

# Dissertation

Submitted to the  
Combined Faculty of Mathematics,  
Engineering and Natural Sciences  
Heidelberg University, Germany  
for the degree of  
Doctor of Natural Sciences (Dr. rer. nat.)

Presented by  
Juan Carlos Pastene Berríos

Oral Examination: March 25<sup>th</sup>, 2025

Fog variability patterns in the Chilean Atacama Desert –  
Enhanced climatological data series and analysis of regional  
climatic interactions and local topographic influence

**Supervisor and Reviewer:**

Prof. Dr. Alexander Siegmund

Department of Geography

Heidelberg University of Education & Ruprecht-Karls-University of Heidelberg, Germany

**Co-Supervisor and Reviewer:**

Prof. Dr. Olaf Bubbenzer

Institute of Geography

Ruprecht-Karls-University of Heidelberg, Germany

## Table of contents

<b>Acknowledgments</b>	<b>I</b>
<b>Abstract</b>	<b>III</b>
<b>Zusammenfassung</b>	<b>V</b>
<b>List of abbreviations</b>	<b>VII</b>
<b>List of figures</b>	<b>IX</b>
<b>List of tables</b>	<b>XII</b>
<b>List of programming flows</b>	<b>XIII</b>
<b>1. Introduction</b>	<b>1</b>
<b>2. Research area</b>	<b>4</b>
2.1. Study area	5
2.2. Study site	9
<b>3. State of the art of fog patterns understanding in the Atacama Desert: multiscale exploration and climatic analysis</b>	<b>11</b>
3.1. Fog research and fog climate measurements background in Chile	11
3.2. Climate data management: quality and control	14
3.3. Fog cycles in the northern Chilean Atacama Desert	16
3.4. Multivariate analysis on topographic and climatic influences on fog variability	18
<b>4. Research objectives, methodological concept, and methodology</b>	<b>21</b>
4.1. General and specific objectives	21
4.2. Research and design methods	23
4.2.1. Quality control	24
4.2.2. Regional analysis	24
4.2.3. Local analysis	26
<b>5. Fog variability characterization in the Chilean Atacama Desert: fog climate data quality assurance and multiscale spatio-temporal analysis</b>	<b>30</b>
5.1. Quality control and data processing	30

## Table of contents

---

5.1.1. Data quality flow	30
5.1.2. Gaps checking: missing data and non-numeric values	32
5.1.3. Outliers checking	37
5.1.4. Filling data process	55
5.2. Regional scale - Temporal analysis	58
5.2.1. Altitudinal and coast-inland variability of fog climate	58
5.2.2. Monthly variability of fog climate	66
5.2.3. Daily variability of fog	74
5.3. Local scale - Spatial analysis	77
5.3.1. Near-surface fog water measurements and explanatory variables	77
5.3.2. Spatial analysis validation	84
5.3.3. Geo-statistical modelling and analysis	90
<b>6. Discussions of multiscale findings on fog variability in the Chilean Atacama Desert</b>	<b>103</b>
<b>7. Conclusions and outlooks in understanding fog climate patterns</b>	<b>110</b>
<b>References</b>	<b>113</b>

## Acknowledgments

This PhD has been the result of an intense, demanding, yet immensely rewarding process made possible by the collective support of institutions, my supervisor, colleagues, friends, and family.

First and foremost, I wish to express my deepest appreciation, admiration, and gratitude to Prof. Dr. Alexander Siegmund for his unwavering confidence in me. His guidance, encouragement, and dedication have been pivotal in shaping this research and fostering my academic and professional development as a geographer. I am profoundly grateful for his steadfast support and for the invaluable scientific knowledge and expertise he has generously shared with me throughout this journey.

I am grateful to have Prof. Dr. Olaf Bubenzer as the second supervisor of this research. His expertise and distinguished contributions to the field provide invaluable depth to the project, and it is an honor to have his involvement. His presence enriches the academic rigor and broadens the scientific perspective of this work.

I also extend my gratitude to Dr. Camilo del Río and Dr. Rosa Lehmann for their esteemed presence as members of my thesis committee. Their involvement not only enriches this research with valuable perspectives but also brings an added honor and depth to the project. I am truly grateful for their support, which lends distinction to this work and underscores its significance.

This research has been supported by ANID and DAAD under the Program ANID-PFCHA/Doctorado con Acuerdo Bilateral en el Extranjero CONICYT-DAAD 2018-62180010. In addition, I am grateful to the FEBiD Project “The Chilean-Peruvian arid coastal fog ecosystems under climate change: understanding biosphere-atmosphere interactions to support biodiversity conservation” whose framework allowed the successful completion of this thesis.

To my colleagues at Department of Geography – Research Group for Earth Observation (rgeo) at the Heidelberg University of Education and the Heidelberg University, I extend my heartfelt thanks for their daily support and collegiality.

I also express my sincere gratitude to the Heidelberg Center for Latin America (HCLA) for their continuous support throughout my postgraduate studies, as well as to the Geosciences team. Special thanks to Dr. Michael Handke, Coordinator of the Master of Science in Governance of Risk and Resources, and to Dr. Johanna Höhl, who was part of the HCLA team during my time there.

## Acknowledgments

---

I am deeply grateful to the Centro UC Desierto de Atacama (CDA), especially to its Director, Dr. Camilo del Río, for his enduring support and warmth throughout the research process and beyond, as well to his family. I also wish to thank the Director of the Atacama Research Station UC Alto Patache, Mag. Pablo Osses, for his support and for generously sharing his valuable field knowledge and expertise.

My thanks to the Deutscher Wetterdienst (DWD) for their assistance in improving the quality of climate data from the Atacama Desert fog network, with special thanks to Dr. Antje Claußnitzer and Mr. Reinhard Spengler.

My appreciation also extends to the Kurt-Hiehle-Foundation for their financial support, which enabled me to carry out fieldwork in the Chilean Atacama Desert and generate high-quality data and insights.

I would like to thank the Military Geographic Institute of Chile (IGM) for their unwavering support throughout my professional and academic development, helping to foster the conditions necessary for advancing geographic knowledge and capabilities.

Finally, my deepest and most heartfelt thanks go to my family and friends in Santiago, Lo Miranda, Concepción, San Pedro de la Paz, Temuco, Valdivia, and Castro in Chile, as well as to my dear friends in Heidelberg and throughout Central Europe. Their unconditional encouragement, motivation, and, above all, their patience in putting up with me have been invaluable throughout this journey. I owe my deepest gratitude to my parents, Agna and Juan, whose love and great efforts made all of this possible.

Thank you so much.

## Abstract

The northern Chilean coastal regions are known for their extreme aridity, with annual precipitation often less than 1 mm. The interaction between ocean-atmosphere systems plays a crucial role in shaping marine advection fog patterns in this area, which is vital for the hyper-arid ecosystem along the Atacama coast. While extensive geographical research has been conducted on fog patterns since the mid-20<sup>th</sup> century, there is limited knowledge in managing fog climate records and understanding their regional and local climatological dynamics. This research is organized into three main sections. The first one outlines a methodological climate data flow, emphasizing data quality control and processing. The second section focuses on analyzing temporal gradients of fog climate at the regional level. The third section delves into local spatial analysis and modeling of Near-surface fog water in the inland desert.

The processing of raw climatic data recorded from extreme environments such as the extreme aridity of the Atacama Desert poses a significant challenge in climatic research, as they exhibit gaps, inconsistencies, and errors. The process of handling climatic data from the eight climatological stations and Standard Fog Collectors (SFCs) comprising the fog climate network in the northern Atacama Desert, Region of Tarapacá, involves data identification, cleansing, correction, and normalization. This is achieved through programming flows guided by standardized parameters, primarily defined by the World Meteorological Organization (WMO). Data gaps and outliers within the network's data series account for less than 2% of the total record. Anomalies in the data series are primarily attributed to electrical and electronic issues inherent in remote/satellite transmission. The precision and quality of climatic data are essential for accurate interpretation and analysis, enhancing the understanding of fog climate dynamics in hyper-arid environments.

The regional approach aims to comprehend the variability between fog collection and its climatological gradient. The datasets were analyzed at different temporal intervals across eight climatological stations. The analysis elucidates how the variability of fog climates and their driving factors in these datasets manifest themselves throughout the region, considering both temporal and geographical determinants. Throughout the year, fog exhibits marked diurnal variation, being more pronounced in the morning and minimal in the afternoon and night. The seasonal influence of the thermal inversion layer (TIL) is significant, with more fog water collected in the winter due to a lower TIL and less collection in the summer due to a higher TIL. High relative humidity is crucial for fog formation in coastal areas, while inland regions are drier. Wind, varying in direction and speed, also

impacts fog. Overall, this regional-temporal analysis shows the multifaceted nature of fog in the Atacama Desert, elucidating its dependence on diurnal, seasonal, and geographical variables. These findings contribute to a better understanding of fog dynamics within this unique arid ecosystem.

In an inland desert at ca. 1,200 meters above sea level (masl) within the "Oyarbide" Site, a local study analyses the link between Near-surface fog water variability and topographical features. The monthly data of ten Mini Fog Collectors (Mini FCs) placed at Oyarbide and Digital Elevation Model (DEM) data, an Ordinary Least Squares (OLS) and Geographically Weighted Regression (GWR) were employed for regression analysis. These methods evaluated the influence of five topographical factors on Near-surface fog water. Results revealed patterns and spatial variability in these associations, emphasizing the impact of topography on Near-surface fog water. Elevation was positively correlated with Near-surface fog water, while Aspect and Slope are related to terrain roughness. Local GWR model outperformed the global OLS model, substantially increasing the coefficient of determination ( $R^2$ ). This provides a comprehensive understanding of how local topography affects Near-surface fog water collection, enhancing the knowledge of atmosphere-biosphere interactions.

## Zusammenfassung

Die Küstenregionen im Norden Chiles sind für ihre extreme Trockenheit bekannt, mit jährlichen Niederschlägen oft unter 1 mm. Die Wechselwirkung zwischen Ozean-Atmosphäre-Systemen spielt eine entscheidende Rolle bei der Formung von marinen Advektionsnebelmustern in dieser Region, die für das hyperaride Ökosystem entlang der Atacama-Küste von großer Bedeutung ist. Obwohl seit Mitte des 20. Jahrhunderts umfangreiche geografische Forschung betrieben wurde, sind die Kenntnisse im Umgang mit Klimaaufzeichnungen des Nebelklimas und das Verständnis ihrer regionalen und lokalen klimatologischen Dynamik begrenzt. Diese Forschungsarbeit ist in drei Hauptabschnitte unterteilt. Der erste Abschnitt skizziert einen methodischen Datenfluss im Klima, wobei die Qualität und Verarbeitung der Daten betont wird. Der zweite Abschnitt konzentriert sich auf die Analyse von zeitlichen Gradienten des Nebelklimas auf regionaler Ebene. Der dritte Abschnitt vertieft die lokale räumliche Analyse und Modellierung von oberflächennahem Nebelwasser in der Binnenwüste.

Die Verarbeitung von Klimarohdaten aus extremen Umweltbedingungen wie sie die extreme Trockenheit der Atacama-Wüste darstellt, stellt die Klimaforschung vor eine erhebliche Herausforderung, da sie Lücken, Inkonsistenzen und Fehler aufweisen. Der Prozess der Bearbeitung von Klimadaten aus den acht klimatologischen Stationen und Standard Fog Collectors (SFCs), die das Nebelklimanetzwerk in der nördlichen Atacama-Wüste, Region Tarapacá, bilden, umfasst die Identifikation, Bereinigung, Korrektur und Normalisierung von Daten. Dies erfolgt durch Programmierungstechniken, die von standardisierten Parametern geleitet werden und hauptsächlich von der Weltorganisation für Meteorologie (WMO) definiert sind. Datenlücken und Ausreißer in den Datenreihen des Netzwerks machen weniger als 2% der Gesamtaufzeichnung aus. Anomalien in den Datenreihen werden hauptsächlich auf elektrische und elektronische Probleme bei der Fern-/Satellitenübertragung zurückgeführt. Die Präzision und Qualität der Klimadaten sind für die genaue Interpretation und Analyse dieser Daten von entscheidender Bedeutung, um unser Verständnis der Dynamik des Nebelklimas in hyperariden Umgebungen zu vertiefen.

Der regionale Ansatz zielt darauf ab, die Variabilität zwischen der Nebelsammlung und ihrem klimatologischen Gradienten zu verstehen. Die Datensätze wurden in verschiedenen Zeitabständen an acht Klimastationen analysiert. Die Analyse verdeutlicht, wie sich die Variabilität des Nebelklimas und seiner treibenden Faktoren in diesen Datensätzen in der gesamten Region manifestiert, wobei sowohl zeitliche als auch geografische Determinanten berücksichtigt werden. Während des ganzen

Jahres weist der Nebel ausgeprägte tageszeitliche Schwankungen auf, die morgens stärker und nachmittags und nachts minimal ist. Der jahreszeitliche Einfluss der thermischen Inversionsschicht (TIL) ist signifikant, wobei im Winter aufgrund einer niedrigeren TIL mehr Nebelwasser und im Sommer aufgrund einer höheren TIL weniger Nebelwasser gesammelt wird. Hohe relative Luftfeuchtigkeit ist entscheidend für die Nebelbildung in Küstengebieten, während die Regionen im Landesinneren trockener sind. Der Wind, der in Richtung und Geschwindigkeit variiert, wirkt sich auch auf den Nebel aus. Insgesamt zeigt diese regional-zeitliche Analyse die vielfältigen Eigenschaften des Nebels in der Atacama-Wüste und verdeutlicht seine Abhängigkeit von tages- und jahreszeitlichen sowie und geografischen Variablen. Diese Erkenntnisse tragen zu einem besseren Verständnis der Nebeldynamik in diesem einzigartigen ariden Ökosystem bei.

