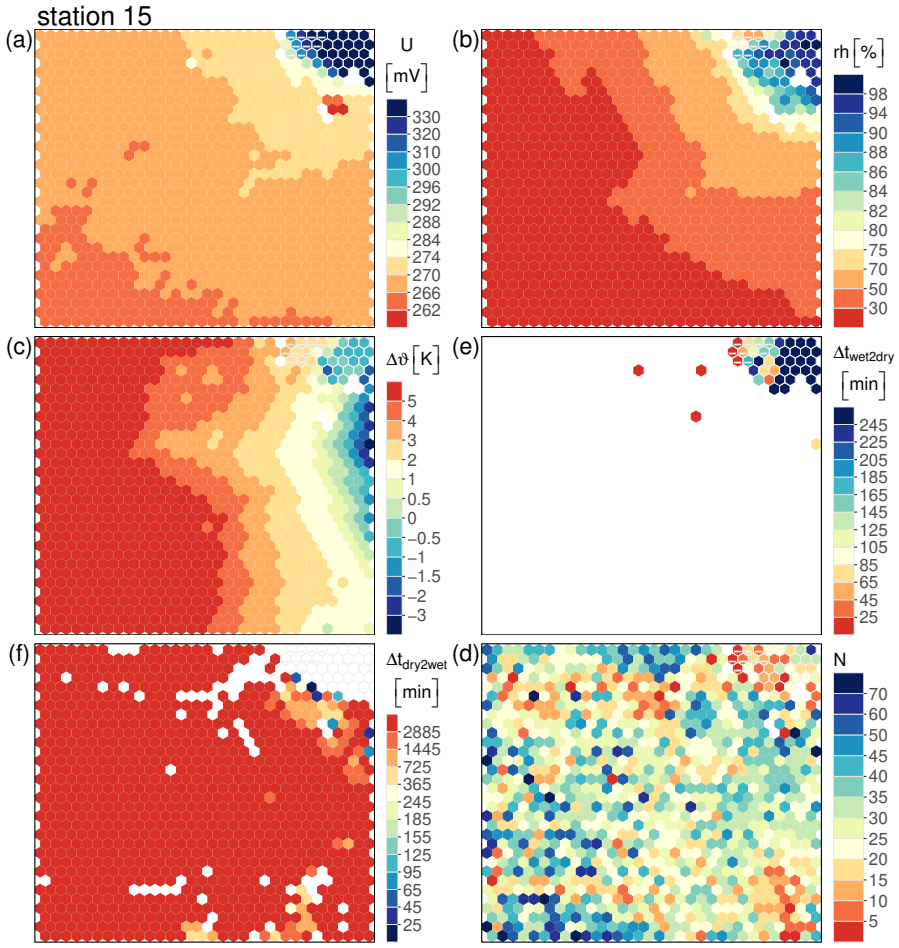
station 15
(a)
U
[mV]
330
320
310
300
296
292
288
284
274
270
266
262
(b)
rh [%]
98
94
90
88
86
84
82
80
75
70
50
30
(c)
Δϑ [K]
5
4
3
2
1
0.5
0
−0.5
−1
−1.5
−2
−3
(e)
Δt wet2dry
[min]
245
225
205
185
165
145
125
105
85
65
45
25
(f)
Δt dry2wet
[min]
2885
1445
725
365
245
185
155
125
95
65
45
25
(d)
N
70
60
50
45
40
35
30
25
20
15
10
5

Figure A.14

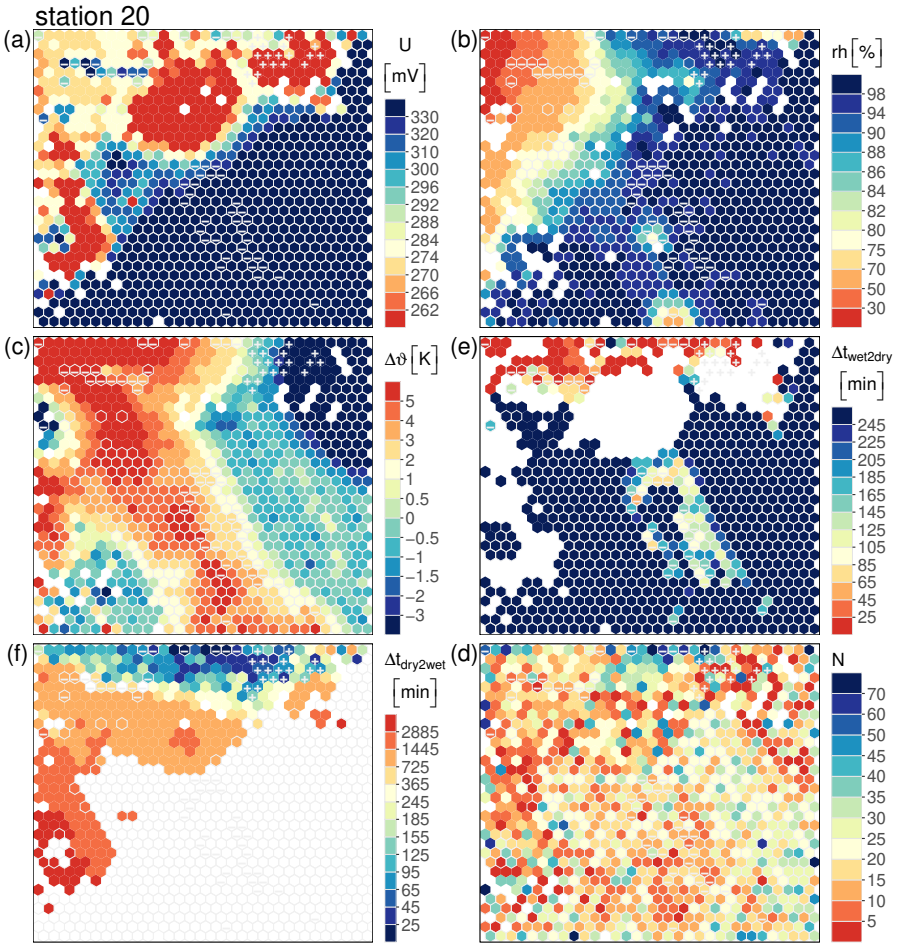
station 20
(a)
U
[mV]
330
320
310
300
296
292
288
284
274
270
266
262
(b)
rh [%]
98
94
90
88
86
84
82
80
75
70
50
30
(c)
Δϑ [K]
5
4
3
2
1
0.5
0
−0.5
−1
−1.5
−2
−3
(e)
Δt wet2dry
[min]
245
225
205
185
165
145
125
105
85
65
45
25
(f)
Δt dry2wet
[min]
2885
1445
725
365
245
185
155
125
95
65
45
25
(d)
N
70
60
50
45
40
35
30
25
20
15
10
5

Figure A.15

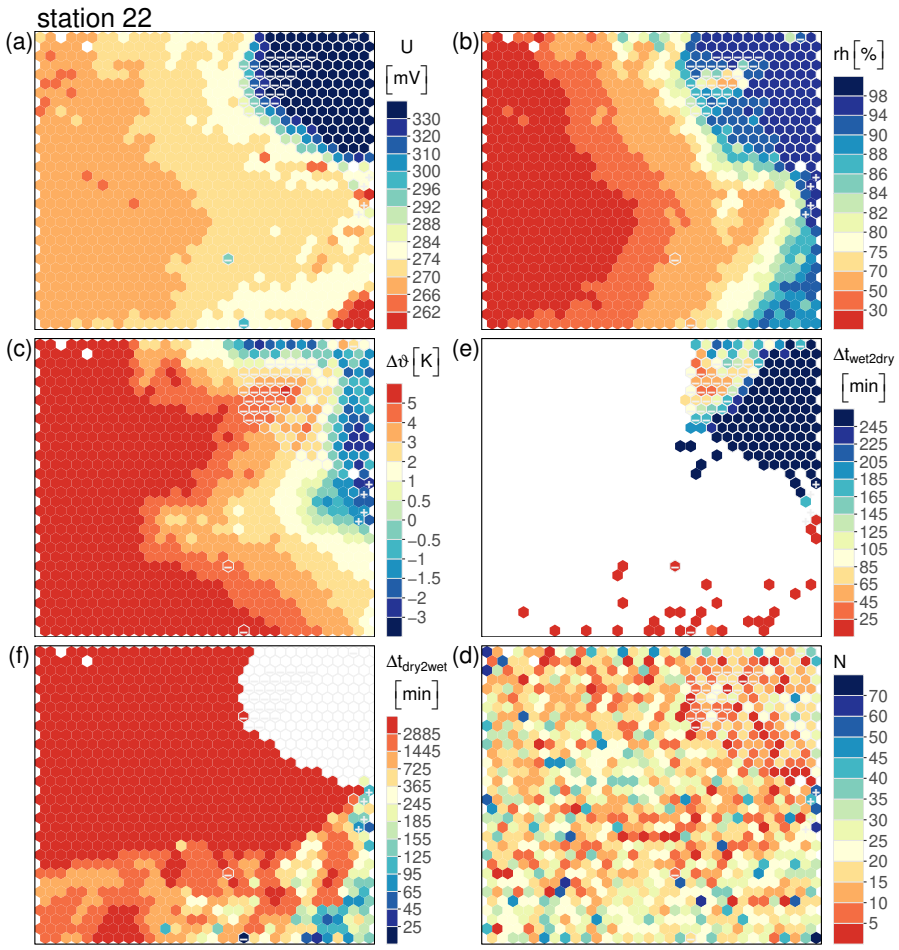
station 22
(a)
U
[mV]
330
320
310
300
296
292
288
284
274
270
266
262
(b)
rh [%]
98
94
90
88
86
84
82
80
75
70
50
30
(c)
Δϑ [K]
5
4
3
2
1
0.5
0
−0.5
−1
−1.5
−2
−3
(e)
Δt wet2dry
[min]
245
225
205
185
165
145
125
105
85
65
45
25
(f)
Δt dry2wet
[min]
2885
1445
725
365
245
185
155
125
95
65
45
25
(d)
N
70
60
50
45
40
35
30
25
20
15
10
5

Figure A.16

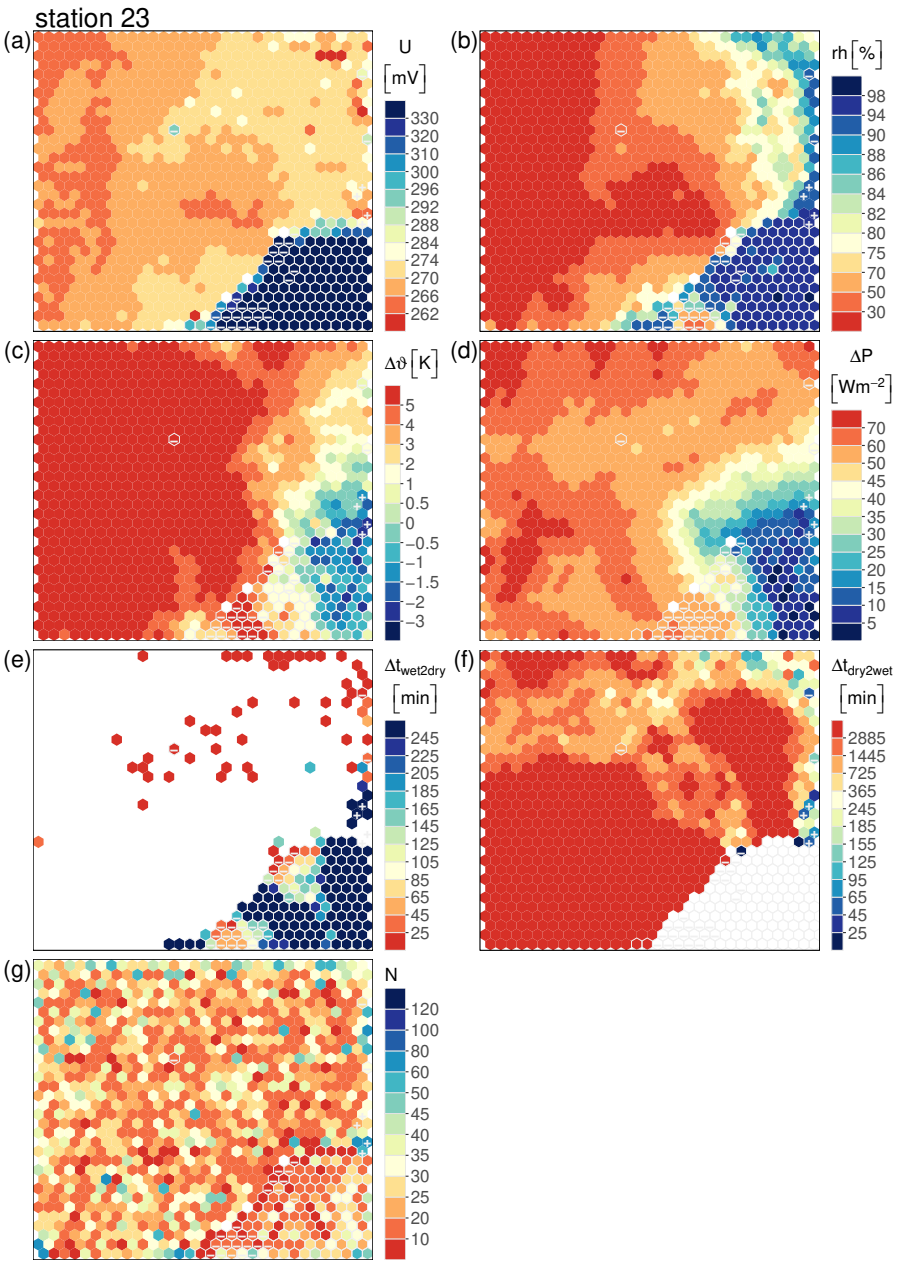
station 23
(a)
U
[mV]
330
320
310
300
296
292
288
284
274
270
266
262
(b)
rh [%]
98
94
90
88
86
84
82
80
75
70
50
30
(c)
Δϑ [K]
5
4
3
2
1
0.5
0
−0.5
−1
−1.5
−2
−3
(d)
ΔP
[Wm−2]
70
60
50
45
40
35
30
25
20
15
10
5
(e)
Δtwet2dry
[min]
245
225
205
185
165
145
125
105
85
65
45
25
(f)
Δtdry2wet
[min]
2885
1445
725
365
245
185
155
125
95
65
45
25
(g)
N
120
100
80
60
50
45
40
35
30
25
20
10

Figure A.17

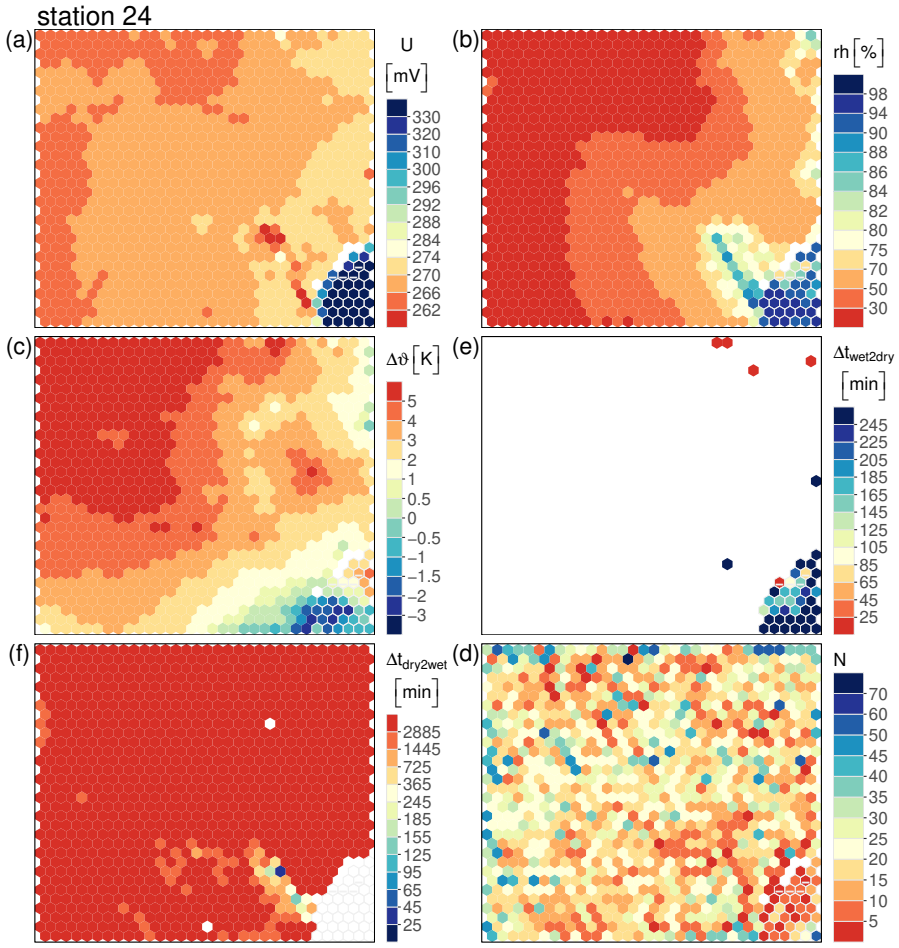
station 24
(a)
U
[mV]
330
320
310
300
296
292
288
284
274
270
266
262
(b)
rh [%]
98
94
90
88
86
84
82
80
75
70
50
30
(c)
Δϑ [K]
5
4
3
2
1
0.5
0
−0.5
−1
−1.5
−2
−3
(e)
Δt wet2dry
[min]
245
225
205
185
165
145
125
105
85
65
45
25
(f)
Δt dry2wet
[min]
2885
1445
725
365
245
185
155
125
95
65
45
25
(d)
N
70
60
50
45
40
35
30
25
20
15
10
5