Am Hauptstandort "Oyarbide", in einer Wüste im Landesinneren auf etwa 1.200 Metern über dem Meeresspiegel (m ü. M.) wird in einer lokalen Studie der Zusammenhang zwischen der Variabilität des oberflächennahen Nebelwassers und den topografischen Merkmalen analysiert. Für die Regressionsanalyse wurden die monatlichen Daten von zehn Mini Fog Collectors (Mini FCs), die in Oyarbide aufgestellt wurden, Daten eines digitalen Höhenmodells (DEM), eine Methode der kleinsten Quadrate (OLS) sowie eine geografisch gewichtete Regression (GWR) verwendet. Mit diesen Methoden wurde der Einfluss von fünf topografischen Faktoren auf das oberflächennahe Nebelwasser untersucht. Die Ergebnisse deckten Muster und räumliche Variabilität in diesen Zusammenhängen auf und betonten die Auswirkung der Topografie auf das oberflächennahe Nebelwasser. Die Höhe war positiv mit dem oberflächennahen Nebelwasser korreliert, während Exposition und Hang in Beziehung zur Geländerauigkeit standen. Das lokale GWR-Modell übertraf das globale OLS-Modell und erhöhte den Bestimmtheitskoeffizienten ( $R^2$ ) erheblich. Dies ermöglicht ein umfassendes Verständnis dafür, wie die lokale Topografie die Sammlung von oberflächennahem Nebelwasser beeinflusst, und erweitert das Wissen über die Wechselwirkungen zwischen Atmosphäre und Biosphäre.

## List of abbreviations

### *Climatological variables*

AP	Air pressure
AT	Air temperature
DP	Dew point
FW360	Fog water 360°
GR	Global radiation
LW	Leaf wetness
RH	Relative humidity
DW	Dew water
SFW	Standard fog water
SoM	Soil moisture
SoT	Soil temperature
ST	Surface temperature
WD2	Wind direction at 2 meters
WD10	Wind direction at 10 meters
WS2	Wind speed at 2 meters
WS10	Wind speed at 10 meters

*Other abbreviations*

AIC	Akaike information criterion
MBL	Marine boundary layer
CDA	Centro UC Desierto de Atacama
DEM	Digital elevation model
ENSO	El Niño-Southern Oscillation
GIS	Geographic Information System
GWR	Geographically Weighted Regression
IPO	Interdecadal Pacific Oscillation
IQR	Interquartile range
JB	Jarque-Bera statistic
K(BP)	K-Bowman tests
masl	Meters above sea level
Mini FC	Mini Fog Collector
OLS	Ordinary Least Squares
QC	Quality Control
RSS	Residual Sum of Squares
r	Pearson correlation
R <sup>2</sup>	Coefficient of determination
SEPH	Southeast Pacific High
SFC	Standard Fog Collector
TIL	Thermal inversion layer
VIF	Variance inflation factor
WMO	World Meteorological Organization

## List of figures

Figure 1: North to south satellite view of the research area – Northern Chilean Atacama Desert	4
Figure 2: View of the transitional desert-ocean gradient from OYA_1069 in Southern Autumn	5
Figure 3: Regional clusters of Tillandsia fields on a 3D perspective	6
Figure 4: Fog climate measurement network at regional scale approach	7
Figure 5: Terrain of the study site “Oyarbide” on a fog event	9
Figure 6: Mini FCs measurements network at local scale in the study site	10
Figure 7: Overview of the methodological approach proposed by ‘geo/CDA for researching fog climate and fog ecosystems in the Chilean Atacama Desert	12
Figure 8: Fog formation and its ocean-continent movement	13
Figure 9: Fog penetration corridors in the Tarapacá Region	17
Figure 10: Average trends using global statistical model (left) masking local trends revealed in local statistical model (right)	19
Figure 11: Multiscale data sources and measurement approaches for analyzing fog variability	22
Figure 12: Methodological framework research on fog variability patterns in the Atacama Desert	23
Figure 13: Climatological station OYA_1211 and SFC at the upper limit of Oyarbide	25
Figure 14: Fog climate network and its measurement sensors	26
Figure 15: Structure and components of a Mini FC	27
Figure 16: Data quality flow to improve regional-scale fog climate network climate data series	31
Figure 17: Standard plots of gaps and non-numeric data by variable in a data series (OYA_518)	35
Figure 18: Missing data across complete time intervals in a data series (OYA_518)	37
Figure 19: Standard plots of the Tukey's test check by variable in a data series (OYA_1211)	40
Figure 20: Standard plots of the plausible value check by variable in a data series (OYA_1211)	43

Figure 21: Standard plots of the check on a maximum allowed variability of an instantaneous value by variable in a data series (OYA_780)	46
Figure 22: Standard plots of the check on a minimum required variability of 1 hr values by variable in a data series (OYA_780)	49
Figure 23: Standard plot of the check on the RH threshold in a data series (OYA_1193)	51
Figure 24: Homogeneous fog climate measurement systems at regional scale	52
Figure 25: Standard plot of the check on the inter-stations correlation in a data series (System 1)	55
Figure 26: Annual cumulative SFW from coast to inland desert	59
Figure 27: Variability of the collected SFW from coast to inland desert	60
Figure 28: Wind patterns gradients from coast (518 masl) to inland desert (862 and 1,211 masl)	62
Figure 29: Relationship between the amounts of SFW collected and wind patterns (1,211 masl)	63
Figure 30: Fog annual variability and its relationship with SFW collection and DP records	64
Figure 31: LW over 70% recording fog above and below TIL (1,354 masl and 1,211 masl)	65
Figure 32: Monthly cumulative (ml) SFW collected inland desert (year 2020)	66
Figure 33: Annual gradient of fog events inland desert	68
Figure 34: RH gradient near coast (518 masl) and inland desert (1,211 masl)	70
Figure 35: AT gradient near coast (518 masl) and inland desert (1,211 masl)	72
Figure 36: Day and night fog events and its relationship with AT and RH	73
Figure 37: Hourly distribution of SFW collection below (1,211 masl) and above TIL (1,354 masl)	75
Figure 38: 10-minute average AT and SFW collection records at 1,211 masl	77
Figure 39: Interpolation of Near-surface fog water variability from June to November 2019	79
Figure 40: Topographic explanatory variables in Oyarbide extracted from the DEM	81
Figure 41: Bivariate correlation between dependent and explanatory variables in Oyarbide	83
Figure 42: Spatial autocorrelation report of Near-surface fog water collection	85
Figure 43: High-low clustering report of Near-surface fog water collection	86
Figure 44: GWR Local $R^2$	94

List of figures

---

Figure 45: GWR t-value for each explanatory variable _____	96
Figure 46: Local influence of explanatory variables on Near-surface fog water variability _____	97
Figure 47: Coefficients of combined explanatory variables and their relationship with Near-surface fog water variability _____	99
Figure 48: Relationship between Elevation and predicted Near-surface fog water _____	100
Figure 49: Wind patterns with (left) and without (right) fog collection inland desert _____	104
Figure 50: GWR standard residual values distribution _____	108

## List of tables

Table 1: Fog climate instrumentations and its features at regional scale _____	8
Table 2: Percentage of filtered data by climate station and data filter _____	57
Table 3: Spatial autocorrelation for the explanatory variables _____	85
Table 4: Exploratory regression overview _____	88
Table 5: Exploratory regression multicollinearity testing _____	89
Table 6: Exploratory regression assessment of variable significance _____	90
Table 7: OLS model summary _____	91
Table 8: OLS diagnostic model _____	92
Table 9: Summary of results of the GWR model analysis _____	93
Table 10: Comparison between global and local regression modelling _____	101
Table 11: Moran's I in the residuals OLS and GWR _____	102

## List of programming flows

Programming flow 1: Gaps checking: missing data _____	34
Programming flow 2: Outliers checking by Tukey's test _____	40
Programming flow 3: Outliers check by plausible value _____	43
Programming flow 4: Outliers check by maximum allowed variability of an instantaneous value __	46
Programming flow 5: Outliers check by minimum required variability of 1 hr values _____	49
Programming flow 6: Outliers checking by RH threshold _____	51
Programming flow 7: Outliers identification by inter-stations correlation _____	54

## 1. Introduction

The western coastal regions of South America are among the driest areas on the planet, with average annual precipitation averaging less than 1 mm annually (Westbeld et al., 2009). Despite these arid conditions, fog plays a vital role in providing natural water resources for local ecosystems and human populations (Cermak et al., 2009). Based on regional studies, the fog in this region is generated by stratocumulus clouds that form over the cold Humboldt Current and humid air masses, which are then transported inland by trade winds and land-sea circulation. When the stratocumulus clouds interact with the coastal topography, they become advection fog. This ocean-atmosphere interaction drives the spatio-temporal dynamics of marine advection fog, which is essential to the development of unique ecosystems and high biodiversity along the Atacama coast (Cereceda et al., 2008b).

Extensive research has been conducted on fog in the Chilean section of the Atacama coastal desert, with a particular focus on its role as a water resource for *Tillandsia landbeckii* ecosystems and human consumption (Cereceda et al., 2002; Schemenauer & Cereceda, 1994a). Geographic studies of this region date back to the mid-20<sup>th</sup> century (Weischet, 1975), and it remains an area of significant interest, with fog research covering approximately 100,000 km<sup>2</sup> in northern Chile (Cereceda et al., 2008a; Cereceda et al., 2002). The distinctive environmental conditions of the Atacama coast have fostered a diverse range of ecological dynamics, making it an ideal setting for in-depth studies of fog processes and their ecological impacts.

Despite significant research on fog in the Chilean Atacama coastal desert, there is still limited understanding of the spatio-temporal dynamics of fog variability on a local and regional scale, especially concerning interactions with the unique fog ecosystems of Tillandsia Lomas dominated by the endemic species *Tillandsia landbeckii*.

This knowledge gap underscores the necessity for a detailed analysis of the spatial and temporal dynamics of fog and its driving factors. Consequently, this study seeks to address the following research questions:

Does the management of the fog climate network database, based on the identification and rectification of gaps and outliers, culminate in an enhancement of database quality, thereby fostering heightened reliability and robustness in ensuing multiscale analyses of fog variability?

How does fog occurrence, duration, and intensity vary across the regional climatological gradient over time, and what is the relationship between these patterns, altitude, and proximity to the coast?

What are the primary topographical factors and their thresholds that drive the spatial variability of Near-surface fog water on a local scale in the inland desert, and how do these interactions influence the variability and accumulation of Near-surface fog water?

By addressing these research questions, this study aims to deepen the understanding of fog variability patterns across multiple scales within the Tarapacá region of northern Chile. Through comprehensive spatio-temporal analyses and modeling of fog-related climatological and topographical data, this research seeks to characterize the variability of fog collection and occurrence, which play a critical role in the region's climatology. Specifically, this work provides insights into effective approaches for modeling fog climate and examining biosphere-atmosphere interactions in the Atacama Desert. This approach is designed to bridge existing knowledge gaps by examining the spatial and temporal dynamics of fog and its primary influencing factors, contributing to a more robust framework for studying hyper-arid, fog-dependent ecosystems.

Firstly, to assess the precision and accuracy of the data provided by the fog climate measurement network and enhance data quality as necessary. Secondly, to define the spatio-temporal variability and patterns of fog on a regional scale and investigate how these are influenced by spatial characteristics such as altitude and proximity to coastal gradients. Finally, to analyze the variability of Near-surface fog and assess the spatial influence of key topographical factors on fog dynamics.

This research has a significant scope and will contribute to advancing the scientific knowledge of fog climatology and its ecological implications in hyper-arid environments. The findings of this study will have implications for regional climate modelling, ecosystem management, and biodiversity conservation efforts and the research results will be relevant not only to researchers but also to policymakers who are interested in climatology, climate change and adaptation in hyper-arid regions through the provision of detailed knowledge on regional-local fog dynamics in the Atacama.

The research adopts a comprehensive approach and is conducted at both regional and local scales. At the regional scale, the study area extends from the coast to approximately 11 km inland into the desert, serving as a representative area for analyzing the fog climate gradient. The local scale study is focused on a representative study site called "Oyarbide" covering an area of ca. 16 km<sup>2</sup>, located near the homonymous hill. The Oyarbide Site is of particular interest owing to the presence of one of the largest *Tillandsia landbeckii* fields in the region, which represents a unique fog-dependent ecosystem covering a range of about 200 m in elevation. It is framed by an east/west directed corridor, where the advective and orographic fog can penetrate far enough inland to reach the study site at around over the upper limit (Cereceda et al., 2007; Westbeld et al., 2009).

To achieve the research objectives, this study is structured around a methodological framework encompassing three main chapters. The first chapter includes an assessment of climatological time series from the fog climate network, applying quality control measures and advanced data processing techniques to ensure data accuracy. The second chapter addresses a regional-scale analysis of fog climate data to identify spatial and temporal patterns that characterize its variability from the coast to inland desert. The final chapter focuses on developing statistical models to evaluate the influence of topographic factors on Near-surface fog water variability at a local scale. This approach facilitates a detailed examination of spatio-temporal patterns in fog variability across both regional and local scales, supporting an enhanced understanding of fog variability in this environment.

Chapter 1 presents the availability, management, and challenges associated with the quality of climatic data obtained from the fog climate measurement network in hyper-arid environments with limited accessibility and data storage/transfer systems. An overview of the regional measurement system in the Atacama Desert is provided, along with a presentation of the data processing and advanced approaches used for the quality control of data transmitted via satellite, as well as the strategies employed to develop high-quality climate data products in such extreme conditions.

Chapter 2 introduces a comprehensive multiscale analysis of the spatio-temporal gradient of the driving parameters that determine the fog climate on a regional scale. This chapter characterizes fog variability through the dynamic of SFW collected by Standard Fog Collectors (SFCs) and the occurrence of fog events in relation to the variability of climatic variables. The analysis is conducted using a transect of climatological stations extending from the coast up to 11 km inland and over 1,300 masl. Advanced statistical techniques, including time series analysis, principal component analysis, and generalized additive models, are employed to identify the spatio-temporal patterns of fog occurrence and its relationship with climatic variables.

Chapter 3 focuses on the spatial analysis and evaluation of the influence of topographic factors affecting fog occurrence and accumulation on the variability of Near-surface fog water, measured at 50 cm from the surface on a local scale. The Ordinary Least Squares (OLS) and Geographically Weighted Regression (GWR) analysis are employed to model the relationship between topography and the Near-surface fog water collection. The differences between the two regression analysis methods are highlighted, and their relevance for modelling this type of interactions is discussed.

This research provides a comprehensive and systematic analysis of data quality management of the regional fog climate measurement system in the Chilean Atacama Desert, regional fog variability and its relationship with climatic gradient, and topographic factors influencing Near-surface fog variability in this hyper-arid environment.

## 2. Research area

The Chilean Atacama Coastal Desert from ca. 18°20'S to 27°S, is part of the hyper-arid geographic complex that extends for more than 3,000 km along a narrow coastal strip from northern Peru to *Norte Chico* of Chile on the west coast of South America. The Atacama Desert's existence is attributed to a stable high-pressure system in the western Pacific Ocean, the Humboldt current's drying effect, and the rain shadow effect of the Andes, restricting moisture penetration from eastern trade winds (Jaksic et al., 1999; Rundel et al., 1991). With an average annual precipitation of less than 1 mm per year, advective sea fog provides water and nutrients to unique ecosystems localized inland Desert (Moat et al., 2021).

In this research area, the focus is specifically on a defined study area approximately 11 km inland from the coast. Within this study area lies a selected study site, covering an area of ca. 16 km<sup>2</sup>, situated ca. 10 km inland. This study area and site is the subject of investigation to understand atmospheric-biosphere dynamics guided by fog collection variability (see Figure 1).

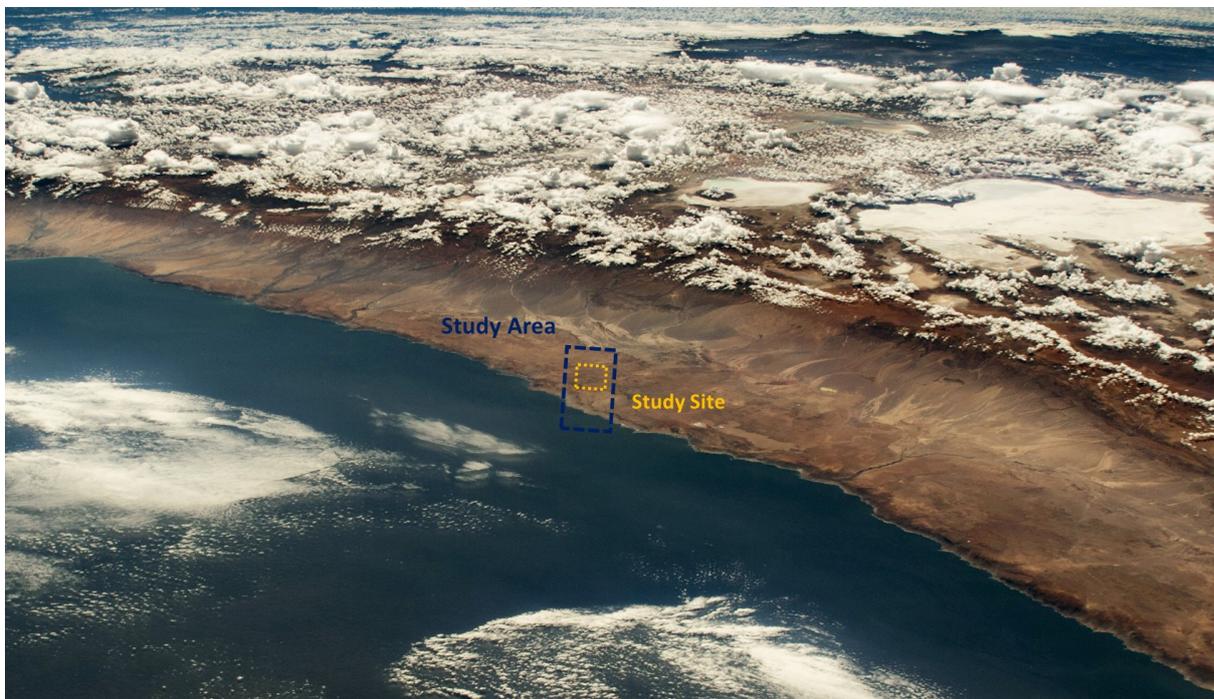


Figure 1: North to south satellite view of the research area – Northern Chilean Atacama Desert.  
Source: Nasa – Earth Observatory<sup>1</sup>

---

<sup>1</sup> [https://earthobservatory.nasa.gov/images/151670/looking-down-on-the-andes?utm\\_source=TWITTER&utm\\_medium=NASAEarth&utm\\_campaign=NASASocial&linkId=229342528](https://earthobservatory.nasa.gov/images/151670/looking-down-on-the-andes?utm_source=TWITTER&utm_medium=NASAEarth&utm_campaign=NASASocial&linkId=229342528)

### 2.1. Study area

The Chilean Atacama Desert, one of the driest regions on Earth, experiences nearly nonexistent rainfall in its core zones. Here, the predominant source of water is a complex water inputs from marine advective fog along the Pacific coast, penetrating deep inland desert from variations along the Atacama coast linked with fluctuations in moisture inputs originating from the Pacific Ocean (Latorre et al., 2011). Multi-scale climatological measurements on a regional approach reveal steep spatial gradients and rapid atmospheric changes in the lower atmosphere. These changes are closely tied to the advection of divergent coastal (humid and cold) and continental (dry and warm) air masses, whose ocean-desert interaction seeks to record the transect of climatic stations that constitute the fog climate network (see Figure 2).

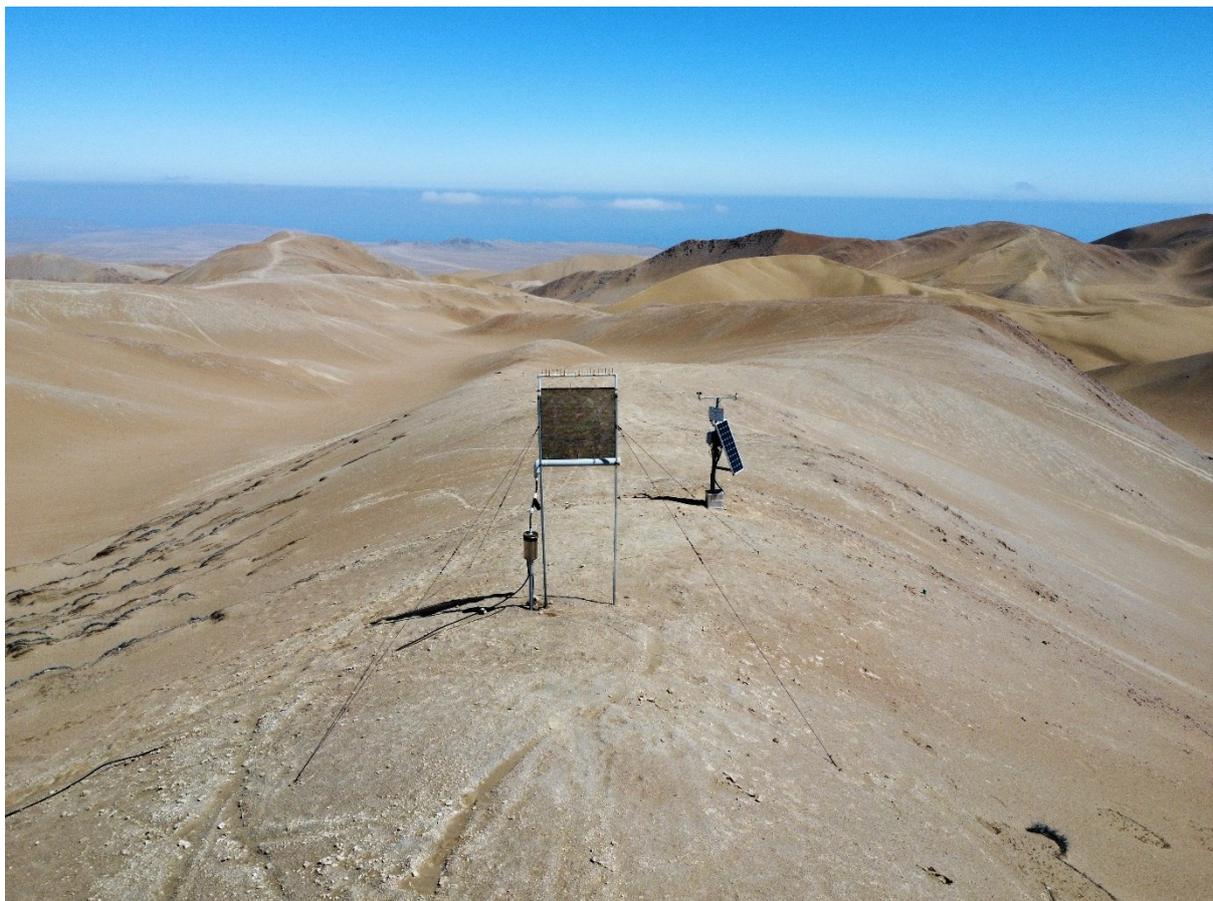


Figure 2: View of the transitional desert-ocean gradient from OYA\_1069 in Southern Autumn.  
Source: author's photographic archive

## 2. Research area

---

This advective fog allows the existence of disjoint patches of vegetation known as Tillandsia Lomas, which are home to endemic species of *Tillandsia landbeckii* that rely exclusively on coastal fog as their primary source of water and constituting important biodiversity hotspot (see Figure 3). The topography of the region generates fog due to the presence of a coastal cliff that intercepts the stratocumulus clouds between 400 and 1,200 masl. Therefore, fog ecosystems exhibit high densities near the ocean and inland locations (Cereceda et al., 1999), and often form linear structures arranged orthogonally to the sloping desert landscape, optimizing fog collection and minimizing self-competition (Wolf et al., 2016).

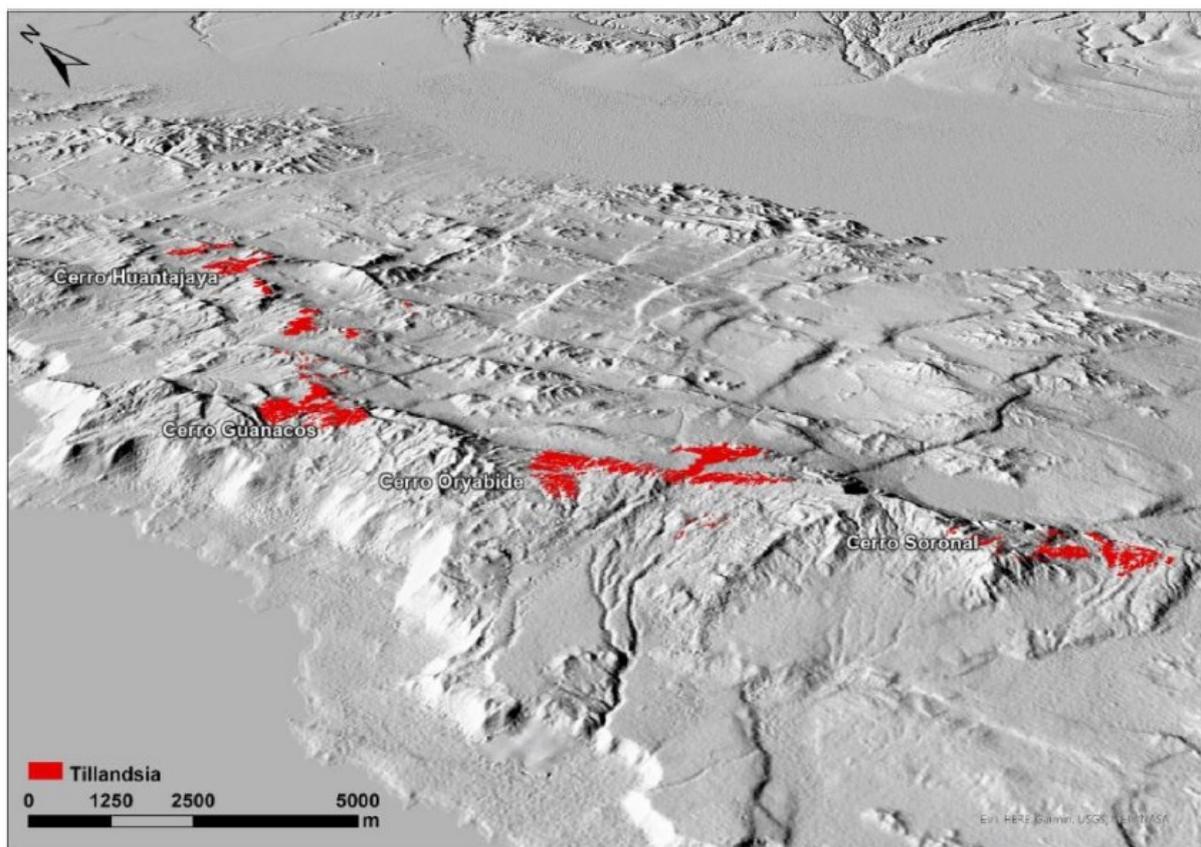


Figure 3: Regional clusters of Tillandsia fields on a 3D perspective. Source: Mikulane et al. (2022)