Figure A.18

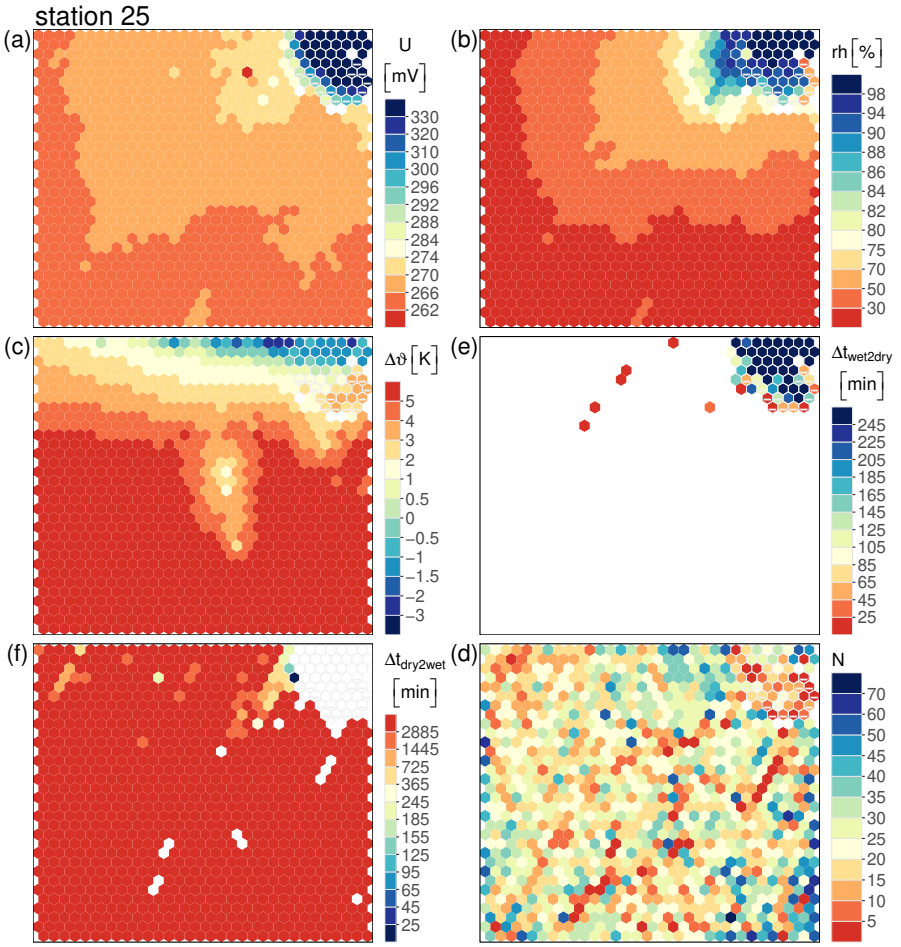
station 25
(a)
U
[mV]
330
320
310
300
296
292
288
284
274
270
266
262
(b)
rh [%]
98
94
90
88
86
84
82
80
75
70
50
30
(c)
Δϑ [K]
5
4
3
2
1
0.5
0
−0.5
−1
−1.5
−2
−3
(e)
Δt wet2dry
[min]
245
225
205
185
165
145
125
105
85
65
45
25
(f)
Δt dry2wet
[min]
2885
1445
725
365
245
185
155
125
95
65
45
25
(d)
N
70
60
50
45
40
35
30
25
20
15
10
5

Figure A.19

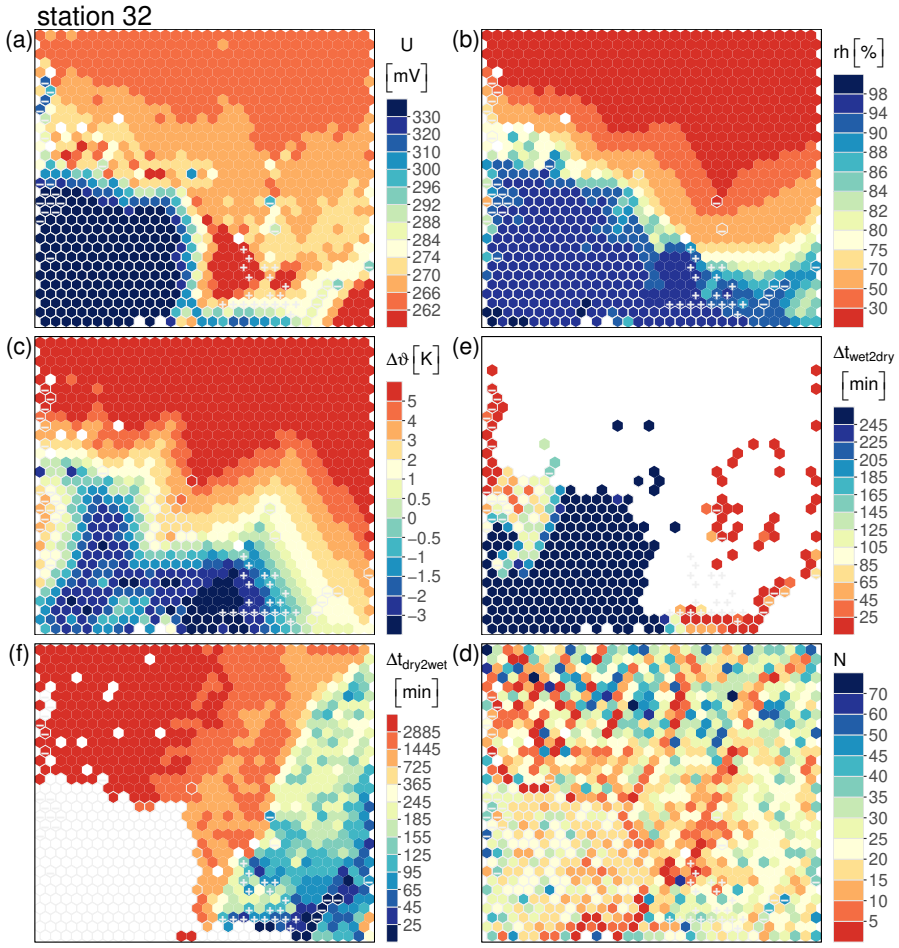
station 32
(a)
U
[mV]
330
320
310
300
296
292
288
284
274
270
266
262
(b)
rh [%]
98
94
90
88
86
84
82
80
75
70
50
30
(c)
Δϑ [K]
5
4
3
2
1
0.5
0
−0.5
−1
−1.5
−2
−3
(e)
Δt wet2dry
[min]
245
225
205
185
165
145
125
105
85
65
45
25
(f)
Δt dry2wet
[min]
2885
1445
725
365
245
185
155
125
95
65
45
25
(d)
N
70
60
50
45
40
35
30
25
20
15
10
5

Figure A.20

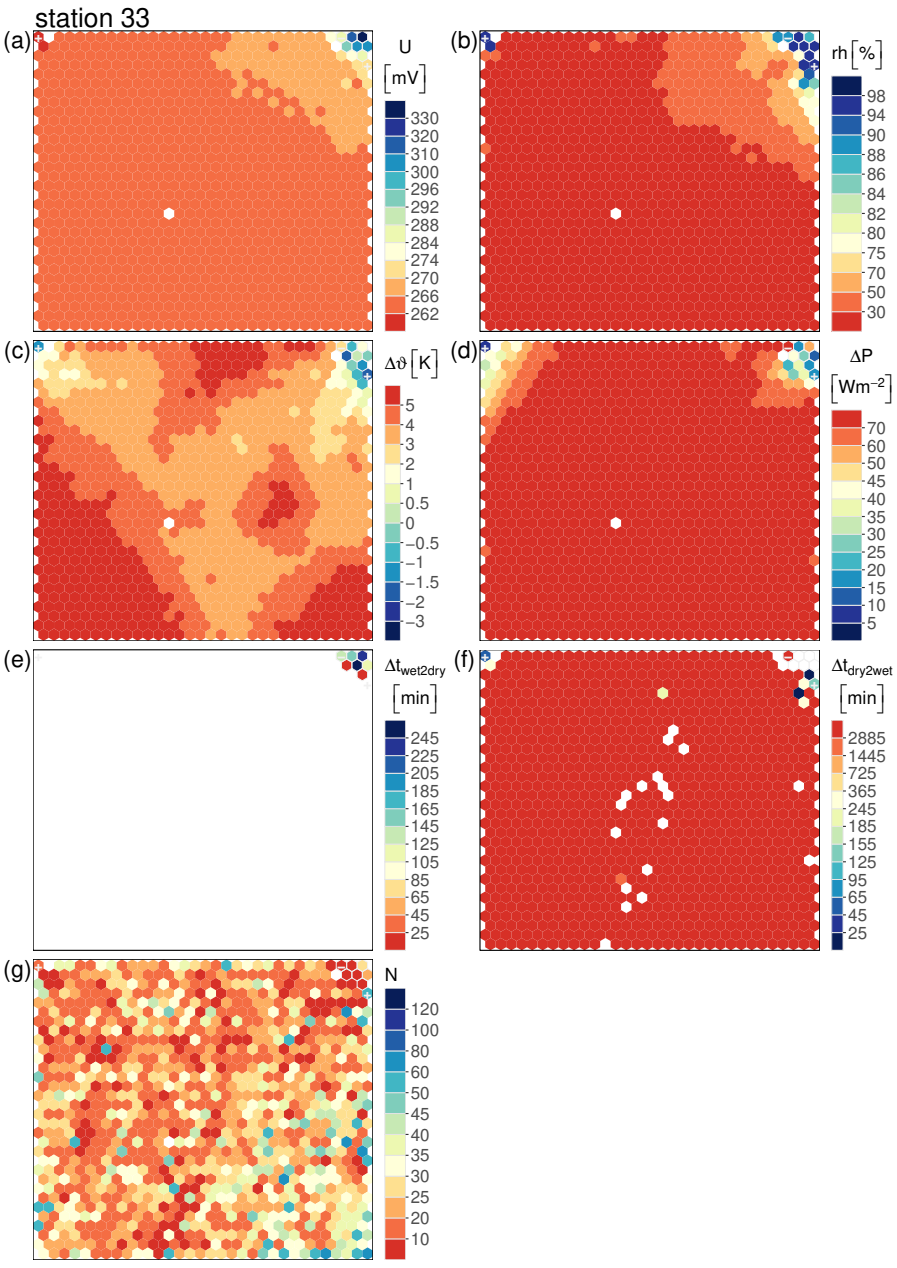
station 33
(a)
U
[mV]
330
320
310
300
296
292
288
284
274
270
266
262
(b)
rh [%]
98
94
90
88
86
84
82
80
75
70
50
30
(c)
Δϑ [K]
5
4
3
2
1
0.5
0
−0.5
−1
−1.5
−2
−3
(d)
ΔP
[Wm−2]
70
60
50
45
40
35
30
25
20
15
10
5
(e)
Δtwet2dry
[min]
245
225
205
185
165
145
125
105
85
65
45
25
(f)
Δtdry2wet
[min]
2885
1445
725
365
245
185
155
125
95
65
45
25
(g)
N
120
100
80
60
50
45
40
35
30
25
20
10

Figure A.21

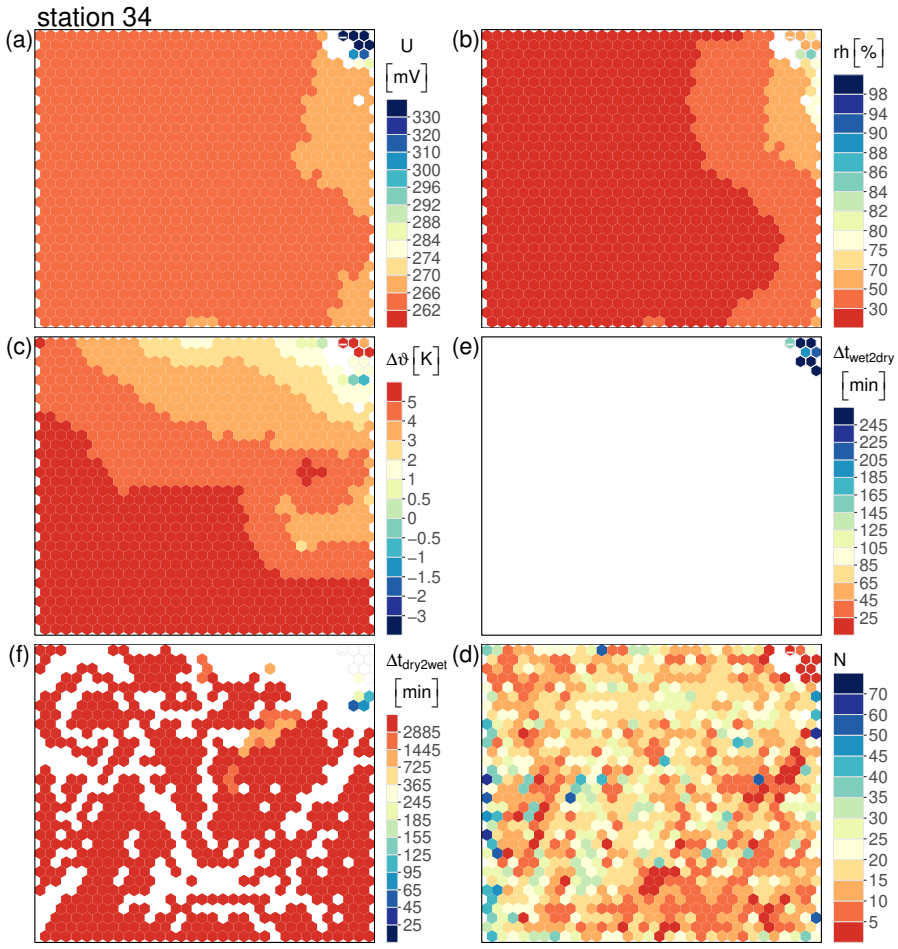

Figure A.22

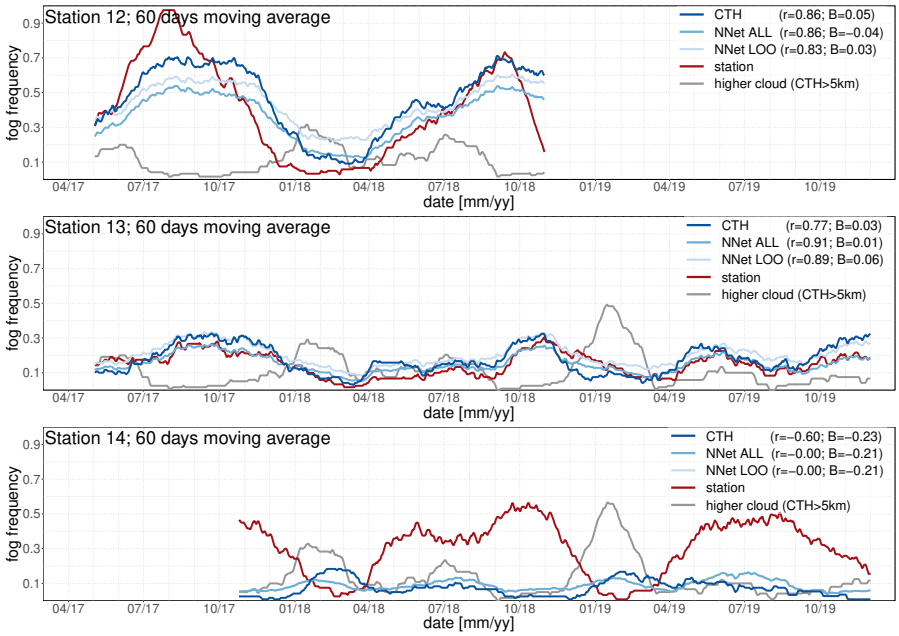

Figure A.23

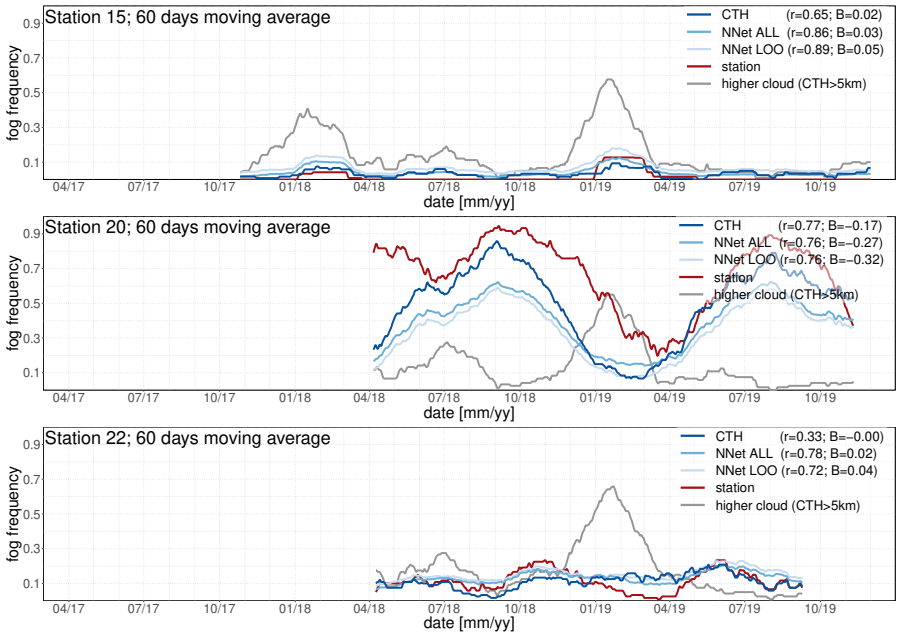

Figure A.24

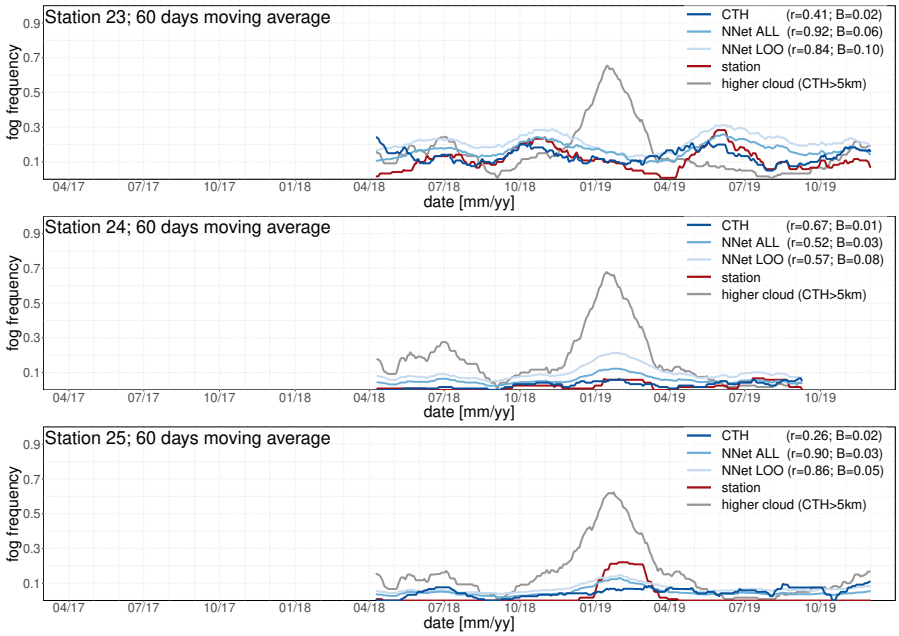

Figure A.25

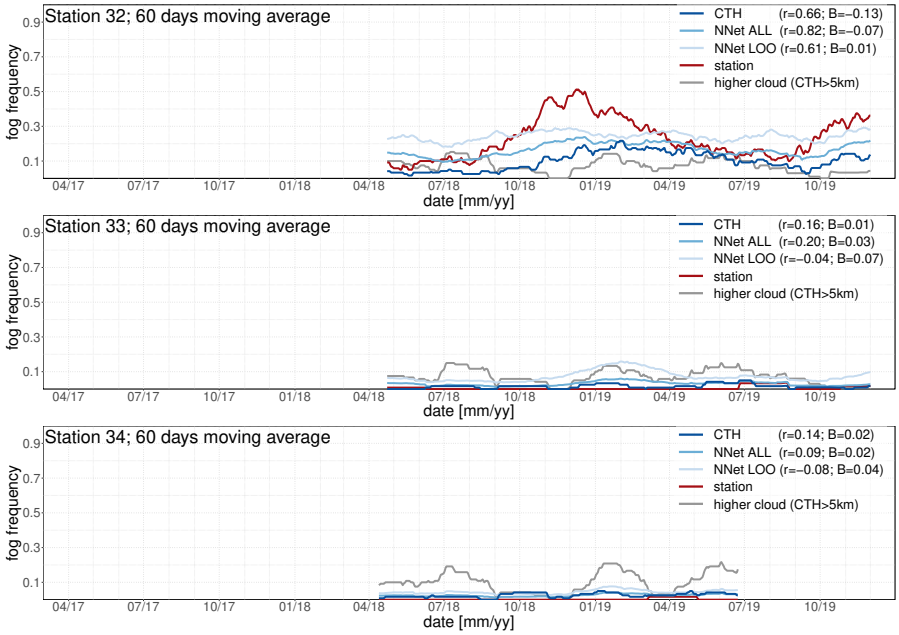
Station 32; 60 days moving average
CTH (r=0.66; B=−0.13)
NNet ALL (r=0.82; B=−0.07)
NNet LOO (r=0.61; B=0.01)
station
higher cloud (CTH>5km)
fog frequency
0.9
0.7
0.5
0.3
0.1
04/17
07/17
10/17
01/18
04/18
07/18
10/18
01/19
04/19
07/19
10/19
date [mm/yy]
Station 33; 60 days moving average
CTH (r=0.16; B=0.01)
NNet ALL (r=0.20; B=0.03)
NNet LOO (r=−0.04; B=0.07)
station
higher cloud (CTH>5km)
fog frequency
date [mm/yy]
Station 34; 60 days moving average
CTH (r=0.14; B=0.02)
NNet ALL (r=0.09; B=0.02)
NNet LOO (r=−0.08; B=0.04)
station
higher cloud (CTH>5km)
fog frequency
date [mm/yy]

Figure A.26

BIBLIOGRAPHY

Abadi, M. et al. (2015). *TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems*. Software available from tensorflow.org.

Ackerman, S., P. Menzel, R. Frey, and B. Baum (2015). "MODIS Atmosphere L2 Cloud Mask Product." In: *NASA MODIS Adaptive Processing System, Goddard Space Flight Center*. DOI: 10.5067/MODIS/MOD35_L2.006.

Albers, C. J. (2013). *Salares*. last access: 07/07/2020 23:04 UTC.

Albrecht, B. A., C. W. Fairall, D. W. Thomson, A. B. White, J. B. Snider, and W. H. Schubert (1990). "Surface-based remote sensing of the observed and the Adiabatic liquid water content of stratocumulus clouds." In: *Geophysical Research Letters* 17.1, pp. 89–92. DOI: 10.1029/GL017i001p00089.

Alexander, M. A., I. Bladé, M. Newman, J. R. Lanzante, N.-C. Lau, and J. D. Scott (Aug. 2002). "The Atmospheric Bridge: The Influence of ENSO Teleconnections on Air–Sea Interaction over the Global Oceans." In: *Journal of Climate* 15.16, pp. 2205–2231. ISSN: 0894-8755. DOI: 10.1175/1520-0442(2002)015<2205:TABTIO>2.0.CO;2.

Alexander, M. A. and C. Deser (Jan. 1995). "A Mechanism for the Recurrence of Wintertime Midlatitude SST Anomalies." In: *Journal of Physical Oceanography* 25.1, pp. 122–137. ISSN: 0022-3670. DOI: 10.1175/1520-0485(1995)025<0122:AMFTRO>2.0.CO;2.

Alexander, M. A. and J. D. Scott (Nov. 2008). "The Role of Ekman Ocean Heat Transport in the Northern Hemisphere Response to ENSO." In: *Journal of Climate* 21.21, pp. 5688–5707. ISSN: 0894-8755. DOI: 10.1175/2008JCLI2382.1.

Amundson, R., A. T. Austin, E. A. G. Schuur, K. Yoo, V. Matzek, C. Kendall, A. Uebersax, D. Brenner, and W. T. Baisden (2003). "Global patterns of the isotopic composition of soil and plant nitrogen." In: *Global Biogeochemical Cycles* 17.1. DOI: 10.1029/2002GB001903.

Andersen, H. and J. Cermak (2018). "First fully diurnal fog and low cloud satellite detection reveals life cycle in the Namib." In: *Atmospheric Measurement Techniques* 11.10, pp. 5461–5470. DOI: 10.5194/amt-11-5461-2018.

Andersson, A., K. Graw, M. Schröder, K. Fennig, J. Liman, S. Bakan, R. Hollmann, and C. Klepp (2017a). *Hamburg Ocean Atmosphere Parameters and Fluxes from Satellite Data - HOAPS 4.0*. DOI: 10.5676/EUM_SAF_CM/HOAPS/V002.

Andersson, A., K. Graw, M. Schröder, K. Fennig, J. Liman, and R. Hollmann (2017b). *Validation Report SSM/I and SSMIS products HOAPS version 4.0*. DOI: 10.5676/EUM_SAF_CM/HOAPS/V002.

Andreoli, R. V. and M. T. Kayano (2005). "ENSO-related rainfall anomalies in South America and associated circulation features during warm and cold Pacific decadal oscillation regimes." In: *International Journal of Climatology* 25.15, pp. 2017–2030. DOI: 10.1002/joc.1222.

Ashok, K., S. K. Behera, S. A. Rao, H. Weng, and T. Yamagata (2007). "El Niño Modoki and its possible teleconnection." In: *Journal of Geophysical Research: Oceans* 112.C11. DOI: 10.1029/2006JC003798.

Astorga-Eló, M., Q. Zhang, G. Larama, A. Stoll, M. J. Sadowsky, and M. A. Jorquera (2020). "Composition, Predicted Functions and Co-occurrence Networks of Rhizobacterial Communities Impacting Flowering Desert Events in the Atacama Desert, Chile." eng. In: *Frontiers in microbiology* 11, pp. 571–571. ISSN: 1664-302X. DOI: https://doi.org/10.3389/fmicb.2020.00571.

Azua-Bustos, A. et al. (2018). "Unprecedented rains decimate surface microbial communities in the hyperarid core of the Atacama Desert." In: *Scientific Reports* 8.1, p. 16706. ISSN: 2045-2322. DOI: 10.1038/s41598-018-35051-w.

Barrett, B. S., D. A. Campos, J. V. Veloso, and R. Rondanelli (2016). "Extreme temperature and precipitation events in March 2015 in central and northern Chile." In: *Journal of Geophysical Research: Atmospheres* 121.9, pp. 4563–4580. DOI: 10.1002/2016JD024835.

Bartz, M., J. Walk, S. A. Binnie, D. Brill, G. Stauch, F. Lehmkuhl, D. Hoffmeister, and H. Brückner (2020). "Late Pleistocene alluvial fan evolution along the coastal Atacama Desert (N Chile)." In: *Global and Planetary Change* 190, p. 103091. ISSN: 0921-8181. DOI: https://doi.org/10.1016/j.gloplacha.2019.103091.

Baum, B. A., W. P. Menzel, R. A. Frey, D. C. Tobin, R. E. Holz, S. A. Ackerman, A. K. Heidinger, and P. Yang (June 2012). "MODIS Cloud-Top Property Refinements for Collection 6." In: *Journal of Applied Meteorology and Climatology* 51.6, pp. 1145–1163. ISSN: 1558-8424. DOI: 10.1175/JAMC-D-11-0203.1.

Bendix, J., B. Thies, T. Nauß, and J. Cermak (2006). "A feasibility study of daytime fog and low stratus detection with TERRA/AQUA-MODIS over land." In: *Meteorological Applications* 13.2, pp. 111–125. DOI: 10.1017/S1350482706002180.

Blamey, R. C., A. M. Ramos, R. M. Trigo, R. Tomé, and C. J. C. Reason (Jan. 2018). "The Influence of Atmospheric Rivers over the South Atlantic on Winter Rainfall in South Africa." In: *Journal of Hydrometeorology* 19.1, pp. 127–142. ISSN: 1525-755X. DOI: 10.1175/JHM-D-17-0111.1.

Böhm, C., O. Sourdeval, J. Mülmenstädt, J. Quaas, and S. Crewell (2019). "Cloud base height retrieval from multi-angle satellite data." In: *Atmospheric Measurement Techniques* 12.3, pp. 1841–1860. DOI: 10.5194/amt-12-1841-2019.

Böhm, C. (2019). *MIBase cloud base height derived from satellite data*. [Accessed 21. February 2019]. DOI: 10.5880/CRC1211DB.19.

Böhm, C., M. Reyers, J. H. Schween, and S. Crewell (2020a). "Water vapor variability in the Atacama Desert during the 20th century." In: *Global and Planetary Change* 190, p. 103192. ISSN: 0921-8181. DOI: 10.1016/j.gloplacha.2020.103192.

Böhm, C., J. H. Schween, M. Reyers, B. Maier, U. Löhnert, and S. Crewell (2020b). "Towards a climatology of fog frequency in the Atacama Desert via multi-spectral satellite data and machine learning techniques." Submitted to: *Journal of Applied Meteorology and Climatology*, date of submission: 10 Sep 2020.

Borbas, E., P. Menzel, and B. Gao (2017a). "MODIS Atmosphere L2 Water Vapor Product." In: *NASA MODIS Adaptive Processing System, Goddard Space Flight Center*. DOI: 10.5067/MODIS/MOD05_L2.061.

Borbas, E., P. Menzel, and B. Gao (2017b). "MODIS Atmosphere L2 Water Vapor Product." In: *NASA MODIS Adaptive Processing System, Goddard Space Flight Center*. DOI: 10.5067/MODIS/MYD05_L2.061.

Boucher, O. et al. (2013). "Clouds and Aerosols." In: *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*. Ed. by T. Stocker, D. Qin, G.-K. Plattner, M. Tignor, S. Allen, J. Boschung, A. Nauels, Y. Xia, V. Bex, and P. M. (eds.) Cambridge, United Kingdom and New York, NY, USA: Cambridge University Press. Chap. 7, pp. 571–657.

Bozkurt, D., R. Rondanelli, R. Garreaud, and A. Arriagada (2016). "Impact of Warmer Eastern Tropical Pacific SST on the March 2015 Atacama Floods." In: *Monthly Weather Review* 144.11, pp. 4441–4460. DOI: 10.1175/MWR-D-16-0041.1.

Bretherton, C. S. and M. C. Wyant (Jan. 1997). "Moisture Transport, Lower-Tropospheric Stability, and Decoupling of Cloud-Topped Boundary Layers." In: *Journal of the Atmospheric Sciences* 54.1, pp. 148–167. ISSN: 0022-4928. DOI: 10.1175/1520-0469(1997)054<0148:MTLTSA>2.0.CO;2.

Bretherton, C. S. (1997). "Convection in Stratocumulus-Topped Atmospheric Boundary Layers." In: *The Physics and Parameterization of Moist Atmospheric Convection*. Ed. by R. Smith. Springer Netherlands, pp. 127–142.

Brown, R. L., W. Wild, and C. Cunningham (2004). "ALMA – the Atacama large millimeter array." In: *Advances in Space Research* 34.3. Astronomy at IR/Submm and the Microwave Background, pp. 555–559. ISSN: 0273-1177. DOI: https://doi.org/10.1016/j.asr.2003.03.028.

Bull, M., J. Matthews, D. McDonald, A. Menzies, C. Moroney, K. Mueller, S. Paradise, and M. Smyth (2011). *Data Products Specifications*. Tech. rep. JPL D-13963, Revision S. Jet Propulsion Laboratory, California Institute of Technology.

Cáceres, L., B. Gómez-Silva, X. Garró, V. Rodríguez, V. Monardes, and C. P. McKay (2007). "Relative humidity patterns and fog water

precipitation in the Atacama Desert and biological implications." In: *Journal of Geophysical Research: Biogeosciences* 112.G4. DOI: 10.1029/2006JG000344.

Campbell Scientific (2018). *LWS Dielectric Leaf Wetness Sensor*. Campbell Scientific, Inc. 815 West 1800 North Logan, Utah 84321-1784.

Canedo-Rosso, C., C. B. Uvo, and R. Berndtsson (2019). "Precipitation variability and its relation to climate anomalies in the Bolivian Altiplano." In: *International Journal of Climatology* 39.4, pp. 2096–2107. DOI: 10.1002/joc.5937.

Cantalloube, F., J. Milli, C. Böhm, S. Crewell, J. Navarrete, K. Rehfeld, M. Sarazin, and A. Sommani (2020). "The impact of climate change on astronomical observations." In: *Nature Astronomy* 4, pp. 826–829. ISSN: 2397-3366. DOI: 10.1038/s41550-020-1203-3.

Capotondi, A. et al. (July 2015). "Understanding ENSO Diversity." In: *Bulletin of the American Meteorological Society* 96.6, pp. 921–938. ISSN: 0003-0007. DOI: 10.1175/BAMS-D-13-00117.1.

Center for Climate and Resilience Research (2018). *Datos de precipitación*. http://www.cr2.cl/datos-de-precipitacion/. File: cr2_prDaily_2018.zip. Accessed: 2019-07-01.

Center for Climate and Resilience Research (2019). *EXPLORADOR CLIMÁTICO*. http://explorador.cr2.cl. Accessed: 2020-09-22.

Cereceda, P., H. Larrain, P. Osses, M. Farías, and I. Egaña (2008a). "The climate of the coast and fog zone in the Tarapacá Region, Atacama Desert, Chile." In: *Atmospheric Research* 87.3. Third International Conference on Fog, Fog Collection and Dew, pp. 301–311. ISSN: 0169-8095. DOI: https://doi.org/10.1016/j.atmosres.2007.11.011.