Embedded within this extreme ecosystem, a network of climatological stations has been set up since 2016, strategically localized from coast to 1,354 masl and extending up to 11 km inland (see Figure 4). These stations systematically record and monitor an array of climatological variables pivotal to study fog dynamics and its driver parameters with a high sensitivity level and high quality of the sensors and store these data at 10-minute intervals. The recorded parameters include standard fog water (SFW) (2 m), fog water 360° (FW360), dew water (DW), air temperature (AT) and relative humidity (RH), soil

## 2. Research area

temperature (SoT) (0.1 m), soil moisture (SoM) (0.1 m), surface temperature (ST) (0.05 m), wind speed and direction (WS and WD at 2/10 m), air pressure (AP), global radiation (GR), leaf wetness (LW), and dew (DP) (see Table 1).

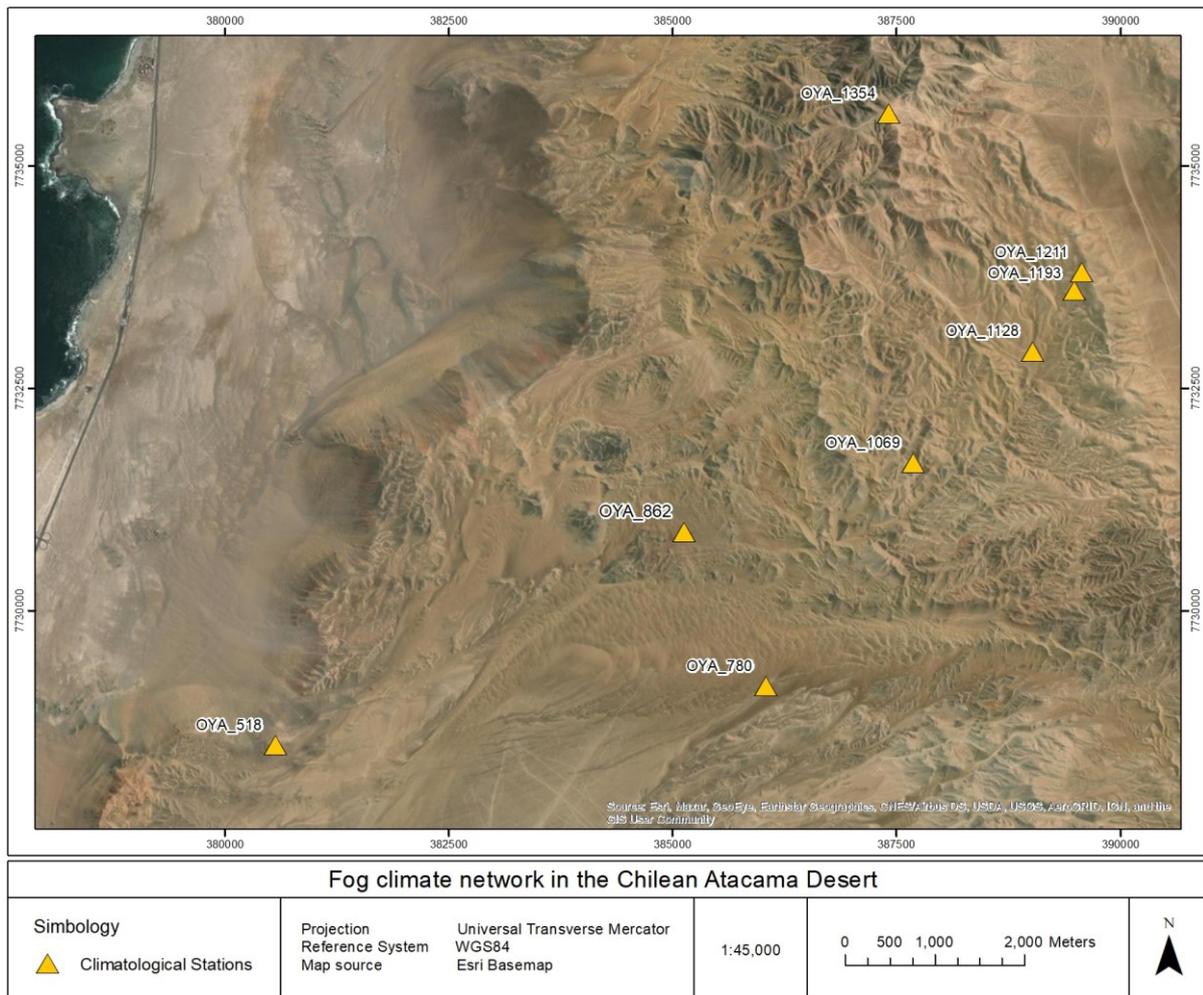


Figure 4: Fog climate measurement network at regional scale approach. Source: author's elaboration based on Esri ArcGIS Basemap

## 2. Research area

Parameter	m	Height	Sensor type	Measuring range	Accuracy
Standard fog water	2		SFC & Precipitation Transmitter/Thies CLIMA	200 cm <sup>2</sup>	0.1mm
Fog water 360°	1		Cylindrical Fog Collector & Precipitation Transmitter/Thies CLIMA	50 cm <sup>2</sup>	0.1mm
Dew water	1		Vertical dew water collector & Precipitation Transmitter/Thies CLIMA	200 cm <sup>2</sup>	0.1mm
Air temperature	2		Hygro-Thermo Transmitter-compact/Thies CLIMA	30... +70°C	± 0.1K
Relative humidity	2		Hygro-Thermo Transmitter-compact/Thies CLIMA	0... 100%	± 2 RH (@ 5... 95% RH and 10... 40°C)
Surface temperature	0.05		Earth Surface Temperature Transmitter Pt 100/Thies CLIMA	30... +50°C	± 0.1K
Soil moisture	0.1		Soil Moisture Probe Trime-Pico 32/IMKO	0... 100%	± 2% (@ 5... 40%), ± 3% (@ 40... 70%)
Soil temperature	0.1		Soil Moisture Probe Trime-Pico 32/IMKO	15... +50°C	± 1.5K
Wind speed (2 m)	2		Wind Transmitter Classic/Thies CLIMA	0.3... 50m/s	< 2% or ± 0.3m/s of meas. value
Wind speed (10 m)	10		Wind Transmitter Classic/Thies CLIMA	0.3... 50m/s	< 2% or ± 0.3m/s of meas. value
Win direction (2 m)	2		Wind Direction Transmitter Classic/Thies CLIMA	0... 360°	± 1.5°
Wind direction (10 m)	10		Wind Direction Transmitter Classic/Thies CLIMA	0... 360°	± 1.5°
Ai pressure	1.5		Baro Transmitter B-278/Thies CLIMA	800... 1060 hPa	± 0.3 hPa @ 20°C
Global radiation	2		Pyranometer CMP 3/Thies Clima	0... 2000 W/m <sup>2</sup>	5... 20 μV/W/m <sup>2</sup>
Leaf wetness	0.05		Leaf Wetness Transmitter/Thies CLIMA	0... 100%	10%
Dew point	0.05		Rain Monitor/Thies CLIMA	0 (No) & 1 (Yes)	Binary

Table 1: Fog climate instrumentations and its features at regional scale. Source: author's elaboration

## 2.2. Study site

In the inland desert, a representative area of ca. 16 km<sup>2</sup> called Oyarbide (20°29'S, 70°03'W) has been selected to analyze and model the spatial gradients and variability of fog Near-surface. The seasonal and diurnal dynamics of fog in this site, which includes the transect of climatological stations OYA\_1128, OYA\_1193 and OYA\_12111, situated proximate to *Cerro Oyarbide*, are intricately interconnected with a complex interplay of regional and local elements, encompassing oceanographic and geographical characteristics that influence its behavior.

The area exhibits a topographical profile characterized by altitudinal gradients ranging from approximately 1,000 m to over 1,200 m. The site is flanked by an east-west oriented corridor that facilitates the inflow of both advective and orographic fog, which can penetrate up to 200 m in elevation, reaching upper limit of Oyarbide (see Figure 5).

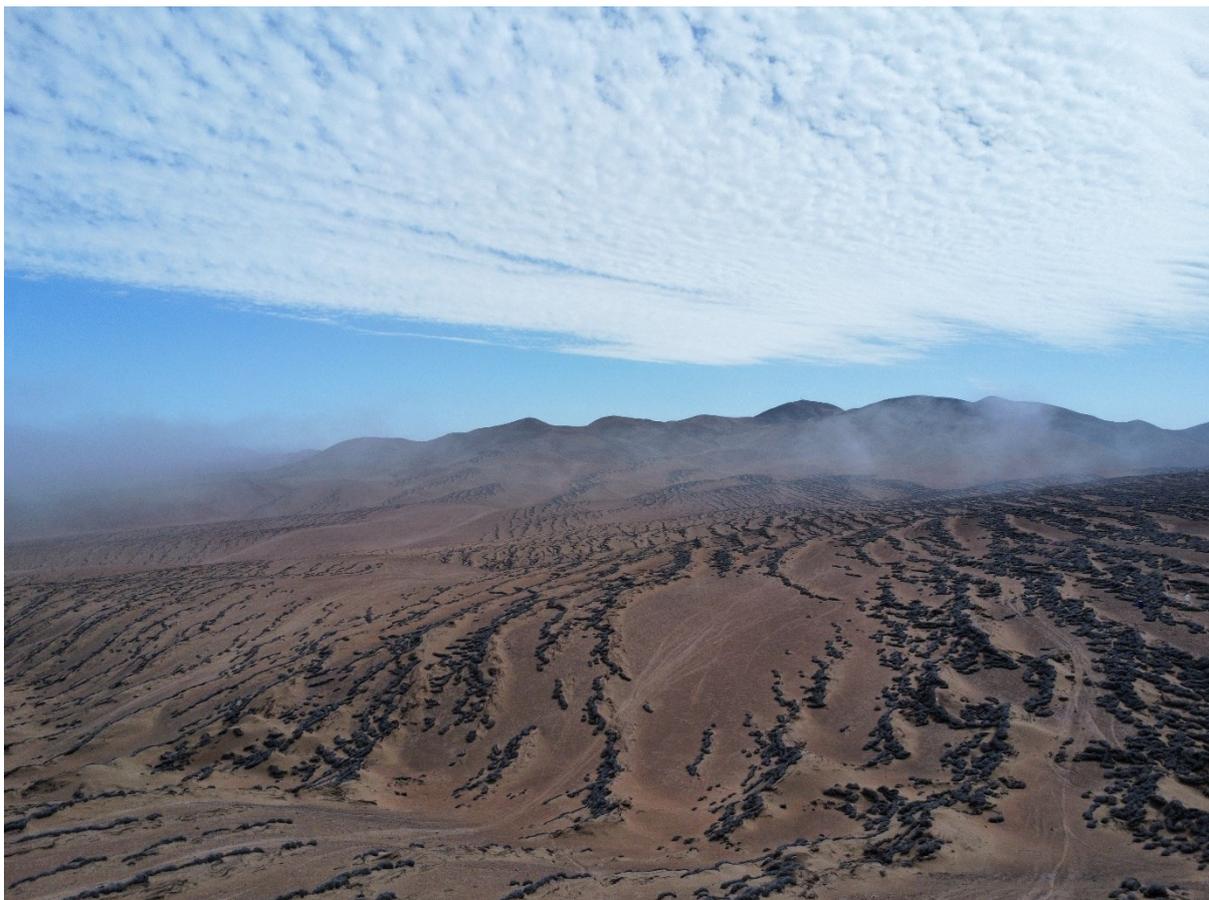


Figure 5: Terrain of the study site “Oyarbide” on a fog event. Source: author’s photographic archive

## 2. Research area

Between the summits of Oyarbide and Soronal hills, a fog corridor is delineated, giving rise to a saddle point that channels the stratocumulus cloud, leading to intermittent fog occurrences (Farías et al., 2005), and comprises one of the most extensive *Tillandsia landbeckii* fields that depends heavily on fog as a primary water source (Cereceda et al., 2007; Westbeld et al., 2009). Oyarbide exhibits the most pronounced variability among *Tillandsia* stands concerning elevation gradients, its topography features an ideal relief characterized by a predominantly uniform incline towards the coast, leading to distinct gradients in stand structure and facilitating modeling and correlation with other climatological variables, especially fog collection (Mikulane et al., 2022).