Cereceda, P., H. Larrain, P. Osses, M. Farías, and I. Egaña (2008b). "The spatial and temporal variability of fog and its relation to fog oases in the Atacama Desert, Chile." In: *Atmospheric Research* 87.3. Third International Conference on Fog, Fog Collection and Dew, pp. 312–323. ISSN: 0169-8095. DOI: http://dx.doi.org/10.1016/j.atmosres.2007.11.012.

Cereceda, P., P Osses, H Larrain, M. Farías, M. Lagos, R Pinto, and R. Schemenauer (Sept. 2002). "Advective, orographic and radiation fog in the Tarapacá Region, Chile." In: *Atmospheric Research* 64.1. 2nd International Conference on Fog and Fog Collection, pp. 261–271. ISSN: 0169-8095. DOI: 10.1016/S0169-8095(02)00097-2.

Cermak, J. (2012). "Low clouds and fog along the South-Western African coast – Satellite-based retrieval and spatial patterns." In: *Atmospheric Research* 116. Remote Sensing of Clouds and Aerosols: Techniques and Applications - Atmospheric Research, pp. 15 –21. ISSN: 0169-8095. DOI: https://doi.org/10.1016/j.atmosres.2011.02.012.

Chahine, M. T. (Jan. 1974). "Remote Sounding of Cloudy Atmospheres. I. The Single Cloud Layer." In: *Journal of the Atmospheric Sciences*

31.1, pp. 233–243. ISSN: 0022-4928. DOI: 10.1175/1520-0469(1974)031<0233:RSOCAI>2.0.CO;2.

Chávez, R., A. Moreira-Muñoz, M. Galleguillos, M. Olea, J. Aguayo, A. Latín, I. Aguilera-Betti, A. Muñoz, and H. Manríquez (2019). "GIMMS NDVI time series reveal the extent, duration, and intensity of "blooming desert" events in the hyper-arid Atacama Desert, Northern Chile." In: *International Journal of Applied Earth Observation and Geoinformation* 76, pp. 193–203. ISSN: 0303-2434. DOI: https://doi.org/10.1016/j.jag.2018.11.013.

Chollet, F., J. Allaire, et al. (2017). *R Interface to Keras*. https://github.com/rstudio/keras.

Chollet, F. et al. (2015). *Keras*. https://keras.io.

Clarke, A. J. (Oct. 1994). "Why Are Surface Equatorial ENSO Winds Anomalously Westerly under Anomalous Large-Scale Convection?" In: *Journal of Climate* 7.10, pp. 1623–1627. ISSN: 0894-8755. DOI: 10.1175/1520-0442(1994)007<1623:WASEEW>2.0.CO;2.

Clarke, A. J., S. Van Gorder, and G. Colantuono (Apr. 2007). "Wind Stress Curl and ENSO Discharge/Recharge in the Equatorial Pacific." In: *Journal of Physical Oceanography* 37.4, pp. 1077–1091. ISSN: 0022-3670. DOI: 10.1175/JPO3035.1.

Cockell, C. S., C. P. McKay, K. Warren-Rhodes, and G. Horneck (2008). "Ultraviolet radiation-induced limitation to epilithic microbial growth in arid deserts – Dosimetric experiments in the hyperarid core of the Atacama Desert." In: *Journal of Photochemistry and Photobiology B: Biology* 90.2, pp. 79–87. ISSN: 1011-1344. DOI: https://doi.org/10.1016/j.jphotobiol.2007.11.009.

Compo, G. P. et al. (2011). "The Twentieth Century Reanalysis Project." In: *Quarterly Journal of the Royal Meteorological Society* 137.654, pp. 1–28. DOI: 10.1002/qj.776.

Cordero, R. et al. (Mar. 2016). "The Solar Spectrum in the Atacama Desert." In: *Scientific Reports* 6, p. 22457. DOI: 10.1038/srep22457.

Costa-Surós, M., J. Calbó, J. A. González, and C. N. Long (2014). "Comparing the cloud vertical structure derived from several methods based on radiosonde profiles and ground-based remote sensing measurements." In: *Atmospheric Measurement Techniques* 7.8, pp. 2757–2773. DOI: 10.5194/amt-7-2757-2014.

Craig, H. and L. I. Gordon (1965). "Deuterium and oxygen 18 variations in the ocean and the marine atmosphere." English. In: *Oceanographic Studies, Conference on Stable Isotopes and Paleotemperatures*. Pisa: Consiglio nazionale delle richerche, Laboratorio de geologia nucleare, pp. 9–130.

Craine, J. M. et al. (2009). "Global patterns of foliar nitrogen isotopes and their relationships with climate, mycorrhizal fungi, foliar nutrient concentrations, and nitrogen availability." In: *New Phytologist* 183.4, pp. 980–992. DOI: 10.1111/j.1469-8137.2009.02917.x.

Crits-Christoph, A., C. K. Robinson, T. Barnum, W. F. Fricke, A. F. Davila, B. Jedynak, C. P. McKay, and J. Diruggiero (2013). "Colonization patterns of soil microbial communities in the Atacama Desert." eng. In: *Microbiome* 1.1, p. 28. ISSN: 2049-2618. DOI: 10.1186/2049-2618-1-28.

Dai, A., J. Wang, P. W. Thorne, D. E. Parker, L. Haimberger, and X. L. Wang (2011). "A New Approach to Homogenize Daily Radiosonde Humidity Data." In: *Journal of Climate* 24.4, pp. 965–991. DOI: 10.1175/2010JCLI3816.1.

Davey, M., A. Brookshaw, and S. Ineson (2014). "The probability of the impact of ENSO on precipitation and near-surface temperature." In: *Climate Risk Management* 1, pp. 5–24. ISSN: 2212-0963. DOI: 10.1016/j.crm.2013.12.002.

Dee, D. P. et al. (2011a). "The ERA-Interim reanalysis: configuration and performance of the data assimilation system." In: *Quarterly Journal of the Royal Meteorological Society* 137.656, pp. 553–597. DOI: 10.1002/qj.828.

Dee, D. P. et al. (2011b). "The ERA-Interim reanalysis: configuration and performance of the data assimilation system." In: *Quarterly Journal of the Royal Meteorological Society* 137.656, pp. 553–597. DOI: 10.1002/qj.828.

Derber, J. C., D. F. Parrish, and S. J. Lord (Dec. 1991). "The New Global Operational Analysis System at the National Meteorological Center." In: *Weather and Forecasting* 6.4, pp. 538–547. ISSN: 0882-8156. DOI: 10.1175/1520-0434(1991)006<0538:TNGOAS>2.0.CO;2.

Deser, C., M. A. Alexander, S.-P. Xie, and A. S. Phillips (2010). "Sea Surface Temperature Variability: Patterns and Mechanisms." In: *Annual Review of Marine Science* 2.1. PMID: 21141660, pp. 115–143. DOI: 10.1146/annurev-marine-120408-151453.

Deser, C., A. S. Phillips, and J. W. Hurrell (Aug. 2004). "Pacific Interdecadal Climate Variability: Linkages between the Tropics and the North Pacific during Boreal Winter since 1900." In: *Journal of Climate* 17.16, pp. 3109–3124. ISSN: 0894-8755. DOI: 10.1175/1520-0442(2004)017<3109:PICVLB>2.0.CO;2.

Deser, C., K. Trenberth, and N. C. for Atmospheric Research Staff (Eds)" (2016). *The Climate Data Guide: Pacific Decadal Oscillation (PDO): Definition and Indices*. Last modified 06 Jan 2016.

Desmons, M., N. Ferlay, F. Parol, L. Mcharek, and C. Vanbauce (2013). "Improved information about the vertical location and extent of monolayer clouds from POLDER3 measurements in the oxygen A-band." In: *Atmospheric Measurement Techniques* 6.8, pp. 2221–2238. DOI: 10.5194/amt-6-2221-2013.

Díaz, F. P., M. Frugone, R. A. Gutiérrez, and C. Latorre (2016). "Nitrogen cycling in an extreme hyperarid environment inferred from δ15N analyses of plants, soils and herbivore diet." In: *Scientific Reports* 6.1, p. 22226. ISSN: 2045-2322. DOI: 10.1038/srep22226.

Díaz, F. P., C. Latorre, A. Maldonado, J. Quade, and J. L. Betancourt (2012). "Rodent middens reveal episodic, long-distance plant colonizations across the hyperarid Atacama Desert over the last 34,000 years." In: *Journal of Biogeography* 39.3, pp. 510–525. DOI: 10.1111/j.1365-2699.2011.02617.x.

Diederich, J. L. et al. (2020). "A 68 ka precipitation record from the hyperarid core of the Atacama Desert in northern Chile." In: *Global and Planetary Change* 184, p. 103054. ISSN: 0921-8181. DOI: https://doi.org/10.1016/j.gloplacha.2019.103054.

Diner, D. (2012). "MISR Level 2 Cloud Heights and Winds HDF-EOS File - Version 1." In: *NASA Langley Atmospheric Science Data Center DAAC*. DOI: 10.5067/Terra/MISR/MIL2TCSP_L2.001.

Dunai, T. J., M. Melles, D. Quandt, C. Knief, and W. Amelung (2020). "Whitepaper: Earth – Evolution at the dry limit." In: *Global and Planetary Change* 193, p. 103275. ISSN: 0921-8181. DOI: https://doi.org/10.1016/j.gloplacha.2020.103275.

Durre, I., Y. Xungang, R. S. Vose, S. Applequist, and J. Arnfield (2016). *Integrated Global Radiosonde Archive (IGRA), Version 2*. Accessed: 2020-09-24. NOAA National Centers for Environmental Information. DOI: https://doi.org/10.7289/V5X63K0Q.

Eastman, R. and S. G. Warren (Feb. 2013). "A 39-Yr Survey of Cloud Changes from Land Stations Worldwide 1971–2009: Long-Term Trends, Relation to Aerosols, and Expansion of the Tropical Belt." In: *Journal of Climate* 26.4, pp. 1286–1303. ISSN: 0894-8755. DOI: 10.1175/JCLI-D-12-00280.1.

Egli, S., B. Thies, J. Drönner, J. Cermak, and J. Bendix (2017). "A 10 year fog and low stratus climatology for Europe based on Meteosat Second Generation data." In: *Quarterly Journal of the Royal Meteorological Society* 143.702, pp. 530–541. DOI: 10.1002/qj.2941.

Egli, S., B. Thies, and J. Bendix (2018). "A Hybrid Approach for Fog Retrieval Based on a Combination of Satellite and Ground Truth Data." In: *Remote Sensing* 10.4. ISSN: 2072-4292. DOI: 10.3390/rs10040628.

Ehlers, T. A. and C. J. Poulsen (2009). "Influence of Andean uplift on climate and paleoaltimetry estimates." In: *Earth and Planetary Science Letters* 281.3, pp. 238–248. ISSN: 0012-821X. DOI: 10.1016/j.epsl.2009.02.026.

Elliott, W. P. and D. J. Gaffen (1991). "On the Utility of Radiosonde Humidity Archives for climate studies." In: *Bulletin of the American Meteorological Society* 72.10, pp. 1507–1520. DOI: 10.1175/1520-0477(1991)072<1507:OTUORH>2.0.CO;2.

Ellrod, G. P. (Sept. 1995). "Advances in the Detection and Analysis of Fog at Night Using GOES Multispectral Infrared Imagery." In: *Weather and Forecasting* 10.3, pp. 606–619. ISSN: 0882-8156. DOI: 10.1175/1520-0434(1995)010<0606:AITDAA>2.0.CO;2.

Falvey, M. and R. D. Garreaud (2009). "Regional cooling in a warming world: Recent temperature trends in the southeast Pacific and

along the west coast of subtropical South America (1979–2006)." In: *Journal of Geophysical Research: Atmospheres* 114.D4. DOI: 10.1029/2008JD010519.

Farías, M. (Mar. 2007). *Estudio Biogeográfico de Tillandsiales de la Región de Tarapacá, Chile y sus Relaciones con el Comportamiento de la Niebla.* published on ResearchGate in 2015. PROYECTO FONDECYT 1051035. DOI: https://doi.org/10.13140/RG.2.1.4975.9201.

Farías, M., P. Cereceda Troncoso, P. Osses Mcintyre, and R. Núñez (Apr. 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." es. In: *Investigaciones Geográficas*, pp. 43 –61. ISSN: 0188-4611.

Farías, M., M. Lagos, P. Cereceda, H. Larrain, P. Osses, R. Pinto, and R. Núñez (2001). "Metodología para el análisis del comportamiento espacio-temporal de nubes estratocúmulos mediante percepción remota, con énfasis en su penetración en el desierto costero de Tarapacá." In: *XXII Congreso Nacional y VII Internacional de Geografía*, pp. 35–43.

Farr, T. G. et al. (2007). "The Shuttle Radar Topography Mission." In: *Reviews of Geophysics* 45.2. DOI: 10.1029/2005RG000183.

Fawcett, T. (2006). "An introduction to ROC analysis." In: *Pattern Recognition Letters* 27.8. ROC Analysis in Pattern Recognition, pp. 861–874. ISSN: 0167-8655. DOI: https://doi.org/10.1016/j.patrec.2005.10.010.

Fennig, K., A. Andersson, S. Bakan, C.-P. Klepp, and M. Schröder (2012). *Hamburg Ocean Atmosphere Parameters and Fluxes from Satellite Data - HOAPS 3.2 - Monthly Means / 6-Hourly Composites*. DOI: 10.5676/EUM_SAF_CM/HOAPS/V001.

Ferlay, N., F. Thieuleux, C. Cornet, A. B. Davis, P. Dubuisson, F. Ducos, F. Parol, J. Riédi, and C. Vanbauce (2010). "Toward New Inferences about Cloud Structures from Multidirectional Measurements in the Oxygen A Band: Middle-of-Cloud Pressure and Cloud Geometrical Thickness from POLDER-3/PARASOL." In: *J. Appl. Meteorol. Clim.* 49, pp. 2492–2507.

Folland, C. K. and D. E. Parker (1990). "Observed Variations of Sea Surface Temperature." In: *Climate-Ocean Interaction*. Dordrecht: Springer Netherlands, pp. 21–52. ISBN: 978-94-009-2093-4. DOI: 10.1007/978-94-009-2093-4_2.

Folland, C. K. and D. E. Parker (1995). "Correction of instrumental biases in historical sea surface temperature data." In: *Quarterly Journal of the Royal Meteorological Society* 121.522, pp. 319–367. DOI: 10.1002/qj.49712152206.

Gao, B.-C. and Y. J. Kaufman (2003). "Water vapor retrievals using Moderate Resolution Imaging Spectroradiometer (MODIS) near-infrared channels." In: *Journal of Geophysical Research: Atmospheres* 108.D13. DOI: 10.1029/2002JD003023.

Gardner, M. and S. Dorling (1998). "Artificial neural networks (the multilayer perceptron)–a review of applications in the atmospheric sciences." In: *Atmospheric Environment* 32.14, pp. 2627–2636. ISSN: 1352-2310. DOI: https://doi.org/10.1016/S1352-2310(97)00447-0.

Garreaud, R. D. (1999). "Multiscale Analysis of the Summertime Precipitation over the Central Andes." In: *Monthly Weather Review* 127.5, pp. 901–921. DOI: 10.1175/1520-0493(1999)127<0901:MAOTSP>2.0.CO;2.

Garreaud, R. D. and P. Aceituno (June 2001). "Interannual Rainfall Variability over the South American Altiplano." In: *Journal of Climate* 14.12, pp. 2779–2789. ISSN: 0894-8755. DOI: 10.1175/1520-0442(2001)014<2779:IRVOTS>2.0.CO;2.

Garreaud, R. D., A. Molina, and M. Farias (2010). "Andean uplift, ocean cooling and Atacama hyperaridity: A climate modeling perspective." In: *Earth and Planetary Science Letters* 292.1, pp. 39–50. ISSN: 0012-821X. DOI: https://doi.org/10.1016/j.epsl.2010.01.017.

Garreaud, R. D. and R. Muñoz (Apr. 2004). "The Diurnal Cycle in Circulation and Cloudiness over the Subtropical Southeast Pacific: A Modeling Study." In: *Journal of Climate* 17.8, pp. 1699–1710. ISSN: 0894-8755. DOI: 10.1175/1520-0442(2004)017<1699:TDCICA>2.0.CO;2.

Garreaud, R. D., M. Vuille, R. Compagnucci, and J. Marengo (2009). "Present-day South American climate." In: *Palaeogeography, Palaeoclimatology, Palaeoecology* 281.3. Long-term multi-proxy climate reconstructions and dynamics in South America (LOTRED-SA): State of the art and perspectives, pp. 180–195. ISSN: 0031-0182. DOI: https://doi.org/10.1016/j.palaeo.2007.10.032.

Garreaud, R. (Oct. 2013). "Warm Winter Storms in Central Chile." In: *Journal of Hydrometeorology* 14.5, pp. 1515–1534. ISSN: 1525-755X. DOI: 10.1175/JHM-D-12-0135.1.