In 2019 ten mini standard fog collectors (Mini FCs) were installed, covering a localized surface area the study site (see Figure 6). These collectors yield monthly data regarding Near-surface fog water collected at a height of 50 cm above the ground. The first results from modeling Near-surface fog variability reveal a significant correlation with the roughness of the terrain.

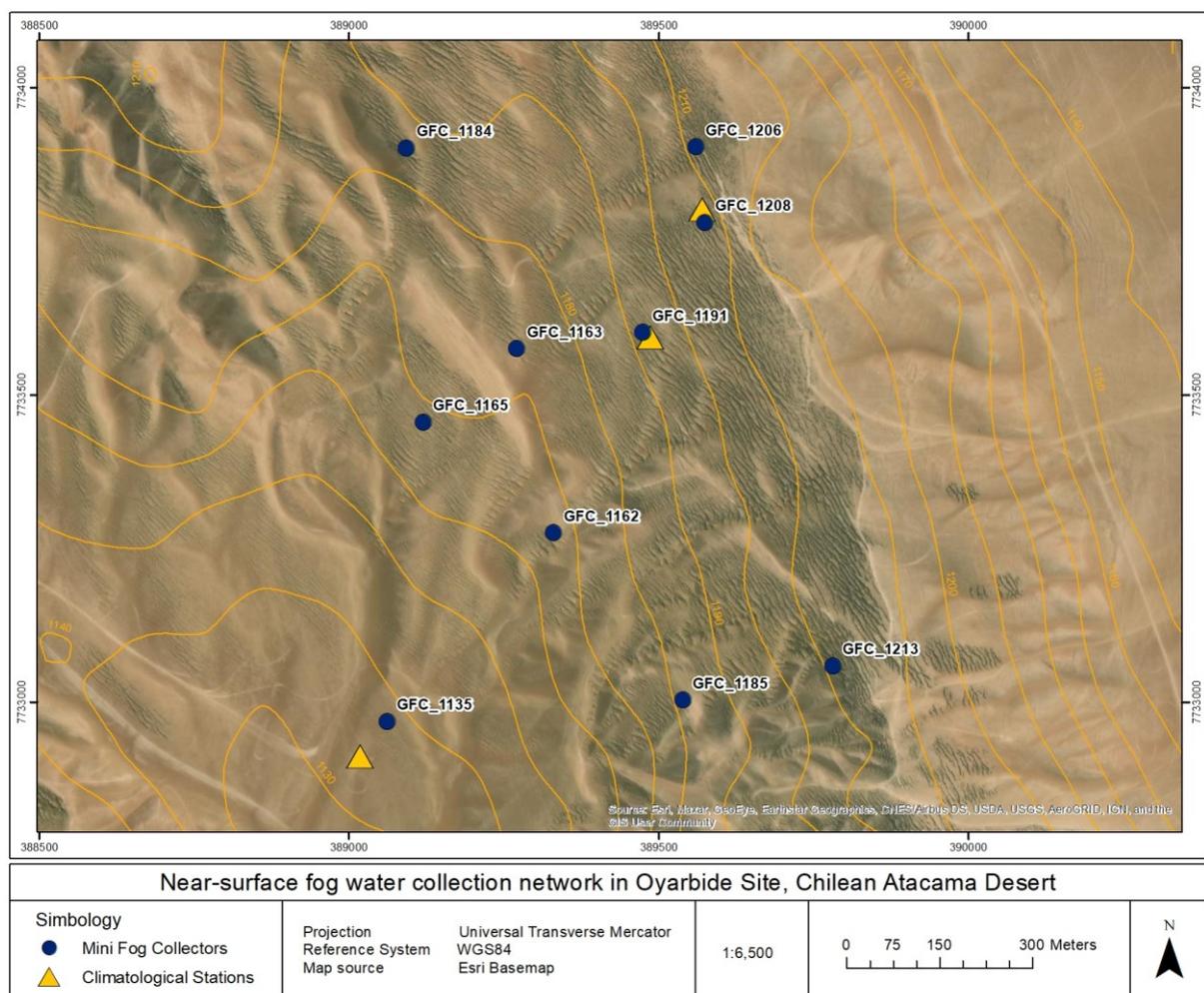


Figure 6: Mini FCs measurements network at local scale in the study site. Source: author's elaboration, Esri ArcGIS Basemap

### 3. State of the art of fog patterns understanding in the Atacama Desert: multiscale exploration and climatic analysis

#### 3.1. Fog research and fog climate measurements background in Chile

Since the 19<sup>th</sup> century, fog has undergone extensive study as a potential natural water resource for human consumption (Fessehaye et al., 2014). In the mid-1980s, several developing countries initiated collecting fog water, with the FogQuest Foundation<sup>2</sup> pioneering the dissemination of knowledge about fog water benefits. Furthermore, the Pontifical Catholic University of Chile has played a pivotal role in the systematic collection and analysis of SFW data within the Atacama Desert. In recent years, monitoring systems like the Satellite-based Operational Fog Observation Scheme (SOFOS)<sup>3</sup>, have enabled interdisciplinary research focused on comprehensively understanding the spatio-temporal variability of fog. Initiatives such as FogNet (FN)<sup>4</sup> in coastal desert regions of the Americas and Africa, the South Africa Science Services Center project for Climate Change and Adaptive Land Management (SASSCAL)<sup>5</sup>, Namib Fog Life Cycle Analysis (NaFoLiCA)<sup>6</sup> in the Namibian Coastal Desert, and the cooperation between the Centro UC Desierto de Atacama (CDA)<sup>7</sup> and the Research Group for Earth Observation<sup>8</sup> from Heidelberg University of Education and University of Heidelberg, collectively seek to advance integrative knowledge on the fog climate of the Atacama Desert and its role in supporting ecosystems that serve as bio-indicators of climate change (see Figure 7).

---

<sup>2</sup> <https://fogquest.org/>

<sup>3</sup> <https://publikationen.bibliothek.kit.edu/1000067688>

<sup>4</sup> [http://the-eis.com/elibrary/sites/default/files/downloads/literature/Monitoring%20the%20lifeblood%20of%20the%20Namib\\_Fog.pdf](http://the-eis.com/elibrary/sites/default/files/downloads/literature/Monitoring%20the%20lifeblood%20of%20the%20Namib_Fog.pdf)

<sup>5</sup> <https://sasscal.org/>

<sup>6</sup> [https://www.imk-asf.kit.edu/english/Projects\\_2885.php](https://www.imk-asf.kit.edu/english/Projects_2885.php)

<sup>7</sup> <https://www.cda.uc.cl/>

<sup>8</sup> <https://www.rgeo.de/de/start/>

### 3. State of the art of fog patterns understanding in the Atacama Desert

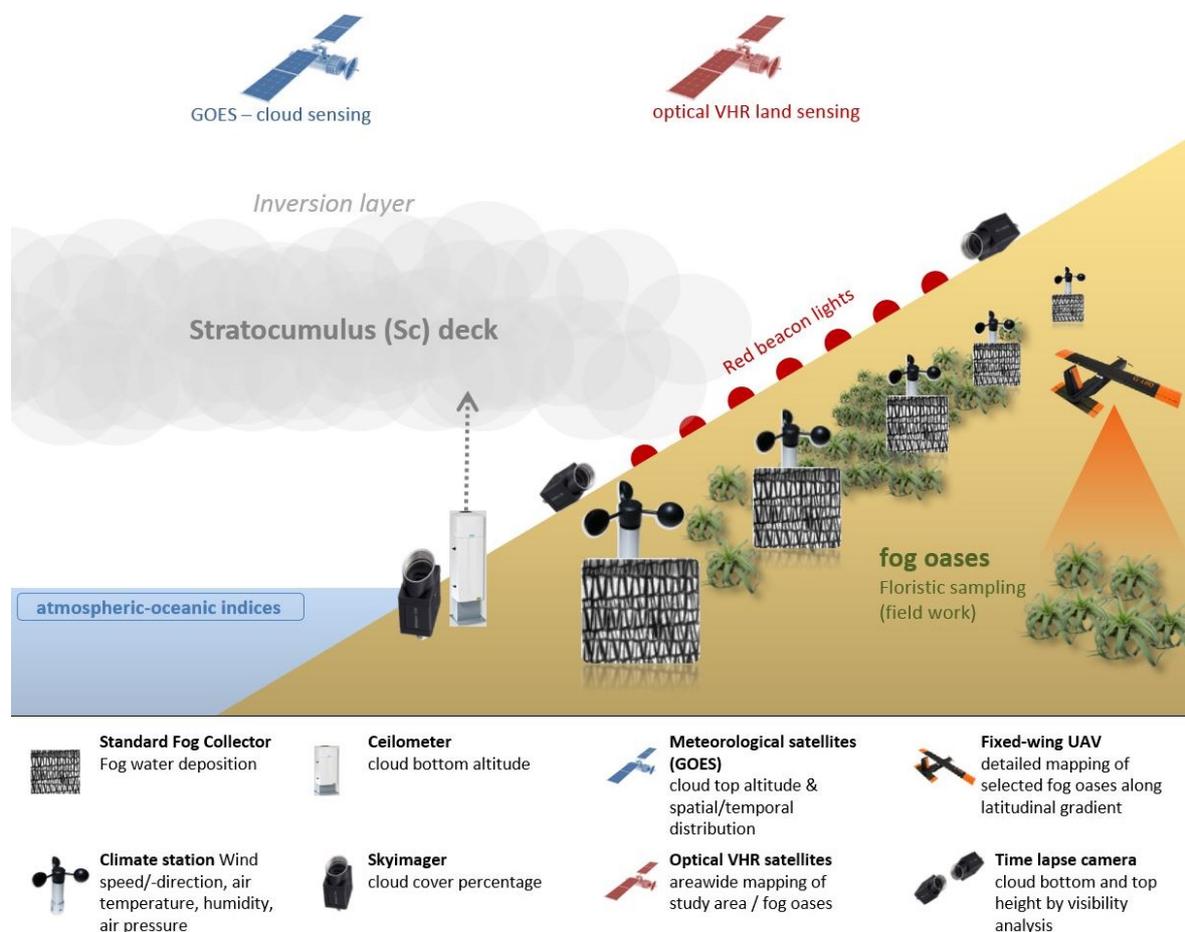


Figure 7: Overview of the methodological approach proposed by 'geo/CDA for researching fog climate and fog ecosystems in the Chilean Atacama Desert. Source: FEBiD Project ref. ELAC2015/T01-0872<sup>9</sup>

Comprehensive research and empirical investigations have revealed that the geographical and topographical characteristics of the Chilean South American coastline create conditions that facilitate the frequent presence of fog for most of the year (Montecinos et al., 2018). The climatic conditions responsible for fog formation in the Atacama Desert are largely driven by the presence of the southeast Pacific anticyclone and the region's prominent topography, the coastal range and the Andean Cordillera (Garreaud et al., 2008). It is influenced by the cold waters of the Humboldt current (Rutllant et al., 2003), and thus also by El Niño-Southern Oscillation (ENSO) and the associated sea surface temperature anomalies in the adjacent ocean (Schulz et al., 2012).

<sup>9</sup> <https://www.eucelac-platform.eu/project/chilean-peruvian-arid-coastal-fog-ecosystems-under-climate-change-understanding-biosphere>

### 3. State of the art of fog patterns understanding in the Atacama Desert

---

A persistent stratocumulus cloud deck forms in the upper part of the marine boundary layer (MBL), producing dense fog where the coastal escarpment interacts with these clouds. The pronounced temperature increase within the inversion layer hampers the vertical development of clouds, while the steep coastal mountain range largely impedes its movement inland (Schulz et al., 2012) (see Figure 8). This ocean-atmosphere system gives rise to a dynamic marine advection fog, providing moisture to a hyper-arid environment and allowing the development of ecosystems and high biodiversity locally called *vegetación de lomas* or fog oases (Cereceda et al., 2008b; Del Río et al., 2018; Garreaud et al., 2008).

A dominant land-sea breeze system prevails in the study area, extending from the coastal zone towards the interior desert, reaching the *Tillandsia* Lomas fields and further into the desert (Miller, 1976). Generally, this local wind system generates orographic fog that develops at the coastal cliff whereas radiation fog occurs further inland (Cereceda et al., 2002).