Garreaud, R., M. Vuille, and A. Clement (May 2003). "The climate of the Altiplano: Observed current conditions and mechanisms of past changes." English (US). In: *Palaeogeography, Palaeoclimatology, Palaeoecology* 194.1-3, pp. 5–22. ISSN: 0031-0182. DOI: 10.1016/S0031-0182(03)00269-4.

Gaurav, S. and P. Jindal (2018). "RADIATIVE TRANSFER MODEL SIMULATIONS TO DETERMINE THE NIGHT TIME FOG DETECTION THRESHOLD." In: *ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences* XLII-5, pp. 511–517. DOI: 10.5194/isprs-archives-XLII-5-511-2018.

Gayo, E. M., C. Latorre, T. E. Jordan, P. L. Nester, S. A. Estay, K. F. Ojeda, and C. M. Santoro (2012). "Late Quaternary hydrological and ecological changes in the hyperarid core of the northern Atacama Desert ($\sim$ 21°S)." In: *Earth-Science Reviews* 113.3, pp. 120–140. ISSN:

0012-8252. DOI: https://doi.org/10.1016/j.earscirev.2012.04.003.

Gázquez, F., M. Morellón, T. Bauska, D. Herwartz, J. Surma, A. Moreno, M. Staubwasser, B. Valero-Garcés, A. Delgado-Huertas, and D. A. Hodell (2018). "Triple oxygen and hydrogen isotopes of gypsum hydration water for quantitative paleo-humidity reconstruction." In: *Earth and Planetary Science Letters* 481, pp. 177–188. ISSN: 0012-821X. DOI: 10.1016/j.epsl.2017.10.020.

Gimeno, L., R. Nieto, M. Vázquez, and D. Lavers (2014). "Atmospheric rivers: a mini-review." In: *Frontiers in Earth Science* 2, p. 2. ISSN: 2296-6463. DOI: 10.3389/feart.2014.00002.

Giovanelli, R. et al. (2001). "The Optical/Infrared Astronomical Quality of High Atacama Sites. II. Infrared Characteristics." In: *Publications of the Astronomical Society of the Pacific* 113.785, pp. 803–813. DOI: 10.1086/322136.

González, A. L., J. M. Fariña, R. Pinto, C. Pérez, K. C. Weathers, J. J. Armesto, and P. A. Marquet (2011). "Bromeliad growth and stoichiometry: responses to atmospheric nutrient supply in fog-dependent ecosystems of the hyper-arid Atacama Desert, Chile." In: *Oecologia* 167.3, pp. 835–845. ISSN: 1432-1939. DOI: 10.1007/s00442-011-2032-y.

Goren, T., D. Rosenfeld, O. Sourdeval, and J. Quaas (2018). "Satellite observations of precipitating marine stratocumulus show greater cloud fraction for decoupled clouds in comparison to coupled clouds." In: *Gephys. Res. Lett.* 45.15378722, pp. 5126–5134. DOI: 10.1029/2018GL078122.

Guan, B. (2020). *ARcatalog*. https://ucla.app.box.com/v/ARcatalog. Accessed: 2020-07-21.

Guan, B. and D. E. Waliser (2015). "Detection of atmospheric rivers: Evaluation and application of an algorithm for global studies." In: *Journal of Geophysical Research: Atmospheres* 120.24, pp. 12514–12535. DOI: 10.1002/2015JD024257.

Guan, B., D. E. Waliser, and F. M. Ralph (Feb. 2018). "An Intercomparison between Reanalysis and Dropsonde Observations of the Total Water Vapor Transport in Individual Atmospheric Rivers." In: *Journal of Hydrometeorology* 19.2, pp. 321–337. ISSN: 1525-755X. DOI: 10.1175/JHM-D-17-0114.1.

Güls, I and J Bendix (1996). "Fog detection and fog mapping using low cost Meteosat-WEFAX transmission." In: *Meteorological Applications* 3.2, pp. 179–187. DOI: 10.1002/met.5060030208.

Güsten, R., L. Å. Nyman, P. Schilke, K. Menten, C. Cesarsky, and R. Booth (2006). "The Atacama Pathfinder EXperiment (APEX) – a new submillimeter facility for southern skies –." In: *Astronomy & Astrophysics* 454.2, pp. L13–L16. DOI: 10.1051/0004-6361:20065420.

Haeffelin, M., S. Crewell, A. J. Illingworth, G. Pappalardo, H. Russchenberg, M. Chiriaco, K. Ebell, R. J. Hogan, and F. Madonna (2016).

"Parallel Developments and Formal Collaboration between European Atmospheric Profiling Observatories and the U.S. ARM Research Program." In: *Meteorological Monographs* 57, pp. 29.1–29.34. DOI: 10.1175/AMSMONOGRAPHS-D-15-0045.1.

Hahn, C. J. and S. G. Warren (2007). *A gridded climatology of clouds over land (1971–96) and ocean (1954–97) from surface observations worldwide. Numeric Data Package NDP-026E ORNL/CDIAC-153*. Tech. rep. CDIAC, Department of Energy, Oak Ridge, TN. DOI: 10.3334/CDIAC/cli.ndp026e.

Handley, L. L., A. T. Austin, G. R. Stewart, D. Robinson, C. M. Scrimgeour, J. A. Raven, T. H. E. Heaton, and S. Schmidt (1999). "The 15N natural abundance (δ15N) of ecosystem samples reflects measures of water availability." In: *Functional Plant Biology* 26.2, pp. 185–199. DOI: 10.1071/PP98146.

Hannay, C., D. L. Williamson, J. J. Hack, J. T. Kiehl, J. G. Olson, S. A. Klein, C. S. Bretherton, and M. Köhler (2009). "Evaluation of Forecasted Southeast Pacific Stratocumulus in the NCAR, GFDL, and ECMWF Models." In: *Journal of Climate* 22.11, pp. 2871–2889. DOI: 10.1175/2008JCLI2479.1.

Haug, E. W., E. R. Kraal, J. O. Sewall, M. Van Dijk, and G. C. Diaz (2010). "Climatic and geomorphic interactions on alluvial fans in the Atacama Desert, Chile." In: *Geomorphology* 121.3, pp. 184–196. ISSN: 0169-555X. DOI: https://doi.org/10.1016/j.geomorph.2010.04.005.

He, J. and Z. Liu (2019). "Comparison of Satellite-Derived Precipitable Water Vapor Through Near-Infrared Remote Sensing Channels." In: *IEEE Transactions on Geoscience and Remote Sensing* 57.12, pp. 10252–10262. ISSN: 1558-0644. DOI: 10.1109/TGRS.2019.2932847.

Heidke, P. (1926). "Berechnung des Erfolges und der Güte der Windstärkevorhersagen im Sturmwarnungsdienst." In: *Geografiska Annaler* 8.4, pp. 301–349.

Hersbach, H., P. Poli, and D. Dee (May 2015). *The observation feedback archive for the ICOADS and ISPD data sets*. Tech. rep. 18. Shinfield Park, Reading: ECMWF, p. 31.

Hersbach, H. et al. (2020). "The ERA5 global reanalysis." In: *Quarterly Journal of the Royal Meteorological Society* 146.730, pp. 1999–2049. DOI: 10.1002/qj.3803.

Hoffmeister, D. (2017a). *Meteorological and soil measurements of the permanent master weather station 13 – Cerros de Calate, Chile*. [Accessed 27. October 2020]. DOI: https://doi.org/10.5880/CRC1211DB.4.

Hoffmeister, D. (2017b). *Meteorological and soil measurements of the permanent weather stations in the Atacama desert, Chile*. [Accessed 26. June 2020]. DOI: http://dx.doi.org/10.5880/CRC1211DB.1.

Hollinger, J. P., J. L. Peirce, and G. A. Poe (1990). "SSM/I instrument evaluation." In: *IEEE Transactions on Geoscience and Remote Sensing* 28.5, pp. 781–790. ISSN: 0196-2892. DOI: 10.1109/36.58964.

Houston, J. (2006). "Variability of precipitation in the Atacama Desert: its causes and hydrological impact." In: *International Journal of Climatology* 26.15, pp. 2181–2198. DOI: 10.1002/joc.1359.

Houston, J. and A. J. Hartley (2003). "The central Andean west-slope rainshadow and its potential contribution to the origin of hyper-aridity in the Atacama Desert." In: *International Journal of Climatology* 23.12, pp. 1453–1464. DOI: 10.1002/joc.938.

Hunt, G. E. (1973). "Radiative properties of terrestrial clouds at visible and infra-red thermal window wavelengths." In: *Quarterly Journal of the Royal Meteorological Society* 99.420, pp. 346–369. DOI: 10.1002/qj.49709942013.

Hyvärinen, O. (Feb. 2014). "A Probabilistic Derivation of Heidke Skill Score." In: *Weather and Forecasting* 29.1, pp. 177–181. ISSN: 0882-8156. DOI: 10.1175/WAF-D-13-00103.1.

Illingworth, A. J. et al. (May 2019). "How Can Existing Ground-Based Profiling Instruments Improve European Weather Forecasts?" In: *Bulletin of the American Meteorological Society* 100.4, pp. 605–619. ISSN: 0003–0007. DOI: 10.1175/BAMS-D-17-0231.1.

Jacques-Coper, M., M. Falvey, and R. C. Muñoz (2015). "Inter-daily variability of a strong thermally-driven wind system over the Atacama Desert of South America: synoptic forcing and short-term predictability using the GFS global model." In: *Theoretical and Applied Climatology* 121.1, pp. 211–223. ISSN: 1434-4483. DOI: 10.1007/s00704-014-1231-y.

Jaeschke, A., C. Böhm, F. F. Merklinger, S. M. Bernasconi, M. Reyers, S. Kusch, and J. Rethemeyer (2019). "Variation in d15N of fog-dependent Tillandsia ecosystems reflect water availability across climate gradients in the hyperarid Atacama Desert." In: *Global and Planetary Change* 183, p. 103029. ISSN: 0921-8181. DOI: 10.1016/j.gloplacha.2019.103029.

Johnson, M. A. and J. J. O'Brien (1990). "The role of coastal Kelvin waves on the northeast Pacific Ocean." In: *Journal of Marine Systems* 1.1, pp. 29–38. ISSN: 0924-7963. DOI: https://doi.org/10.1016/0924-7963(90)90085-O.

Jones, D. L., S. Olivera-Ardid, E. Klumpp, C. Knief, P. W. Hill, E. Lehndorff, and R. Bol (2018). "Moisture activation and carbon use efficiency of soil microbial communities along an aridity gradient in the Atacama Desert." In: *Soil Biology and Biochemistry* 117, pp. 68–71. ISSN: 0038-0717. DOI: https://doi.org/10.1016/j.soilbio.2017.10.026.

Jordan, T. E., C. H. L., L. V. Godfrey, S. J. Colucci, C. G. P., J. U. M., G. G. L., and J. F. Paul (Jan. 2019). "Isotopic characteristics and paleoclimate implications of the extreme precipitation event of March 2015 in northern Chile." In: *Andean Geology* 46.1, pp. 1–31. DOI: https://doi.org/10.5027/andgeoV46n1-3087.

Jung, P. et al. (2020). "Desert breath–How fog promotes a novel type of soil biocenosis, forming the coastal Atacama Desert's living skin." In: *Geobiology* 18.1, pp. 113–124. DOI: 10.1111/gbi.12368.

Kayano, M. T. and R. V. Andreoli (2007). "Relations of South American summer rainfall interannual variations with the Pacific Decadal Oscillation." In: *International Journal of Climatology* 27.4, pp. 531–540. DOI: 10.1002/joc.1417.

Kerber, F., R. R. Querel, R. Rondanelli, R. Hanuschik, M. van den Ancker, O. Cuevas, A. Smette, J. Smoker, T. Rose, and H. Czekala (Feb. 2014). "An episode of extremely low precipitable water vapour over Paranal observatory." In: *Monthly Notices of the Royal Astronomical Society* 439.1, pp. 247–255. ISSN: 0035-8711. DOI: 10.1093/mnras/stt2404.

Khaniani, A. S., Z. Nikraftar, and S. Zakeri (2020). "Evaluation of MODIS Near-IR water vapor product over Iran using ground-based GPS measurements." In: *Atmospheric Research* 231, p. 104657. ISSN: 0169-8095. DOI: https://doi.org/10.1016/j.atmosres.2019.104657.

Kingma, D. P. and J. Ba (2014). *Adam: A Method for Stochastic Optimization*.

Klein, S. A. and D. L. Hartmann (Aug. 1993). "The Seasonal Cycle of Low Stratiform Clouds." In: *Journal of Climate* 6.8, pp. 1587–1606. ISSN: 0894-8755. DOI: 10.1175/1520-0442(1993)006<1587:TSCOLS>2.0.CO;2.

Knief, C. et al. (2020). "Tracing elevational changes in microbial life and organic carbon sources in soils of the Atacama Desert." In: *Global and Planetary Change* 184, p. 103078. ISSN: 0921-8181. DOI: https://doi.org/10.1016/j.gloplacha.2019.103078.

Koch, M., D. Kleinpeter, E. Auer, A. Siegmund, C. del Río, P. Osses, J. García, M. Marzol, G. Zizka, and C. Kiefer (Nov. 2019). "Living at the dry limits: ecological genetics of Tillandsia landbeckii lomas in the Chilean Atacama Desert." In: *Plant Systematics and Evolution* 305, pp. 1041–1053. DOI: https://doi.org/10.1007/s00606-019-01623-0.

Kohonen, T. (2001). *Self-Organizing Maps*. 3rd ed. Vol. 30. Springer Series in Information Sciences. Springer-Verlag Berlin Heidelberg. DOI: 10.1007/978-3-642-56927-2.

Kug, J.-S., F.-F. Jin, and S.-I. An (Mar. 2009). "Two Types of El Niño Events: Cold Tongue El Niño and Warm Pool El Niño." In: *Journal of Climate* 22.6, pp. 1499–1515. ISSN: 0894-8755. DOI: 10.1175/2008JCLI2624.1.

Laloyaux, P. et al. (2018). "CERA-20C: A Coupled Reanalysis of the Twentieth Century." In: *Journal of Advances in Modeling Earth Systems* 10.5, pp. 1172–1195. DOI: 10.1029/2018MS001273.

Lamy, F., R. Kilian, H. W. Arz, J.-P. Francois, J. Kaiser, M. Prange, and T. Steinke (2010). "Holocene changes in the position and intensity of

the southern westerly wind belt." In: *Nature Geoscience* 3.10, pp. 695–699. ISSN: 1752-0908. DOI: 10.1038/ngeo959.

Larraín, H, F Velásquez, P Cereceda, R Espejo, R Pinto, P Osses, and R. Schemenauer (2002). "Fog measurements at the site "Falda Verde" north of Chañaral compared with other fog stations of Chile." In: *Atmospheric Research* 64.1. 2nd International Conference on Fog and Fog Collection, pp. 273–284. ISSN: 0169-8095. DOI: https://doi.org/10.1016/S0169-8095(02)00098-4.

Lary, D. J., A. H. Alavi, A. H. Gandomi, and A. L. Walker (2016). "Machine learning in geosciences and remote sensing." In: *Geoscience Frontiers* 7.1. Special Issue: Progress of Machine Learning in Geosciences, pp. 3–10. ISSN: 1674-9871. DOI: https://doi.org/10.1016/j.gsf.2015.07.003.

Latorre, C., J. L. Betancourt, K. A. Rylander, and J. Quade (Mar. 2002). "Vegetation invasions into absolute desert: A 45,000 yr rodent midden record from the Calama-Salar de Atacama basins, northern Chile (lat 22°–24°S)." In: *GSA Bulletin* 114.3, pp. 349–366. ISSN: 0016-7606. DOI: 10.1130/0016-7606(2002)114<0349:VIIADA>2.0.CO;2.

Latorre, C., A. L. González, J. Quade, J. M. Fariña, R. Pinto, and P. A. Marquet (2011). "Establishment and formation of fog-dependent Tillandsia landbeckii dunes in the Atacama Desert: Evidence from radiocarbon and stable isotopes." In: *Journal of Geophysical Research: Biogeosciences* 116.G3. DOI: 10.1029/2010JG001521.

Lau, M. W., Y. L. Yung, and D. L. Wu (2012). "Determining Cloud Base and Thickness from Spaceborne Stereoscopic Imaging and Lidar Profiling Techniques." In: *Accepted by Caltech Undergraduate Research Journal* Spring Issue.

Lee, T. and M. J. McPhaden (2010). "Increasing intensity of El Niño in the central-equatorial Pacific." In: *Geophysical Research Letters* 37.14. DOI: 10.1029/2010GL044007.

Lehnert, L. W., P. Jung, W. A. Obermeier, B. Büdel, and J. Bendix (2018a). "Estimating Net Photosynthesis of Biological Soil Crusts in the Atacama Using Hyperspectral Remote Sensing." In: *Remote Sensing* 10.6. ISSN: 2072-4292. DOI: 10.3390/rs10060891.

Lehnert, L. W. et al. (2018b). "A Case Study on Fog/Low Stratus Occurrence at Las Lomitas, Atacama Desert (Chile) as a Water Source for Biological Soil Crusts." In: *Aerosol and Air Quality Research* 18.1, pp. 254–269. DOI: 10.4209/aaqr.2017.01.0021.

Lenters, J. D. and K. H. Cook (Mar. 1997). "On the Origin of the Bolivian High and Related Circulation Features of the South American Climate." In: *Journal of the Atmospheric Sciences* 54.5, pp. 656–678. ISSN: 0022-4928. DOI: 10.1175/1520-0469(1997)054<0656:OTOOTB>2.0.CO;2.

Levine, A. F. Z. and M. J. McPhaden (2015). "The annual cycle in ENSO growth rate as a cause of the spring predictability barrier."

In: *Geophysical Research Letters* 42.12, pp. 5034–5041. DOI: 10.1002/2015GL064309.

Lilly, D. K. (1968). "Models of cloud-topped mixed layers under a strong inversion." In: *Quarterly Journal of the Royal Meteorological Society* 94.401, pp. 292–309. DOI: 10.1002/qj.49709440106.

Liu, Z. and M. Alexander (2007). "Atmospheric bridge, oceanic tunnel, and global climatic teleconnections." In: *Reviews of Geophysics* 45.2. DOI: 10.1029/2005RG000172.

Lobos Roco, F., J. V.-G. de Arellano, and X. Pedruzo-Bagazgoitia (2018). "Characterizing the influence of the marine stratocumulus cloud on the land fog at the Atacama Desert." In: *Atmospheric Research* 214, pp. 109–120. ISSN: 0169-8095. DOI: 10.1016/j.atmosres.2018.07.009.

Locarnini, R. A. et al. (2013). *World ocean atlas 2013. Volume 1, Temperature*. eng. Atlas.

Lübbecke, J. F. and M. J. McPhaden (Aug. 2013). "A Comparative Stability Analysis of Atlantic and Pacific Niño Modes." In: *Journal of Climate* 26.16, pp. 5965–5980. ISSN: 0894-8755. DOI: 10.1175/JCLI-D-12-00758.1.

MODIS Characterization Support Team (MCST) (2017a). *MODIS 1km Calibrated Radiances Product*. NASA MODIS Adaptive Processing System, Goddard Space Flight Center, USA. DOI: http://dx.doi.org/10.5067/MODIS/MOD021KM.061.

MODIS Characterization Support Team (MCST) (2017b). *MODIS 1km Calibrated Radiances Product*. NASA MODIS Adaptive Processing System, Goddard Space Flight Center, USA. DOI: http://dx.doi.org/10.5067/MODIS/MYD021KM.061.

MODIS Characterization Support Team (MCST) (2017c). *MODIS Geolocation Fields Product*. NASA MODIS Adaptive Processing System, Goddard Space Flight Center, USA. DOI: http://dx.doi.org/10.5067/MODIS/MOD03.061.

MODIS Characterization Support Team (MCST) (2017d). *MODIS Geolocation Fields Product*. NASA MODIS Adaptive Processing System, Goddard Space Flight Center, USA. DOI: http://dx.doi.org/10.5067/MODIS/MYD03.061.

Mantua, N. J. and S. R. Hare (2002). "The Pacific Decadal Oscillation." In: *Journal of Oceanography* 58.1, pp. 35–44. ISSN: 1573-868X. DOI: 10.1023/A:1015820616384.

Mantua, N. J., S. R. Hare, Y. Zhang, J. M. Wallace, and R. C. Francis (1997). "A Pacific Interdecadal Climate Oscillation with Impacts on Salmon Production." In: *Bulletin of the American Meteorological Society* 78.6, pp. 1069–1080. DOI: 10.1175/1520-0477(1997)078<1069:APICOW>2.0.CO;2.

Marchand, R. T., T. P. Ackerman, and C. Moroney (2007). "An assessment of Multiangle Imaging Spectroradiometer (MISR) stereo-derived cloud top heights and cloud top winds using ground-based

radar, lidar, and microwave radiometers." In: *Journal of Geophysical Research: Atmospheres* 112.D6. DOI: 10.1029/2006JD007091.

Marín, J. C. and B. S. Barrett (2017). "Seasonal and intraseasonal variability of precipitable water vapour in the Chajnantor plateau, Chile." In: *International Journal of Climatology* 37.S1, pp. 958–971. DOI: 10.1002/joc.5049.

May, M. and D. Hoffmeister (2018). *Hourly time-lapse images at weather station 13 - Cerros de Calate, Chile from April 2017 to September 2017*. [Accessed 20. October 2020].

McHugh, T. A., E. Morrissey, S. Reed, B. Hungate, and E. Schwartz (Sept. 2015). "Water from air: An overlooked source of moisture in arid and semiarid regions." In: *Scientific Reports* 5.13767. DOI: https://doi.org/10.1038/srep13767.

McIntyre, P. E. O., M. F. Salvador, R. N. Cárdenas, P. C. Troncoso, and H. L. Barros (2005). "Coastal Fog, Satellite Imagery, and Drinking Water: Student Fieldwork in the Atacama Desert." In: *Geocarto International* 20.1, pp. 69–74. DOI: 10.1080/10106040508542338.

Meerkötter, R. and T. Zinner (2007). "Satellite remote sensing of cloud base height for convective cloud fields: A case study." In: *Geophysical Research Letters* 34.17. L17805, n/a–n/a. ISSN: 1944-8007. DOI: 10.1029/2007GL030347.

Menzel, W. P., R. A. Frey, H. Zhang, D. P. Wylie, C. C. Moeller, R. E. Holz, B. Maddux, B. A. Baum, K. I. Strabala, and L. E. Gumley (Apr. 2008). "MODIS Global Cloud-Top Pressure and Amount Estimation: Algorithm Description and Results." In: *Journal of Applied Meteorology and Climatology* 47.4, pp. 1175–1198. ISSN: 1558-8424. DOI: 10.1175/2007JAMC1705.1.

Merk, D., H. Deneke, B. Pospichal, and P. Seifert (2016). "Investigation of the adiabatic assumption for estimating cloud micro- and macrophysical properties from satellite and ground observations." In: *Atmospheric Chemistry and Physics* 16.2, pp. 933–952. DOI: 10.5194/acp-16-933-2016.

Montecino, V. and C. B. Lange (2009). "The Humboldt Current System: Ecosystem components and processes, fisheries, and sediment studies." In: *Progress in Oceanography* 83.1. Eastern Boundary Upwelling Ecosystems: Integrative and Comparative Approaches, pp. 65–79. ISSN: 0079-6611. DOI: https://doi.org/10.1016/j.pocean.2009.07.041.

Montecinos, A. and P. Aceituno (Jan. 2003). "Seasonality of the ENSO-Related Rainfall Variability in Central Chile and Associated Circulation Anomalies." In: *Journal of Climate* 16.2, pp. 281–296. ISSN: 0894-8755. DOI: 10.1175/1520-0442(2003)016<0281:SOTERR>2.0.CO;2.

Mörchen, R., E. Lehndorff, F. A. Diaz, G. Moradi, R. Bol, B. Fuentes, E. Klumpp, and W. Amelung (2019). "Carbon accrual in the Atacama Desert." In: *Global and Planetary Change* 181, p. 102993. ISSN: 0921-8181. DOI: https://doi.org/10.1016/j.gloplacha.2019.102993.

Moroney, C., R. Davies, and J. P. Muller (2002). "Operational retrieval of cloud-top heights using MISR data." In: *IEEE Transactions on Geoscience and Remote Sensing* 40.7, pp. 1532–1540. ISSN: 0196-2892. DOI: 10.1109/TGRS.2002.801150.

Moroney, C. and K. Mueller (2012). *Data Product Specification for the MISR Level 2 Cloud Product*. Tech. rep. JPL D-72327. Jet Propulsion Laboratory, California Institute of Technology.

Muñoz-Schick, M., R. Pinto, A. Mesa, and A. Moreira-Muñoz (June 2001). ""Oasis de neblina" en los cerros costeros del sur de Iquique, región de Tarapacá, Chile, durante el evento El Niño 1997-1998." es. In: *Revista chilena de historia natural* 74, pp. 389–405. ISSN: 0716-078X. DOI: 10.4067/S0716-078X2001000200014.

Muñoz, R. C., J. Quintana, M. J. Falvey, J. A. Rutllant, and R. Garreaud (June 2016). "Coastal Clouds at the Eastern Margin of the Southeast Pacific: Climatology and Trends." In: *Journal of Climate* 29.12, pp. 4525–4542. ISSN: 0894-8755. DOI: 10.1175/JCLI-D-15-0757.1.

Muñoz, R. C., R. A. Zamora, and J. A. Rutllant (2011). "The Coastal Boundary Layer at the Eastern Margin of the Southeast Pacific (23.4°S, 70.4°W): Cloudiness-Conditioned Climatology." In: *Journal of Climate* 24.4, pp. 1013–1033. DOI: 10.1175/2010JCLI3714.1.

Mueller, K., C. Moroney, V. Jovanovic, M. Garay, J.-P. Muller, L. Di Girolamo, and R. Davies (2013). *MISR Level 2 Cloud Product Algorithm Theoretical Basis*. Tech. rep. JPL D-73327. Jet Propulsion Laboratory, California Institute of Technology.

Mülmenstädt, J., O. Sourdeval, D. S. Henderson, T. S. L'Ecuyer, C. Unglaub, L. Jungandreas, C. Böhm, L. M. Russell, and J. Quaas (2018). "Using CALIOP to estimate cloud-field base height and its uncertainty: The Cloud Base Altitude Spatial Extrapolator (CBASE) algorithm and dataset." In: *Earth Syst. Sci. Data* 10, pp. 2279–2293. DOI: 10.5194/essd-10-2279-2018.

Nash, D., D. Waliser, B. Guan, H. Ye, and F. M. Ralph (2018). "The Role of Atmospheric Rivers in Extratropical and Polar Hydroclimate." In: *Journal of Geophysical Research: Atmospheres* 123.13, pp. 6804–6821. DOI: 10.1029/2017JD028130.

National Oceanic and Atmospheric Administration, Department of Defense, Federal Aviation Administration, and United States Navy (1998). *Automated Surface Observing System User's Guide*.

Ndarana, T. and D. W. Waugh (2010). "The link between cut-off lows and Rossby wave breaking in the Southern Hemisphere." In: *Quarterly Journal of the Royal Meteorological Society* 136.649, pp. 869–885. DOI: 10.1002/qj.627.

Neiman, P. J., F. M. Ralph, G. A. Wick, J. D. Lundquist, and M. D. Dettinger (Feb. 2008). "Meteorological Characteristics and Overland Precipitation Impacts of Atmospheric Rivers Affecting the West Coast of North America Based on Eight Years of SSM/I Satellite

Observations." In: *Journal of Hydrometeorology* 9.1, pp. 22–47. ISSN: 1525-755X. DOI: 10.1175/2007JHM855.1.

Newell, R. E., N. E. Newell, Y. Zhu, and C. Scott (1992). "Tropospheric rivers? – A pilot study." In: *Geophysical Research Letters* 19.24, pp. 2401–2404. DOI: 10.1029/92GL02916.

Newman, M., S.-I. Shin, and M. A. Alexander (2011). "Natural variation in ENSO flavors." In: *Geophysical Research Letters* 38.14. DOI: 10.1029/2011GL047658.

Newman, M. et al. (June 2016). "The Pacific Decadal Oscillation, Revisited." In: *Journal of Climate* 29.12, pp. 4399–4427. ISSN: 0894-8755. DOI: 10.1175/JCLI-D-15-0508.1.

Osses, P., C. Barría, M. Farías, and P. Cereceda (Jan. 2005). "La nube estratocúmulo en Tarapacá, Chile. Validación de imágenes GOES mediante observación en tiempo real (17 al 26 de julio del año 2002)." In: *Revista de Geografía, Norte Grande*, pp. 131–143.

Osses, P., R. S. Schemenauer, P. Cereceda, H. Larraín, and C. Correa (2000). "Los atrapanieblas del Santuario Padre Hurtado y sus proyecciones en el combate a la desertificación." In: *Revista de geografía Norte Grande* 000196632.27, pp. 61–67. ISSN: 0379-8682.

Otarola, A., C. D. Breuck, T. Travouillon, S. Matsushita, L.-Å. Nyman, A. Wootten, S. J. E. Radford, M. Sarazin, F. Kerber, and J. P. Pérez-Beaupuits (2019). "Precipitable Water Vapor, Temperature, and Wind Statistics At Sites Suitable for mm and Submm Wavelength Astronomy in Northern Chile." In: *Publications of the Astronomical Society of the Pacific* 131.998, p. 045001. DOI: 10.1088/1538-3873/aafb78.

Paltan, H., D. Waliser, W. H. Lim, B. Guan, D. Yamazaki, R. Pant, and S. Dadson (2017). "Global Floods and Water Availability Driven by Atmospheric Rivers." In: *Geophysical Research Letters* 44.20, pp. 10,387–10,395. DOI: 10.1002/2017GL074882.

Paluch, I. R. and D. H. Lenschow (Oct. 1991). "Stratiform Cloud Formation in the Marine Boundary Layer." In: *Journal of the Atmospheric Sciences* 48.19, pp. 2141–2158. ISSN: 0022-4928. DOI: 10.1175/1520-0469(1991)048<2141:SCFITM>2.0.CO;2.

Payne, A. E., M.-E. Demory, L. R. Leung, A. M. Ramos, C. A. Shields, J. J. Rutz, N. Siler, G. Villarini, A. Hall, and F. M. Ralph (2020). "Responses and impacts of atmospheric rivers to climate change." In: *Nature Reviews Earth & Environment* 1.3, pp. 143–157. ISSN: 2662-138X. DOI: 10.1038/s43017-020-0030-5.

Peters, D. and D. W. Waugh (Nov. 2003). "Rossby Wave Breaking in the Southern Hemisphere Wintertime Upper Troposphere." In: *Monthly Weather Review* 131.11, pp. 2623–2634. ISSN: 0027-0644. DOI: 10.1175/1520-0493(2003)131<2623:RWBITS>2.0.CO;2.

Pinto, R., I. Barría, and P. Marquet (2006). "Geographical distribution of Tillandsia lomas in the Atacama Desert, northern Chile." In: *Journal of Arid Environments* 65.4, pp. 543–552. ISSN: 0140-1963. DOI: https://doi.org/10.1016/j.jaridenv.2005.08.015.

Pizarro, J. G. and A. Montecinos (1999). "Cutoff cyclones off the subtropical coast of Chile." In: *Sixth International Conference on Southern Hemisphere Meteorology and Oceanography*. American Meteorological Society.

Platnick, S., S. Ackerman, M. King, G. Wind, K. Meyer, P. Menzel, R. Frey, R. Holz, B. Baum, and P. Yang (2017a). "MODIS atmosphere L2 cloud product (06_L2)." In: *NASA MODIS Adaptive Processing System, Goddard Space Flight Center*. DOI: https://doi.org/10.5067/MODIS/MOD06_L2.061.

Platnick, S., S. Ackerman, M. King, G. Wind, K. Meyer, P. Menzel, R. Frey, R. Holz, B. Baum, and P. Yang (2017b). "MODIS atmosphere L2 cloud product (06_L2)." In: *NASA MODIS Adaptive Processing System, Goddard Space Flight Center*. DOI: https://doi.org/10.5067/MODIS/MYD06_L2.061.

Pliscoff, P., N. Zanetta, J. Hepp, and J. Machuca (2017). "Efectos sobre la flora y vegetación del evento de precipitación extremo de agosto 2015 en Alto Patache, Desierto de Atacama, Chile." es. In: *Revista de geografía Norte Grande* 68, pp. 91–103. ISSN: 0718-3402. DOI: https://dx.doi.org/10.4067/S0718-34022017000300091.

Poli, P. et al. (2016). "ERA-20C: An Atmospheric Reanalysis of the Twentieth Century." In: *Journal of Climate* 29.11, pp. 4083–4097. DOI: 10.1175/JCLI-D-15-0556.1.

Priego, E., J. Jones, M. Porres, and A. Seco (2017). "Monitoring water vapour with GNSS during a heavy rainfall event in the Spanish Mediterranean area." In: *Geomatics, Natural Hazards and Risk* 8.2, pp. 282–294. DOI: 10.1080/19475705.2016.1201150.

Qiu, B. and S. Chen (Nov. 2005). "Variability of the Kuroshio Extension Jet, Recirculation Gyre, and Mesoscale Eddies on Decadal Time Scales." In: *Journal of Physical Oceanography* 35.11, pp. 2090–2103. ISSN: 0022-3670. DOI: 10.1175/JPO2807.1.

Rahn, D. A. and R. Garreaud (2010). "Marine boundary layer over the subtropical southeast Pacific during VOCALS-REx – Part 1: Mean structure and diurnal cycle." In: *Atmospheric Chemistry and Physics* 10.10, pp. 4491–4506. DOI: 10.5194/acp-10-4491-2010.

Ralph, F. M. and M. D. Dettinger (2011). "Storms, floods, and the science of atmospheric rivers." In: *Eos, Transactions American Geophysical Union* 92.32, pp. 265–266. DOI: 10.1029/2011EO320001.

Ralph, F. M., M. D. Dettinger, M. M. Cairns, T. J. Galarneau, and J. Eylander (May 2018). "Defining "Atmospheric River": How the Glossary of Meteorology Helped Resolve a Debate." In: *Bulletin of the American Meteorological Society* 99.4, pp. 837–839. ISSN: 0003-0007. DOI: 10.1175/BAMS-D-17-0157.1.