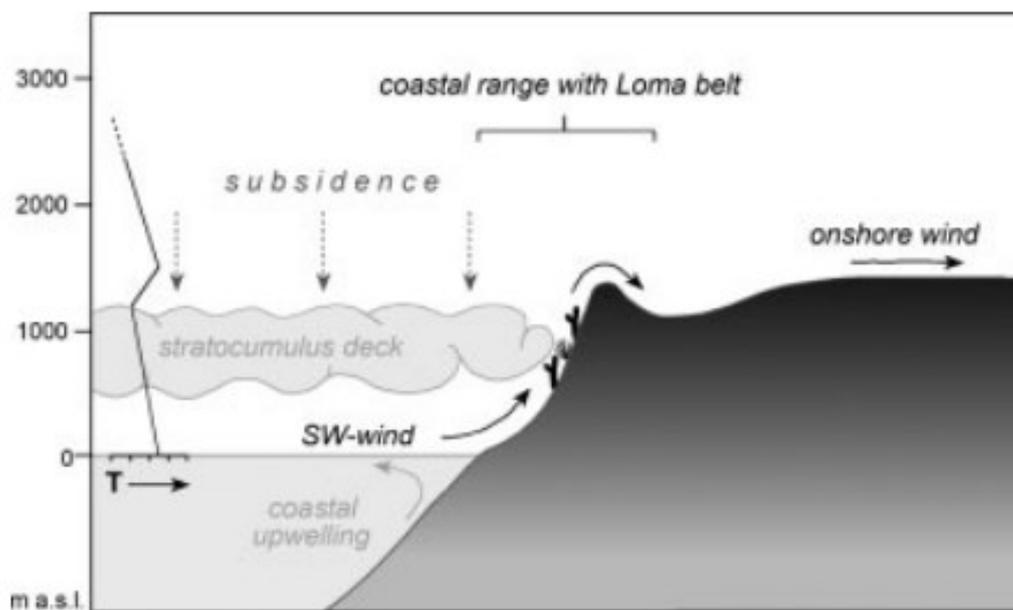


Figure 8: Fog formation and its ocean-continent movement. Source: Schulz et al. (2012)

In northern Chile, particularly in the *Norte Grande*, the meteorological stations have limited and sparse data records, with sixty climate stations, mostly belonging to the Chilean General Directorate of Water (DGA), and six corresponding to the Chilean Meteorological Office (DMC) (Sarricolea et al., 2017). On regional level, climatological measurements and observations have been recorded at Diego Aracena Airport of Iquique (20°32'57"S 70°10'52"W) from 1981 until nowadays. The initial studies on fog

collection in the Atacama Desert were conducted in the 1960s, when a team of physicists from the University of Antofagasta began to experiment with forms of collecting SFW following the Grunow Method (Grunow, 1952), which set the foundation for the development of modern Standard Fog Collectors (Schemenauer & Cereceda, 1994b). During the 1980s, the Institute of Geography of the Pontifical Catholic University of Chile carried out initiatives for fog water collection and measurement for some coastal towns and since 1997, through the Centro del Desierto de Atacama – Alto Patache, has recorded fog data and driving factors up nowadays. However, there is a knowledge gap about fog weather in its interaction of climate variables with topographic parameters and layer inversion height, at different scales, particularly in a region with one of the largest Tillandsias fields, where the variability of the fog in these environments responds to a complex biosphere-atmosphere interaction framed in a complex topographic landscape.

#### 3.2. Climate data management: quality and control

High-quality and homogenous long-term data series play a pivotal role for climatological research. The availability of comprehensive historical climate datasets spanning extended periods, coupled with data homogeneity, stands as paramount for meeting the exigencies of climate research endeavors. Such datasets not only serve as indispensable prerequisites for conducting reliable and representative research but also support the credibility and robustness of climatological investigations (WMO, 2011). However, a significant portion of long-term climatological data series contains outliers, missing values, and inhomogeneity (Cao & Yan, 2012; Trewin, 2013).

In estimating climatological data, various methods are employed. From advanced technologies such as the use of satellites, statistical programs, and hydrological modeling (Smith et al., 1998), which, due to their high cost, are not massively used, to traditional statistical methods, which are easier to access and widely used. The latter are based on simple mathematical formulas, such as simple averages, multiple linear regressions, or time series models (Kim et al., 2007), where relationships are established between standard or nearby stations and the station with a lack of climatic information.

To enhance the estimation of missing data and outliers within a temporal series of values, statistical procedures exist that incorporate the physical properties inherent to the climatological station or climate network under investigation. These procedures aim to yield realistic and coherent values consistent with the underlying physical principles (WMO, 2008).

In general, the necessity for implementing a QC process can be attributed to three key factors (Doraiswamy et al., 2000). Firstly, such a system ensures the proper generation of information. Secondly, it serves to identify and rectify erroneous records, thereby averting potential misinterpretations and erroneous decision-making. Lastly, it facilitates the timely detection of issues, enabling prompt resolutions through maintenance, repair, and calibration of climate sensors.

The World Meteorological Organization (WMO) Guide to Climatological Practices (WMO, 2011) proposes various methodologies employed for filling missing data or outliers in climatological databases.

The Quality Control (QC) procedures are imperative for the various phases of the data collection process. However, comprehensive QC measures are lacking across all levels (Zahumensky, 2004). The QC procedures widely employed worldwide are rooted in meticulously formulated general protocols extensively documented across various WMO guidelines and publications, particularly catering to synoptic networks and climate data. The different manuscript of Practices, Observing, Instrumentation, Methods, and Processing System Guides, collectively represent the basic literature of the WMO<sup>10</sup>, delineating the international benchmarks for data QC.

Numerous techniques have been developed for detecting irregularities and adjusting them in climatic datasets. These include the Buishand range test (Buishand, 1982), Kruskal-Wallis test (Kruskal & Wallis, 1952), Mann-Kendall test (Kendall, 1975; Mann, 1945), Multiple Analysis of Series for Homogenization (Szentimrey, 1999), Pettit test (Pettitt, 1979), Regression-Based methods (Easterling & Peterson, 1995; Lund & Reeves, 2002), and the Standard Normal Homogeneity Test (Alexandersson, 1986), among others.

The Normal Ratio calculation (RN) is a non-geostatistical method widely used to fill in missing non-continuous climatological data; it estimates erroneous or missing values at the climatological station under review by employing a weighted average of stations within the homogeneous hydrological area (Paulhus & Kohler, 1952).

A widely accepted proposal is based on a contribution by Zahumensky (2004) who proposed classification system of anomalies in a database based on internal consistency between stations and/or in time, differentiating between random errors, which exhibit a symmetric distribution around zero, systematic errors, characterized by an asymmetric distribution and primarily caused by equipment malfunctions or data processing errors, and micro-meteorological errors, representing inconsistencies in local records compared to surrounding regions.

---

<sup>10</sup> <https://library.wmo.int/>

In the process of homogenization, irregularities identified within a time series using various techniques are subsequently adjusted to render the dataset homogeneous (Trewin, 2013). Although a wide quantity of techniques exists, no single procedure is universally recommended. Instead, the choice of method depends on the specific characteristics of the data series and the nature of the anomalies recorded.

It is imperative in the process of identification and homogenization within a climatological database that statistical procedures incorporate the physical properties of the measurement station under study, facilitating the derivation of realistic and coherent values consistent with physical principles (WMO, 2011).

#### 3.3. Fog cycles in the northern Chilean Atacama Desert

Numerous studies on fog occurrence and water availability have been conducted at various Atacama Desert places (Cereceda & Schemenauer, 1991) and have shown that fog drifts inland and cover the Coastal Range up to ca. 25 km, and occasionally reach as far as 50 km, reaching the *depression intermedia* (Farías et al., 2005) (see Figure 9). Fog frequency is generally highest in the coastal areas, peaking between 35° and 40°S with approximately 38 fog days per year. However, frequencies with 189 days of fog are particularly high at the more elevated parts of the coast. Requiring dense networks of observations, fog water availability in the Atacama Desert is a complex function of the topography, seasonal wind patterns and trade inversion height, which partially contribute to patchy small-scale fog oases (Lehnert et al., 2018).

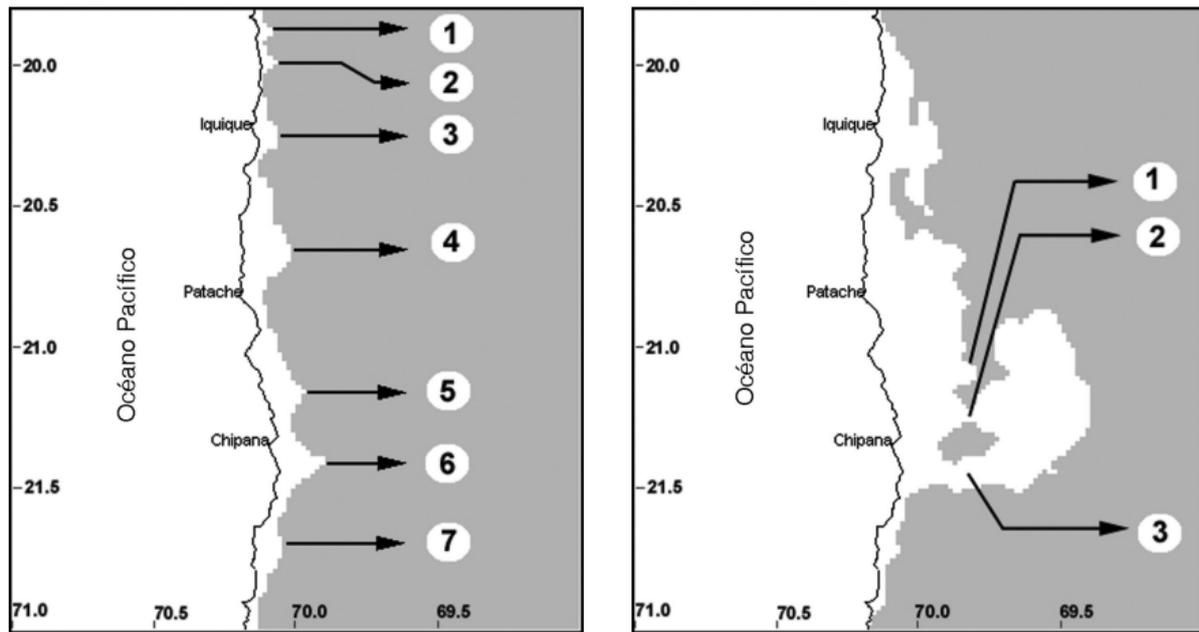


Figure 9: Fog penetration corridors in the Tarapacá Region. Source: Farías et al. (2005)

Fog data showed significant seasonal characteristics and interannual variability in the coastal Chilean Atacama Desert. In austral winter, 90% of the fog over land is of advective origin, while 10% is orographic. The situation changes in austral summer when there is less fog and orographic events are more frequent than advective events (Farías et al., 2005). High fog frequency at higher elevations, however, is usually a product of orographic cooling (Lange, 2003). The presence of stratocumulus clouds and fog represents a well-defined annual cycle with less presence from December to April and greater presence from May to November (Cereceda et al., 2008a, 2008b; Farías et al., 2005). On average, SFW yields based on Standard Fog Collector measurements between  $0.8$  to  $7 \text{ L m}^{-2} \text{ day}^{-1}$  (Cereceda et al., 2008a; Larrain et al., 2002). Likewise, there is a predominant daily cycle, with higher fog presence during the night and dawn, while fog water is generally dissipated and evaporated by solar radiation during the day (Cáceres et al., 2007; Larrain et al., 2002; Osses et al., 2017).

Despite these investigations, there is still limited understanding of spatial and inter-annual variability of fog clouds and their associated water content, especially on a local level. The reasons for the spatiotemporal variability of fog and the relationship between this variability and the Southeast Pacific oceanographic and atmospheric indices remains mostly unknown (Del Río et al., 2018).

High temporal (daily and hourly data) and spatial (local scale) climatological data, based on standardized high-quality sensors for the analyses of fog climate are not available in the Atacama Desert due to sparse observation networks and the limited systematization of existing data.

Consequently, it is practically impossible to carry out a reliable and spatially coherent analysis of fog distribution based only on existing surface observation data (Bendix, 2002).

#### 3.4. Multivariate analysis on topographic and climatic influences on fog variability

Topographic features such as saddlebacks, depressions, fog-corridors, aspect and elevation above sea level affect the local geographic distribution, frequency and water-content of fog and hence the distribution of fog ecosystems (Osses et al., 2007; Westbeld et al., 2009). At a local/stand level, *Tillandsia Lomas* are known to form distinct spatial patterns which provide indication on the local fog-water inputs and their relation to vegetation density and vitality (Borthagaray et al., 2010). The isolated spatial distribution of fog in these extremely arid conditions provides a test bed for observing, analyzing, and modeling biosphere-atmospheric dependencies and relating them to meso-climate regimes and changes on a global scale (Vautard et al., 2009).

Regression analysis is the most widely used statistical technique for examining and modeling the relationships between variables (Montgomery et al., 2021), and has a widely application in almost every domain, including ecological and remote sensing studies (Foody, 2003).

Miller (2012) emphasizes the OLS model as a 'global' method that assumes spatial stationarity, positing constant parameters across geographic space. This approach, reliant on a singular equation, dominates ecological applications for parameter (Berterretche et al., 2005; Zhang & Shi, 2004). However, the OLS model remains a-spatial, disregarding geographical information in parameter estimation, thus averaging all parameters across the dataset (Fotheringham et al., 2002).

Key assumptions underpinning OLS regression include normality, homogeneity, and independence of residuals (Montgomery et al., 2021), and the violations of these assumptions yield biased estimators, compromising parameter accuracy (Fox et al., 2001; Montgomery et al., 2021). Furthermore, the assumption of spatially random observations and invariant relationships between variables across geographical locations in OLS models masks local characteristics, portraying global results as average effects across diverse regions.