Ramos, A. M., R. Tomé, R. M. Trigo, M. L. R. Liberato, and J. G. Pinto (2016). "Projected changes in atmospheric rivers affecting Europe in CMIP5 models." In: *Geophysical Research Letters* 43.17, pp. 9315–9323. DOI: 10.1002/2016GL070634.

Randall, D. A. and M. J. Suarez (Oct. 1984). "On the Dynamics of Stratocumulus Formation and Dissipation." In: *Journal of the Atmospheric Sciences* 41.20, pp. 3052–3057. ISSN: 0022-4928. DOI: 10.1175/1520-0469(1984)041<3052:OTDOSF>2.0.CO;2.

Rauber, R. M., H. Hu, F. Dominguez, S. W. Nesbitt, G. M. McFarquhar, T. J. Zaremba, and J. A. Finlon (2020). "Structure of an Atmospheric River Over Australia and the Southern Ocean. Part I: Tropical and Midlatitude Water Vapor Fluxes." In: *Journal of Geophysical Research: Atmospheres* 125.18. e2020JD032513 2020JD032513, e2020JD032513. DOI: 10.1029/2020JD032513.

Rayner, N. A., D. E. Parker, E. B. Horton, C. K. Folland, L. V. Alexander, D. P. Rowell, E. C. Kent, and A. Kaplan (2003). "Global analyses of sea surface temperature, sea ice, and night marine air temperature since the late nineteenth century." In: *Journal of Geophysical Research: Atmospheres* 108.D14. DOI: 10.1029/2002JD002670.

Remy, D., M. Falvey, S. Bonvalot, M. Chlieh, G. Gabalda, J.-L. Froger, and D. Legrand (2011). "Variability of atmospheric precipitable water in northern Chile: Impacts on interpretation of InSAR data for earthquake modeling." In: *Journal of South American Earth Sciences* 31.2, pp. 214–226. ISSN: 0895-9811. DOI: https://doi.org/10.1016/j.jsames.2011.01.003.

Reyers, M. (2018). *WRF Output daily accumulated total precipitation 10km resolution Atacama*. Accessed: 2020-10-23. DOI: https://dx.doi.org/10.5880/CRC1211DB.20.

Reyers, M., C. Böhm, L. Knarr, Y. Shao, and S. Crewell (Aug. 2020). "Synoptic-to-regional scale analysis of rainfall in the Atacama Desert (18°S–26°S) using a long-term simulation with WRF." In: *Monthly Weather Review*, pp. 1–51. ISSN: 0027-0644. DOI: 10.1175/MWR-D-20-0038.1.

Reyers, M. and Y. Shao (2019). "Cutoff lows off the coast of the Atacama Desert under present day conditions and in the Last Glacial Maximum." In: *Global and Planetary Change* 181, p. 102983. ISSN: 0921-8181. DOI: https://doi.org/10.1016/j.gloplacha.2019.102983.

Río, C. del et al. (2018). "ENSO Influence on Coastal Fog-Water Yield in the Atacama Desert, Chile." In: *Aerosol and Air Quality Research* 18.1, pp. 127–144. DOI: 10.4209/aaqr.2017.01.0022.

Ritter, B. et al. (2019). "Climatic fluctuations in the hyperarid core of the Atacama Desert during the past 215 ka." In: *Scientific Reports* 9.1, p. 5270. ISSN: 2045-2322. DOI: 10.1038/s41598-019-41743-8.

Ritterbach, L. and P. Becker (2020). "Temperature and humidity dependent formation of $CaSO_4 \cdot xH_2O$ $(x = 0 \cdots 2)$ phases." In: *Global and Planetary Change* 187, p. 103132. ISSN: 0921-8181. DOI: 10.1016/j.gloplacha.2020.103132.

Rogers, D. P. and D. Koraccn (Aug. 1992). "Radiative Transfer and Turbulence in the Cloud-topped Marine Atmospheric Boundary Layer." In: *Journal of the Atmospheric Sciences* 49.16, pp. 1473–1486.

ISSN: 0022-4928. DOI: 10.1175/1520-0469(1992)049<1473:RTATIT>2.0.CO;2.

Rondanelli, R., A. Molina, and M. Falvey (2015). "The Atacama Surface Solar Maximum." In: *Bulletin of the American Meteorological Society* 96.3, pp. 405–418. DOI: 10.1175/BAMS-D-13-00175.1.

Ruhm, J., T. Böhnert, M. Weigend, F. F. Merklinger, A. Stoll, D. Quandt, and F. Luebert (May 2020). "Plant life at the dry limit—Spatial patterns of floristic diversity and composition around the hyperarid core of the Atacama Desert." In: *PLOS ONE* 15.5, pp. 1–21. DOI: 10.1371/journal.pone.0233729.

Rundel, P., B. Palma, M. Dillon, M. R. Sharifi, and K. Boonpragob (Jan. 1997). "Tillandsia landbeckii in the coastal Atacama Desert of northern Chile." In: *Revista Chilena de Historia Natural* 70, pp. 341–349.

Rutllant, J. A., R. C. Muñoz, and R. D. Garreaud (2013). "Meteorological observations on the northern Chilean coast during VOCALS-REx." In: *Atmospheric Chemistry and Physics* 13.6, pp. 3409–3422. DOI: 10.5194/acp-13-3409-2013.

Rutllant, J. A., H. Fuenzalida, and P. Aceituno (2003). "Climate dynamics along the arid northern coast of Chile: The 1997–1998 Dinámica del Clima de la Región de Antofagasta (DICLIMA) experiment." In: *Journal of Geophysical Research: Atmospheres* 108.D17. 4538, n/a–n/a. ISSN: 2156-2202. DOI: 10.1029/2002JD003357.

Saavedra, F., G. Cortés, M. Viale, S. Margulis, and J. McPhee (2020). "Atmospheric Rivers Contribution to the Snow Accumulation Over the Southern Andes (26.5°S–37.5°S)." In: *Frontiers in Earth Science* 8, p. 261. ISSN: 2296-6463. DOI: 10.3389/feart.2020.00261.

Sasaki, Y. N., S. Minobe, and N. Schneider (Mar. 2013). "Decadal Response of the Kuroshio Extension Jet to Rossby Waves: Observation and Thin-Jet Theory." In: *Journal of Physical Oceanography* 43.2, pp. 442–456. ISSN: 0022-3670. DOI: 10.1175/JPO-D-12-096.1.

Sasaki, Y. N. and N. Schneider (May 2011). "Decadal Shifts of the Kuroshio Extension Jet: Application of Thin-Jet Theory." In: *Journal of Physical Oceanography* 41.5, pp. 979–993. ISSN: 0022-3670. DOI: 10.1175/2011JPO4550.1.

Schemenauer, R. S. and P. Cereceda (Nov. 1994a). "A Proposed Standard Fog Collector for Use in High-Elevation Regions." In: *Journal of Applied Meteorology* 33.11, pp. 1313–1322. ISSN: 0894-8763. DOI: 10.1175/1520-0450(1994)033<1313:APSFCF>2.0.CO;2.

Schemenauer, R. S. and P. Cereceda (1994b). "Fog collection's role in water planning for developing countries." In: *Natural Resources Forum* 18.2, pp. 91–100. DOI: 10.1111/j.1477-8947.1994.tb00879.x.

Schröder, M. et al. (2017). "GEWEX water vapor assessment (G-VAP)." In: *World Climate Research Programme (WCRP): Geneva Switzerland* WCRP report 16/2017, 216 pp.

Schröder, M., M. Lockhoff, J. M. Forsythe, H. Q. Cronk, T. H. Vonder Haar, and R. Bennartz (2016). "The GEWEX Water Vapor Assessment: Results from Intercomparison, Trend, and Homogeneity Analysis of Total Column Water Vapor." In: *Journal of Applied Meteorology and Climatology* 55.7, pp. 1633–1649. DOI: 10.1175/JAMC-D-15-0304.1.

Schulz, N., J. P. Boisier, and P. Aceituno (2012). "Climate change along the arid coast of northern Chile." In: *International Journal of Climatology* 32.12, pp. 1803–1814. DOI: 10.1002/joc.2395.

Schween, J. H., D. Hoffmeister, and U. Löhnert (2020). "Filling the observational gap in the Atacama Desert with a new network of climate stations." In: *Global and Planetary Change* 184, p. 103034. ISSN: 0921-8181. DOI: https://doi.org/10.1016/j.gloplacha.2019.103034.

Seethala, C., J. R. Norris, and T. A. Myers (Oct. 2015). "How Has Subtropical Stratocumulus and Associated Meteorology Changed since the 1980s?" In: *Journal of Climate* 28.21, pp. 8396–8410. ISSN: 0894-8755. DOI: 10.1175/JCLI-D-15-0120.1.

Sepulchre, P., L. C. Sloan, and F. Fluteau (2011). "Modelling the Response of Amazonian Climate to the Uplift of the Andean Mountain Range." In: *Amazonia: Landscape and Species Evolution*. John Wiley & Sons, Ltd. Chap. Thirteen, pp. 211–222. ISBN: 9781444306408. DOI: 10.1002/9781444306408.ch13.

Skinner, C. B., J. M. Lora, A. E. Payne, and C. J. Poulsen (2020). "Atmospheric river changes shaped mid-latitude hydroclimate since the mid-Holocene." In: *Earth and Planetary Science Letters* 541, p. 116293. ISSN: 0012-821X. DOI: https://doi.org/10.1016/j.epsl.2020.116293.

Smith, W. L. and C. M. R. Platt (1978). "Comparison of Satellite-Deduced Cloud Heights with Indications from Radiosonde and Ground-Based Laser Measurements." In: *J. Appl. Meteorol.* 17, pp. 1796–1802. DOI: 10.1175/1520-0450(1978)017\%3C1796:COSDCH\%3E2.0.CO;2.

Soderberg, K. S. (2010). "The Role of Fog in the Ecohydrology and Biogeochemistry of the Namib Desert." PhD thesis. University of Virginia. DOI: 10.18130/V3KZ88.

Stein, A. F., R. R. Draxler, G. D. Rolph, B. J. B. Stunder, M. D. Cohen, and F. Ngan (Jan. 2016). "NOAA's HYSPLIT Atmospheric Transport and Dispersion Modeling System." In: *Bulletin of the American Meteorological Society* 96.12, pp. 2059–2077. ISSN: 0003-0007. DOI: 10.1175/BAMS-D-14-00110.1.

Steinke, S., S. Eikenberg, U. Löhnert, G. Dick, D. Klocke, P. Di Girolamo, and S. Crewell (2015). "Assessment of small-scale integrated water vapour variability during HOPE." In: *Atmospheric Chemistry and Physics* 15.5, pp. 2675–2692. DOI: 10.5194/acp-15-2675-2015.

Stevens, B. and S. Bony (June 2013). "Water in the Atmosphere." In: *Physics Today* 66, pp. 29–34. DOI: https://doi.org/10.1063/PT.3.2009.

Strabala, K. I., S. A. Ackerman, and W. P. Menzel (Feb. 1994). "Cloud Properties inferred from 8–12-μm Data." In: *Journal of Applied Meteorology* 33.2, pp. 212–229. ISSN: 0894-8763. DOI: 10.1175/1520-0450(1994)033<0212:CPIFD>2.0.CO;2.

Suen, J. Y., M. T. Fang, and P. M. Lubin (2014). "Global Distribution of Water Vapor and Cloud Cover—Sites for High-Performance THz Applications." In: *IEEE Transactions on Terahertz Science and Technology* 4.1, pp. 86–100. ISSN: 2156-342X. DOI: 10.1109/TTHZ.2013.2294018.

Sunuararajan, R. and M. Tjernström (2000). "Observations and simulations of a non-stationary coastal atmospheric boundary layer." In: *Quarterly Journal of the Royal Meteorological Society* 126.563, pp. 445–476. DOI: 10.1002/qj.49712656305.

Surma, J., S. Assonov, D. Herwartz, C. Voigt, and M. Staubwasser (2018). "The evolution of 17O-excess in surface water of the arid environment during recharge and evaporation." In: *Scientific Reports* 8.4972, pp. 1–10. DOI: 10.1038/s41598-018-23151-6.

Taguchi, B., H. Nakamura, M. Nonaka, N. Komori, A. Kuwano-Yoshida, K. Takaya, and A. Goto (Jan. 2012). "Seasonal Evolutions of Atmospheric Response to Decadal SST Anomalies in the North Pacific Subarctic Frontal Zone: Observations and a Coupled Model Simulation." In: *Journal of Climate* 25.1, pp. 111–139. ISSN: 0894-8755. DOI: 10.1175/JCLI-D-11-00046.1.

Taguchi, B., S.-P. Xie, N. Schneider, M. Nonaka, H. Sasaki, and Y. Sasai (June 2007). "Decadal Variability of the Kuroshio Extension: Observations and an Eddy-Resolving Model Hindcast." In: *Journal of Climate* 20.11, pp. 2357–2377. ISSN: 0894-8755. DOI: 10.1175/JCLI4142.1.

Tang, Y., J. Gao, C. Liu, X. Chen, and Y. Zhao (2019). "Dehydration Pathways of Gypsum and the Rehydration Mechanism of Soluble Anhydrite $\gamma - CaSO_4$." In: *ACS Omega* 4.4, pp. 7636–7642. DOI: https://doi.org/10.1021/acsomega.8b03476.

Tanimoto, Y., H. Nakamura, T. Kagimoto, and S. Yamane (2003). "An active role of extratropical sea surface temperature anomalies in determining anomalous turbulent heat flux." In: *Journal of Geophysical Research: Oceans* 108.C10. DOI: 10.1029/2002JC001750.

Taylor, S., P. Stier, B. White, S. Finkensieper, and M. Stengel (2017). "Evaluating the diurnal cycle in cloud top temperature from SEVIRI." In: *Atmospheric Chemistry and Physics* 17.11, pp. 7035–7053. DOI: 10.5194/acp-17-7035-2017.

Tedeschi, R. G., I. F. A. Cavalcanti, and A. M. Grimm (2013). "Influences of two types of ENSO on South American precipitation." In: *International Journal of Climatology* 33.6, pp. 1382–1400. DOI: 10.1002/joc.3519.

Tharwat, A. (2018). "Classification assessment methods." In: *Applied Computing and Informatics*. ISSN: 2210-8327. DOI: https://doi.org/10.1016/j.aci.2018.08.003.

Timmermann, A. et al. (July 2018). "El Niño–Southern Oscillation complexity." In: *Nature* 559, pp. 535–545. DOI: 10.1038/s41586-018-0252-6.

UNEP (2011). *Global Drylands: A UN system-wide response*. Tech. rep. United Nations Environment Management Group.

Valdés-Pineda, R., J. Cañón, and J. B. Valdés (2018). "Multi-decadal 40- to 60-year cycles of precipitation variability in Chile (South America) and their relationship to the AMO and PDO signals." In: *Journal of Hydrology* 556, pp. 1153–1170. ISSN: 0022-1694. DOI: https://doi.org/10.1016/j.jhydrol.2017.01.031.

Valdés-Pineda, R., J. B. Valdés, H. F. Diaz, and R. Pizarro-Tapia (2016). "Analysis of spatio-temporal changes in annual and seasonal precipitation variability in South America-Chile and related ocean–atmosphere circulation patterns." In: *International Journal of Climatology* 36.8, pp. 2979–3001. DOI: 10.1002/joc.4532.

Van Beusekom, A. E., G. González, and M. A. Scholl (2017). "Analyzing cloud base at local and regional scales to understand tropical montane cloud forest vulnerability to climate change." In: *Atmospheric Chemistry and Physics* 17.11, pp. 7245–7259. DOI: 10.5194/acp-17-7245-2017.

Vargas, G., J. Rutllant, and L. Ortlieb (2006). "ENSO tropical–extratropical climate teleconnections and mechanisms for Holocene debris flows along the hyperarid coast of western South America (17°–24°S)." In: *Earth and Planetary Science Letters* 249.3, pp. 467–483. ISSN: 0012-821X. DOI: https://doi.org/10.1016/j.epsl.2006.07.022.

Vecchi, G. A. and D. E. Harrison (June 2000). "Tropical Pacific Sea Surface Temperature Anomalies, El Niño, and Equatorial Westerly Wind Events." In: *Journal of Climate* 13.11, pp. 1814–1830. ISSN: 0894-8755. DOI: 10.1175/1520-0442(2000)013<1814:TPSSTA>2.0.CO;2.

Viale, M. and M. N. Nuñez (Aug. 2011). "Climatology of Winter Orographic Precipitation over the Subtropical Central Andes and Associated Synoptic and Regional Characteristics." In: *Journal of Hydrometeorology* 12.4, pp. 481–507. ISSN: 1525-755X. DOI: 10.1175/2010JHM1284.1.

Viale, M., R. Valenzuela, R. D. Garreaud, and F. M. Ralph (Oct. 2018). "Impacts of Atmospheric Rivers on Precipitation in Southern South America." In: *Journal of Hydrometeorology* 19.10, pp. 1671–1687. ISSN: 1525-755X. DOI: 10.1175/JHM-D-18-0006.1.

Voigt, C. (2020). "Tracing the water cycle in the Atacama Desert using water isotopes (δ2H, δ17O, δ18O) and pedogenic salt distributions." PhD thesis. Köln: Universität zu Köln.