Addressing the limitations posed by spatial stationarity, the GWR emerges as a solution to address spatially autocorrelated error terms in OLS models (Wang et al., 2005; Zhao et al., 2010). OLS models, rooted in spatial stationarity, universalize derived parameters across sampled space, overlooking local variations (Foody, 2003). Consequently, the adoption of local analyses becomes imperative to unveil

nuanced spatial realities, diverging from the homogeneity espoused by global analyses (Foody, 2003; Fotheringham et al., 2002) (see Figure 10).

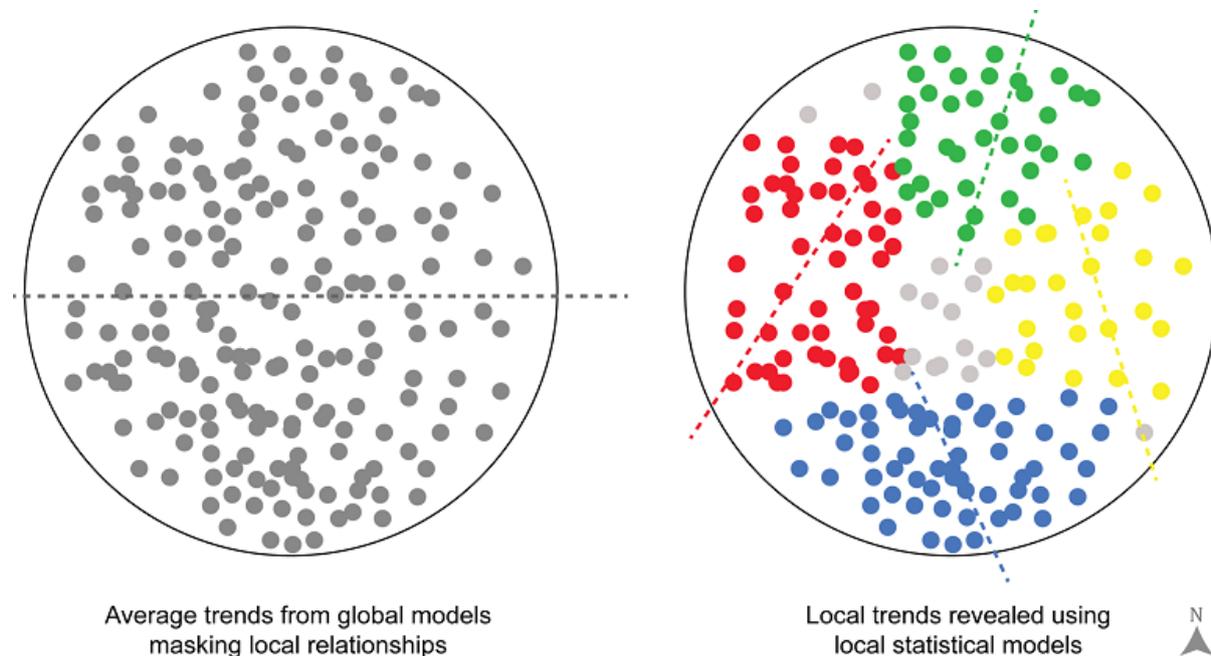


Figure 10: Average trends using global statistical model (left) masking local trends revealed in local statistical model (right). Source: Sachdeva & Fotheringham (2020)

The GWR, introduced in geographical literature by Brunson et al. (1996) and developed theoretically by Fotheringham et al. (2001) investigates spatial variability of regression model correlations, addressing parametric nonstationarity. Spatial heterogeneity, stemming from diverse environmental processes operating at various scales (Legendre, 1993), challenges models assuming uniformity across space. Fotheringham et al. (2001) propose that GWR model leverages local and variable parameter regression techniques, integrating spatial location information into the regression framework. Utilizing local smoothness via distance-weighted regression points and surrounding observations based on geographical positions (Zhou et al., 2019), GWR emerges as a potent tool to account for spatial heterogeneity, notably in climate research.

Operational within local windows centered on a regular grid, GWR employs a spatial kernel to weight observations, affording higher influence on spatially proximate data points (Windle et al., 2010). The bandwidth of this kernel, whether fixed or adaptive, crucially determines the extent of influence,

impacting parameter estimates and the scale of relationships between variables (Brunsdon et al., 1996; Osborne et al., 2007; Windle et al., 2010).

Autocorrelation, widely observed in ecological variables across spatial and temporal domains (Legendre, 1993; Legendre, P., & Legendre, L., 1998), underscores the necessity of accounting for spatial autocorrelation and spatial non-stationarity in modelling research (Brunsdon et al., 1996; Fotheringham et al., 2002; Fotheringham et al., 1998). Spatial non-stationarity, inherent in natural environments due to dynamic biological and large-scale physical processes (Fernandes & Leblanc, 2005; Frescino et al., 2001; Legendre, P., & Legendre, L., 1998), demands models capable of accommodating such intricacies for enhanced predictive capabilities (Foody, 2003; Zhang & Gove, 2005).

As a local regression method, GWR extends global regression principles, employing local statistics to link spatial autocorrelation and heterogeneity across space, aligning with Tobler's first law in Geography (Tobler, 1970), emphasizing the significance of spatial relationships.

## 4. Research objectives, methodological concept, and methodology

### 4.1. General and specific objectives

Several studies have been conducted on the variability of fog water collection and about dynamics of fog and its role with fog ecosystems on the place. However, the understanding of the driving parameters and characteristics of fog climate at regional and local levels remains limited. This knowledge gap persists due to constraints such as the scarcity and low-level quality of available data, alongside the diverse temporal dynamics and spatial gradients of fog climatology.

This study aims to contribute to the understanding of the atmosphere-biosphere interaction in fog ecosystems by employing a multiscale data collection network in the Atacama Desert (see Figure 11). Specifically, this research seeks to determine the variability and availability of SFW through regional and local scale fog climate modeling. The expected outcome of this study is to generate new knowledge that can inform the ecological niche of *Tillandsia landbeckii* fog ecosystems, serving as a bioindicator for SFW availability and the potential impacts of climate change in hyperarid regions like the Atacama Desert. The insights gained from this research are expected to contribute significantly to the study of fog climatology and provide valuable guidance for the management and conservation of ecosystems in arid regions.

A methodological design is proposed which is determined by the overcome of three secondary objectives detailed below:

a) Implementing a data quality flow aimed at enhancing the reliability and precision of the climatic data series originating from the fog climate network established in the Atacama Desert since 2016. This flow is based on a rigorous process through programming flows that identifies and record the gaps, non-numeric values, and outliers. Subsequently, filling or correcting the recorded data will be conducted, thereby ensuring the integrity and utility of the climatic datasets.

b) Conducting a comprehensive analysis of the climatological data from the fog climate network in the Atacama Desert, following the implementation of a data quality workflow. This analysis will specifically focus on comprehending the fog climate across its climatic gradient from the coastal regions until ca. 11 km inland into the desert, up to an elevation of about 1,354 masl. A detailed temporal analysis will be performed to scrutinize fog variability in the region through its SFW collection records from the SFC.

#### 4. Research objectives, methodological concept, and methodology

c) Conducting a comprehensive analysis of the spatial distribution of Near-surface fog at the Oyarbide study site based on the record of Mini FCs network and its correlation with the local topography. This will be achieved through Geographic Information System (GIS)-based regression analysis, which will assess the relationship between Near-surface fog water variability and topographic features. The OLS and GWR models will be compared to determine the best fit. By utilizing disaggregate statistics to test and model this relationship, the pattern of association can be visually represented through cartographic techniques, with statistical values spatially represented on raster maps and plots.

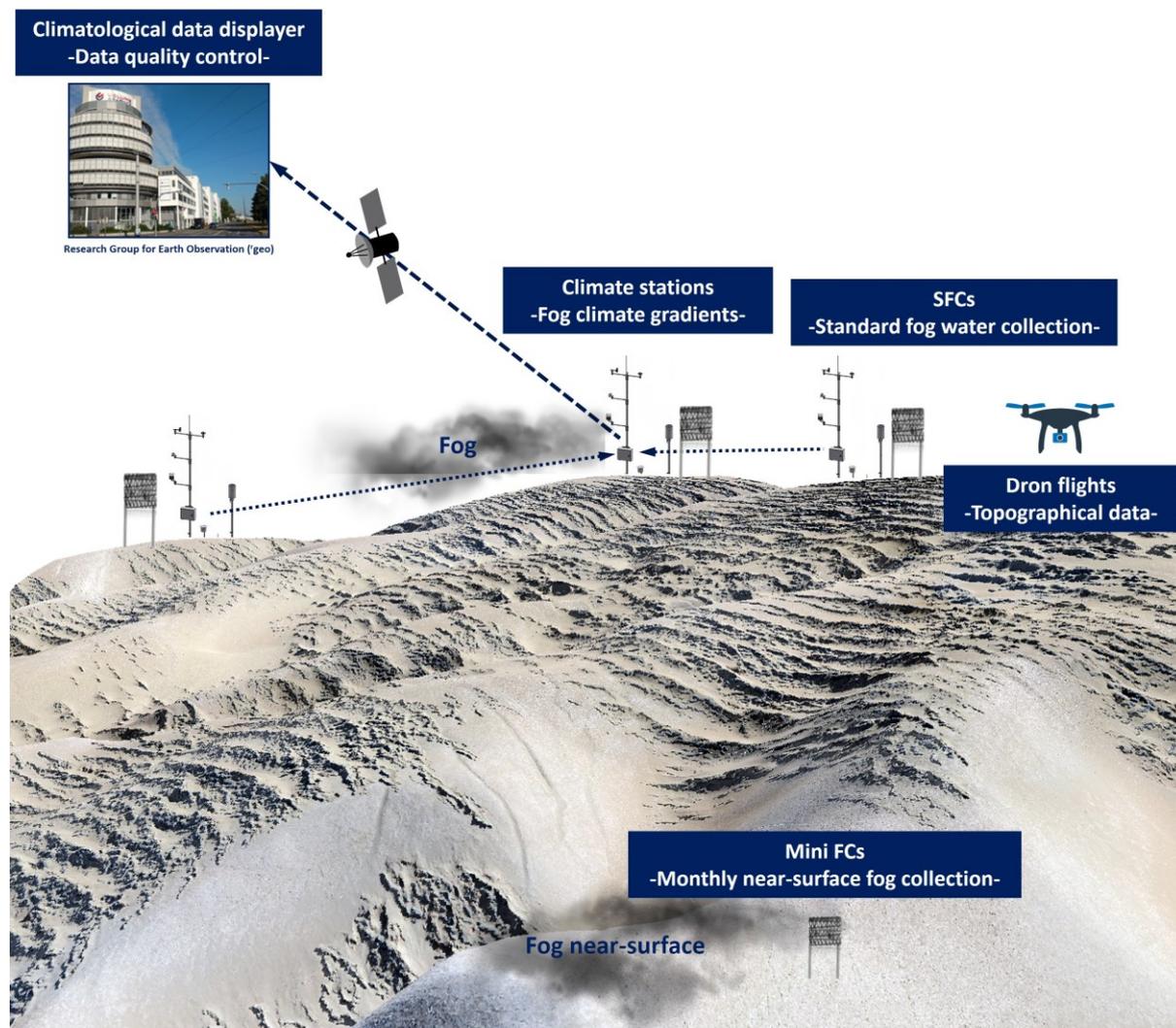


Figure 11: Multiscale data sources and measurement approaches for analyzing fog variability.  
Source: author's elaboration

## 4.2. Research and design methods

The proposed research adopts a multi-scalar, integrative analytical approach—both temporal and spatial—in conjunction with an inductive methodology, as outlined in stages in Figure 12.