Voigt, C., S. Klipsch, D. Herwartz, G. Chong, and M. Staubwasser (2020). "The spatial distribution of soluble salts in the surface soil

of the Atacama Desert and their relationship to hyperaridity." In: *Global and Planetary Change* 184, p. 103077. ISSN: 0921-8181. DOI: https://doi.org/10.1016/j.gloplacha.2019.103077.

Vuille, M. (1999). "Atmospheric circulation over the Bolivian Altiplano during dry and wet periods and extreme phases of the Southern Oscillation." In: *International Journal of Climatology* 19.14, pp. 1579–1600. DOI: 10.1002/(SICI)1097-0088(19991130)19:14<1579::AID-JOC441>3.0.CO;2-N.

Vuille, M. and C. Ammann (1997). "REGIONAL SNOWFALL PATTERNS IN THE HIGH, ARID ANDES." In: *Climatic Change* 36.3, pp. 413–423. ISSN: 1573-1480. DOI: https://doi.org/10.1023/A:1005330802974.

Waliser, D. and B. Guan (2017). "Extreme winds and precipitation during landfall of atmospheric rivers." In: *Nature Geoscience* 10.3, pp. 179–183. ISSN: 1752-0908. DOI: 10.1038/ngeo2894.

Walk, J., G. Stauch, M. Reyers, P. Vásquez, F. A. Sepúlveda, M. Bartz, D. Hoffmeister, H. Brückner, and F. Lehmkuhl (2020). "Gradients in climate, geology, and topography affecting coastal alluvial fan morphodynamics in hyperarid regions – The Atacama perspective." In: *Global and Planetary Change* 185, p. 102994. ISSN: 0921-8181. DOI: https://doi.org/10.1016/j.gloplacha.2019.102994.

Wang, S., J. Huang, Y. He, and Y. Guan (2014). "Combined effects of the Pacific Decadal Oscillation and El Niño-Southern Oscillation on Global Land Dry–Wet Changes." In: *Scientific Reports* 4.1, p. 6651. ISSN: 2045-2322. DOI: 10.1038/srep06651.

Warren-Rhodes, K. A., K. L. Rhodes, S. B. Pointing, S. A. Ewing, D. C. Lacap, B. Gómez-Silva, R. Amundson, E. I. Friedmann, and C. P. McKay (2006). "Hypolithic Cyanobacteria, Dry Limit of Photosynthesis, and Microbial Ecology in the Hyperarid Atacama Desert." In: *Microbial Ecology* 52.3, pp. 389–398. ISSN: 1432-184X. DOI: 10.1007/s00248-006-9055-7.

Wehrens, R. and L. M. C. Buydens (2007). "Self- and Super-Organizing Maps in R: The kohonen Package." In: *Journal of Statistical Software* 21.5, pp. 1–19. DOI: 10.18637/jss.v021.i05.

Wehrens, R. and J. Kruisselbrink (2018). "Flexible Self-Organizing Maps in kohonen 3.0." In: *Journal of Statistical Software* 87.7, pp. 1–18. DOI: 10.18637/jss.v087.i07.

Wengel, C., M. Latif, W. Park, J. Harlaß, and T. Bayr (2018). "Seasonal ENSO phase locking in the Kiel Climate Model: The importance of the equatorial cold sea surface temperature bias." In: *Climate Dynamics* 50.3, pp. 901–919. ISSN: 1432-0894. DOI: 10.1007/s00382-017-3648-3.

Wentz, F. J. (2013). *SSM/I Version-7 Calibration Report*. RSS Technical Report 011012.

Westbeld, A., O. Klemm, F. Grießbaum, E. Sträter, H. Larrain, P. Osses, and P. Cereceda (2009). "Fog deposition to a Tillandsia carpet in the

Atacama Desert." In: *Annales Geophysicae* 27.9, pp. 3571–3576. DOI: 10.5194/angeo-27-3571-2009.

Wilcox, A. C., C. Escauriaza, R. Agredano, E. Mignot, V. Zuazo, S. Otárola, L. Castro, J. Gironás, R. Cienfuegos, and L. Mao (2016). "An integrated analysis of the March 2015 Atacama floods." In: *Geophysical Research Letters* 43.15, pp. 8035–8043. DOI: 10.1002/2016GL069751.

Wilson, A. M. et al. (June 2020). "Training the Next Generation of Researchers in the Science and Application of Atmospheric Rivers." In: *Bulletin of the American Meteorological Society* 101.6, E738–E743. ISSN: 0003-0007. DOI: 10.1175/BAMS-D-19-0311.1.

Winker, D. M. et al. (2010). "The CALIPSO Mission." In: *Bulletin of the American Meteorological Society* 91.9, pp. 1211–1230. DOI: 10.1175/2010BAMS3009.1.

Wittenberg, A. T., A. Rosati, T. L. Delworth, G. A. Vecchi, and F. Zeng (Mar. 2014). "ENSO Modulation: Is It Decadally Predictable?" In: *Journal of Climate* 27.7, pp. 2667–2681. ISSN: 0894-8755. DOI: 10.1175/JCLI-D-13-00577.1.

Wood, R. et al. (2011). "The VAMOS Ocean-Cloud-Atmosphere-Land Study Regional Experiment (VOCALS-REx): goals, platforms, and field operations." In: *Atmospheric Chemistry and Physics* 11.2, pp. 627–654. DOI: 10.5194/acp-11-627-2011.

Wood, R. (2012). "Stratocumulus Clouds." In: *Monthly Weather Review* 140.8, pp. 2373–2423. DOI: 10.1175/MWR-D-11-00121.1.

World Meteorological Organization (2013). *Technical Regulations Volume II: Meteorological service for international air navigation*.

Wyant, M. C. et al. (2010). "The PreVOCA experiment: modeling the lower troposphere in the Southeast Pacific." In: *Atmospheric Chemistry and Physics* 10.10, pp. 4757–4774. DOI: 10.5194/acp-10-4757-2010.

Xiong, X., K. Chiang, A. Wu, W. L. Barnes, B. Guenther, and V. V. Salomonson (2008). "Multiyear On-Orbit Calibration and Performance of Terra MODIS Thermal Emissive Bands." In: *IEEE Transactions on Geoscience and Remote Sensing* 46.6, pp. 1790–1803. ISSN: 1558-0644. DOI: 10.1109/TGRS.2008.916217.

Zhang, Y., J. M. Wallace, and D. S. Battisti (May 1997). "ENSO-like Interdecadal Variability: 1900–93." In: *Journal of Climate* 10.5, pp. 1004–1020. ISSN: 0894-8755. DOI: 10.1175/1520-0442(1997)010<1004:ELIV>2.0.CO;2.

Zhu, Y. and R. E. Newell (Mar. 1998). "A Proposed Algorithm for Moisture Fluxes from Atmospheric Rivers." In: *Monthly Weather Review* 126.3, pp. 725–735. ISSN: 0027-0644. DOI: 10.1175/1520-0493(1998)126<0725:APAFMF>2.0.CO;2.

ACKNOWLEDGMENTS

I would like to thank Prof. Dr. Johannes Quaas for forwarding me the job posting for the PhD position in Cologne and for the cloud base height topic which he gave to me. The latter was probably a door opener for my engagement at the INFERNO group. Furthermore, I like to thank Dr. Odran Sourdeval and Dr. Johannes Mülmenstädt for speeding up my R game.

I would like to thank Prof. Dr. Tibor Dunai and everybody who was involved in the initiation of the CRC 1211. The past years were very inspiring and enabled me to engage in a unique science community of all flavors.

A gigantic "Thank you" is shouted to Prof. Susanne Crewell who helped me very much especially in the beginning. She always gave me the feeling that my work was her biggest priority even though she has tons of other things to do. She also gave me the freedom to take my own turn around the CRC 1211 community.

I would like to thank Prof. Dr. Stephanie Fiedler and Prof. Dr. Jörg Bendix for agreeing to review this thesis.

I would also like to thank Prof. Dr. Martin Melles to take take the head chair of the defense committee. Furthermore, I like to thank him for guiding the review process towards the second paper.

A big thank you goes to Dr. Jan Schween who was always ready discuss Atacama-related issues. He always gave me the feeling that he has read every single paper anyone has ever written on the subject. So new inspiration was coming after a talk with him. Also thank you, for reviewing Chapter 2!

For always cheering me up and spreading an incredibly positive aura, I would like to thank Dr. Mark Reyers. Many collaborations arose, because the chemistry is just right. Also thank you for reviewing Chapter 7 and helping me with so many other manuscripts and endless discussions.

I also like to thank Dr. Claudia Voigt from the geology department for multiple things. Thank you for reviewing Chapter 1, thank you for initiating many PhD activities around the CRC 1211, thank you for patiently explaining paleo-humidity proxies and the development thereof.

Another thank you goes to Prof. Dr. Ulrich Löhnert who always had an open door for questions and who would always give feedback on presentations and thus moving me forward.

Of course, I would like the entire INFERNO group for bringing so many different topics and approaches to the group. While the diversity in topics makes it sometimes tough to follow, it also forces one to

constantly think outside the own comfort zone which is beneficial in so many ways.

I would also like to thank the administration of the Institute and the University of Cologne. I never had an issue which means you are doing an incredible job!

I would like to thank Dr. Benedikt Maier for bringing me to the KERAS package and agreeing to be part of the fog study. Thank you, for taking the time to discuss all the neural network issues and jointly draining any obstacles down to laughter.

Thank you, Lucas Mittenentzwei for the Pizza and the Tiramisu last Saturday. It was exactly what I needed to get through those conclusions!

Thanks to my family who unconditionally support me and always feel with me with great empathy. They always know what I need, even though I never say it.

The biggest "Thank you" of all is going to Antanina, who read the entire thesis and most likely found every missing comma! You also turned my life around when I was stuck in Chapter 1. I cannot imagine how I would have done it without you. Thank you for all your understanding, patience and encouragement.

DECLARATION

Hiermit versichere ich an Eides statt, dass ich die vorliegende Dissertation selbstständig und ohne die Benutzung anderer als der angegebenen Hilfsmittel und Literatur angefertigt habe. Alle Stellen, die wörtlich oder sinngemäßaus veröffentlichten und nicht veröffentlichten Werken dem Wortlaut oder dem Sinn nach entnommen wurden, sind als solche kenntlich gemacht. Ich versichere an Eides statt, dass diese Dissertation noch keiner anderen Fakultät oder Universität zur Prüfung vorgelegen hat; dass sie - abgesehen von unten angegebenen Teilpublikationen und eingebundenen Artikeln und Manuskripten - noch nicht veröffentlicht worden ist sowie, dass ich eine Veröffentlichung der Dissertation vor Abschluss der Promotion nicht ohne Genehmigung des Promotionsausschusses vornehmen werde. Die Bestimmungen dieser Ordnung sind mir bekannt. Darüber hinaus erkläre ich hiermit, dass ich die Ordnung zur Sicherung guter wissenschaftlicher Praxis und zum Umgang mit wissenschaftlichem Fehlverhalten der Universität zu Köln gelesen und sie bei der Durchführung der Dissertation zugrundeliegenden Arbeiten und der schriftlich verfassten Dissertation beachtet habe und verpflichte mich hiermit, die dort genannten Vorgaben bei allen wissenschaftlichen Tätigkeiten zu beachten und umzusetzen. Ich versichere, dass die eingereichte elektronische Fassung der eingereichten Druckfassung vollständig entspricht.

Köln, September 2020

Christoph Böhm

FULLY INTEGRATED PUBLICATIONS AND SUBMISSIONS:

Böhm, C., O. Sourdeval, J. Mülmenstädt, J. Quaas, and S. Crewell (2019). "Cloud base height retrieval from multi-angle satellite data." In: *Atmospheric Measurement Techniques* 12.3, pp. 1841–1860. DOI: 10.5194/amt-12-1841-2019.

PERSONAL CONTRIBUTION: The idea for this study was presented from Prof. Johannes Quaas, Odran Sourdeval and Johannes Mülmenstädt. Susanne Crewell suggested to add the application on the southeast Pacific and the comparison to reanalysis data. The study was designed by myself and fine tuned in discussions among the co-authors. MISR data was gathered by myself, whereas the ceilometer reference data was gathered and provided by Johannes Mülmenstädt. I processed the data and created all graphics by myself. The manuscript was written by me with helpful suggestions from all co-authors. Furthermore, seven anonymous reviewers gave valuable input which initiated the thorough investigation of the scene structure influence which was carried out by me.

Böhm, C., M. Reyers, J. H. Schween, and S. Crewell (2020b). "Water vapor variability in the Atacama Desert during the 20th century." In: *Global and Planetary Change* 190, p. 103192. ISSN: 0921-8181. DOI: 10.1016/j.gloplacha.2020.103192.

PERSONAL CONTRIBUTION: The idea for this study came from Prof. Susanne Crewell. The study was designed by myself and fine tuned in discussions among the co-authors. All data was gathered by myself except for the cut-off lows which were identified by Mark Reyers who also provided the dates of cut-off low situations. Furthermore, I processed the data and created all graphics by myself. The manuscript was written by me with helpful suggestions from all co-authors. Furthermore, two anonymous reviewers gave valuable input which initiated an additional bias assessment and a study of lagged correlations which were carried out by me.

Böhm, C., J. H. Schween, M. Reyers, B. Maier, U. Löhnert, and S. Crewell (2020c). "Towards a climatology of fog frequency in the Atacama Desert via multi-spectral satellite data and machine learning techniques." Submitted to: *Journal of Applied Meteorology and Climatology*, date of submission: 10 Sep 2020.

PERSONAL CONTRIBUTION: The idea for this study was my own. The study was designed by myself and fine tuned in discussions

among the co-authors. All data was gathered by myself except the meteorological data from the ground-base stations which were compiled and preprocessed by Jan Schween. Furthermore, I processed the data and created all graphics by myself. Benedikt Maier suggested the use of the KERAS package to set up the neural network which proved to outperform other packages which I tried before. Furthermore, he assisted in finding a useful set up of the neural network. Mark Reyers inspired me to apply Self-Organizing Maps as a helpful tool to create the ground-based reference data set. The manuscript was written by me with helpful suggestions from all co-authors.

PARTIAL PUBLICATIONS:

Jaeschke, A., C. Böhm, F. F. Merklinger, S. M. Bernasconi, M. Reyers, S. Kusch, and J. Rethemeyer (2019). "Variation in d15N of fog-dependent Tillandsia ecosystems reflect water availability across climate gradients in the hyperarid Atacama Desert." In: *Global and Planetary Change* 183, p. 103029. ISSN: 0921-8181. DOI: 10.1016/j.gloplacha.2019.103029.

PERSONAL CONTRIBUTION: I created Figures 3 and 4a and wrote section 2.4. (remote sensing section). Furthermore, I created provided the cloud height data and gathered the precipitation data utilized in this study. I participated in various discussions with the first author, making suggestions how the satellite remote sensing data should be interpreted and how it could be relevant for this study.

Cantalloube, F., J. Milli, C. Böhm, S. Crewell, J. Navarrete, K. Rehfeld, M. Sarazin, and A. Sommani (2020). "The impact of climate change on astronomical observations." In: *Nature Astronomy* 4, pp. 826–829. ISSN: 2397-3366. DOI: 10.1038/s41550-020-1203-3.

PERSONAL CONTRIBUTION: I gathered and preprocessed ERA-20C IWV and temperature data which is utilized in this study. Furthermore, I participated in discussions among the co-authors making suggestions how the data should be interpreted and what potential pitfalls are. Furthermore, I reviewed the manuscript and made some few improvements on the text.

Reyers, M., C. Böhm, L. Knarr, Y. Shao, and S. Crewell (Aug. 2020). "Synoptic-to-regional scale analysis of rainfall in the Atacama Desert (18°S–26°S) using a long-term simulation with WRF." In: *Monthly Weather Review*, pp. 1–51. ISSN: 0027-0644. DOI: 10.1175/MWR-D-20-0038.1.

PERSONAL CONTRIBUTION: I gathered and preprocessed the atmospheric river data to extract dates which are associated with atmospheric river conditions in the Atacama Desert. I gathered and provided the gauge measurement data utilized in this study. I assisted in interpreting the figures and results and in putting them in a broader perspective regarding the existing literature body.

UNPUBLISHED WORK:

Böhm, C., M. Reyers, L. Knarr, and S. Crewell (2020a). "The role of atmospheric rivers for precipitation in the Atacama Desert." Unpublished.

PERSONAL CONTRIBUTION: The idea for this study arose within the work group of Prof. Susanne Crewell after atmospheric rivers become a hot topic. The study was designed by myself and fine tuned in discussions among the co-authors. All data were gathered by myself except for all data derived from the WRF simulation which was carried out by Mark Reyers as well as the IGRA data which were gathered by Mark Reyers and the data relevant for the back trajectory analysis which were gathered by Leon Knarr. Mark Reyers created the individual panels which are compiled in Figure 7.5. The two panels compiled in Fig. 7.4 were created by Leon Knarr who carried out the trajectory analysis using HYSPLIT. Aside from the WRF vertical cross sections and the trajectory analysis, I processed the data and created all other graphics by myself. The Chapter was written entirely by me with some helpful suggestions from Mark Reyers.