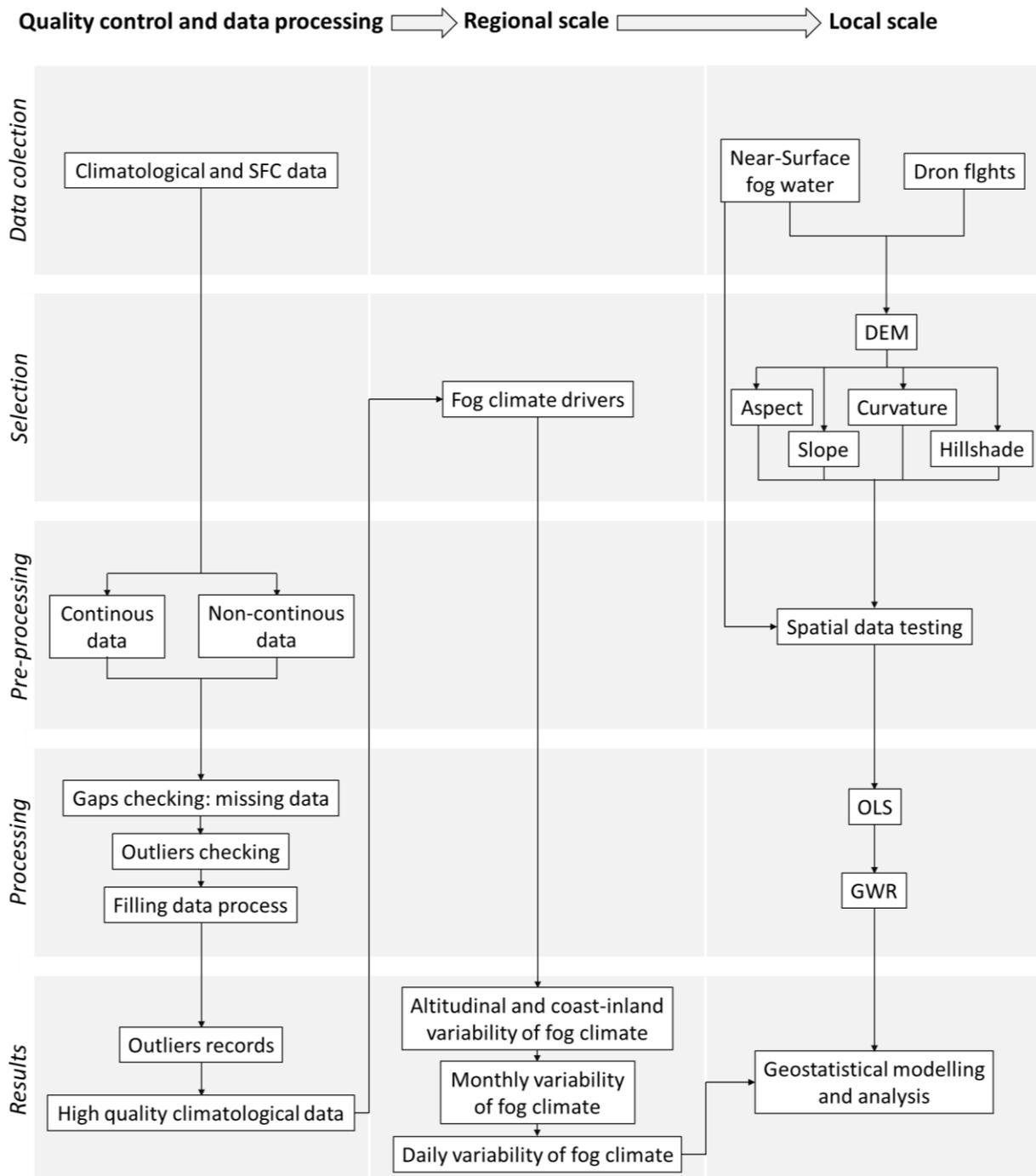


Figure 12: Methodological framework research on fog variability patterns in the Atacama Desert. Source: author's elaboration

### 4.2.1. Quality control

The first stage "Quality control and data processing" seeks to homogenize and systematize the climatological data to eliminate non-climate factors in order that the data variability is strictly a reflection of the regional-local climatological variability. To establish a homogeneity of the climatological data series, considering the recommendations of the WMO, a non-parametric test will be applied, called the sequence method. With this test the number of data sequences will be evaluated in their temporal order whose record must be included between the critical values of the sample distribution table. This process will allow to determine the homogeneity of the data series when these vary according to the natural climatic factors that affect regional meteorology, also if the environmental conditions remain constant and the climate sensors are in good condition. If these conditions are not met, it will be necessary to delete the anomalous data from the data series, in which all the fluctuations contained in their time series reflect the variability and trends of the climatic variables.

The process of data filling or correction is divided into three distinct intervals. For intervals characterized by short durations, specifically those containing six or fewer climate records, each of which spans 10 minutes, the arithmetic mean method is employed to estimate missing data. In this approach, the estimation aligns with the model  $y = \alpha$ , wherein "y" denotes the estimated value for the missing data, and " $\alpha$ " signifies the mean derived from observations of the variable of interest before and after interval. For middle and long-term intervals, the Neighboring Station Interpolation Method, initially proposed by Paulhus and Kohler (1952) and referenced by Searcy and Hardison (1960), is applied. This method estimates climatic variable data by averaging observations from neighboring stations that experience the same biosphere-atmosphere influence during the respective period.

The start of data processing for quality control dates from the start of each climate station's transmissions until June 30<sup>th</sup>, 2023, at 23:50 hr.

### 4.2.2. Regional analysis

The measured data from climatological stations and SFC reveal the temporal and spatial dynamics of the fog climate variability from coast to inland desert, over the Coastal Cordillera.

The coastal-to-interior gradient of the desert, along with the altitudinal gradient, is represented by the transect of climatic stations from OYA\_518 to OYA\_1211 at the innermost point of the desert and

#### 4. Research objectives, methodological concept, and methodology

---

from OYA\_518 to OYA\_1354 at the highest altitudinal point. The climate data considered on this transect are the primary driving parameters upon which fog water collection predominantly depends (Montecinos et al., 2018) and are essential for characterizing the coast-inland and altitudinal gradients. Considering that OYA1211 has the most quantity of sensors, was installed farthest from coast and at the top of a *Tillandsia landbeckii* field, has relevance in the fog measurement network and will be considered for deeper analysis (see Figure 13).



Figure 13: Climatological station OYA\_1211 and SFC at the upper limit of Oyarbide. Source: author's photographic archive

Analysis have included a multitemporal approach (see Figure 14), considering main drivers as air temperature (AT), relative humidity (RH), wind speed (WS), wind direction (WD), and standard fog water (SFW), as well as relevant variables for the inland desert analysis, considering the penetration of oceanic humidity conditions and the variability of the TIL, such as leaf wetness (LW) and dew (DP). While the primary emphasis will be placed on the driving climate data previously mentioned, a

#### 4. Research objectives, methodological concept, and methodology

comprehensive analysis of all climatic data provided by the network of stations has been conducted to achieve a holistic characterization.

The start of data processing for regional analysis dates from the start of each climate station's transmissions until June 30<sup>th</sup>, 2023, at 23:50 hr.

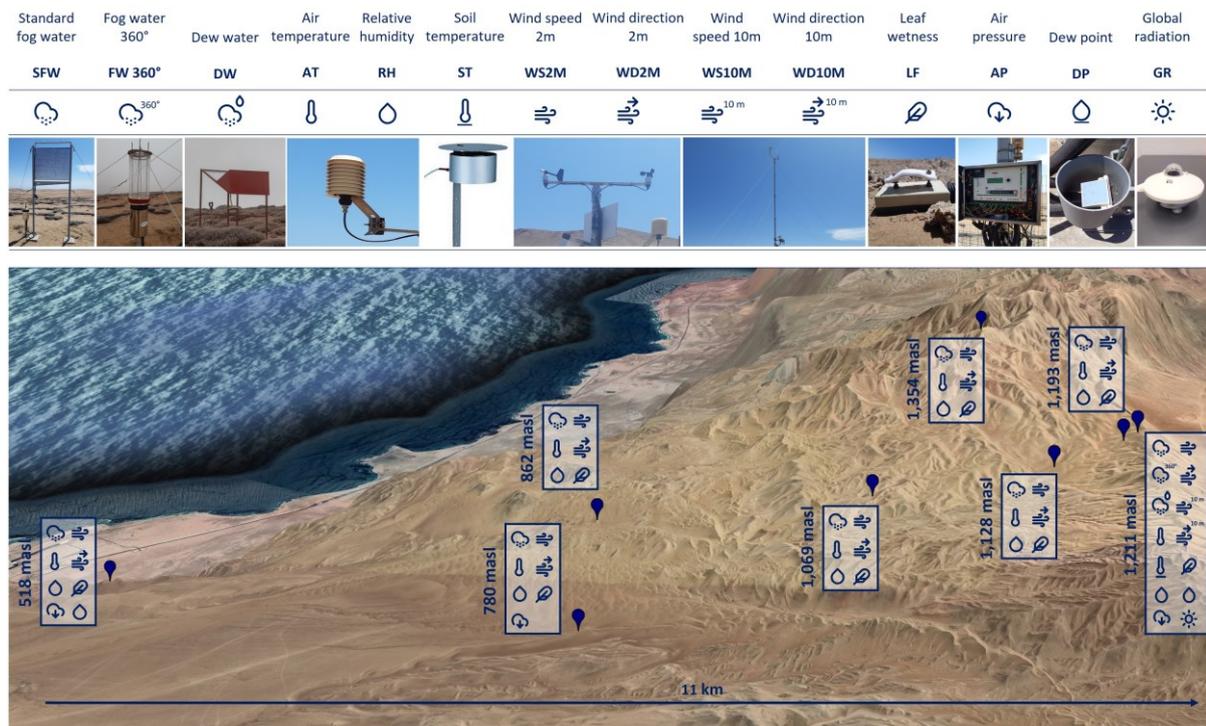


Figure 14: Fog climate network and its measurement sensors. Source: author's elaboration

#### 4.2.3. Local analysis

In the Oyarbide Site, a complex terrain and home to one of the region's largest fields of *Tillandsia landbeckii* of the region, it's been conducted Near-surface fog water measurements using a network of Mini FCs with the primary aim of analyzing and assessing the geospatial relationship between Near-surface fog water variability and the topographical characteristics of the study site.

A Mini FC is a passive collector of fog water positioned near the surface, based on the (Schemenauer & Cereceda, 1994b) method, which employs a "Raschel" type mesh with 35% shade, placed in two layers at a height of 50cm from its base. The collected fog water is stored in a 60L tank, linked to the device through a plastic hose (see Figure 15). These measurements have been conducted monthly since February 2019 and continue until nowadays.

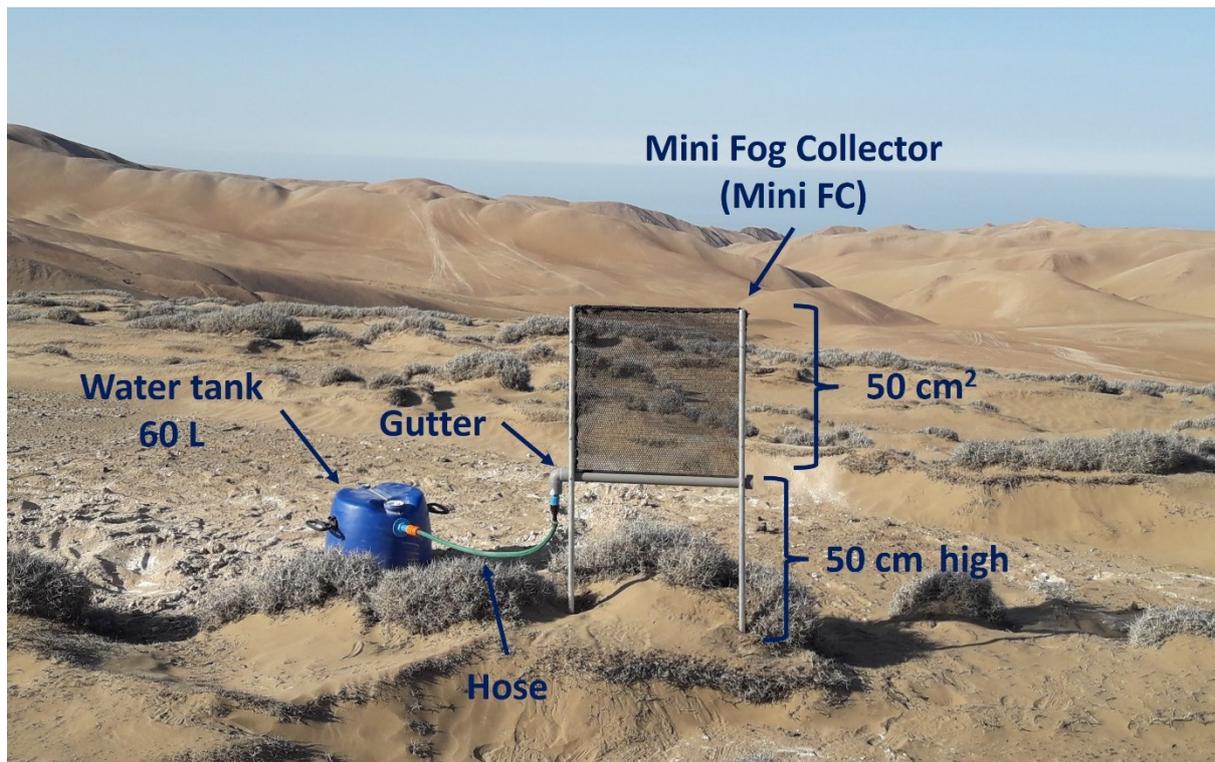


Figure 15: Structure and components of a Mini FC. Source: Elaboration by the author based on the author's photographic archive

To gain insights into the variability of Near-surface fog water, which constitutes the focal dependent variable within the local analytical framework, a carefully designed monthly field data collection protocol was implemented on-site at Oyarbide. The fog water collected by the Mini FCs is measured monthly in-situ using a milliliter-scale (with a margin of error of 20 ml) to ensure accurate measurement quantification. Subsequently, this data is transmitted to the central data repository for systematic organization.

To gather data for the topographical explanatory variables, a Digital Elevation Model (DEM) has been made from the results from the processing of ca. 6800 aerial photos captured during a flight campaigns November 2018 in the study site by a consumer digital SLR Camera DJI Zenmuse X5S Gimbal stabilized mounted on a DJI Matrice 200 UAV/Drone, with a height set to 90 m and draped 1.42 cm/pix resolution. By geoprocessing tools from the DEM were extracted the topographical factors analyzed to potentially predict the Near-surface fog water variability: Altitude, Slope, Aspect, Curvature and Hillshade. Given that the Aspect variable is circular in nature (ranging from 0 to 359°), it requires transformation into incremental values based on trigonometric functions to northness and eastness (Olaya, 2009; Roberts, 1986), as  $northness = \cos(Aspect)$  and  $eastness = \sin(Aspect)$ .

#### 4. Research objectives, methodological concept, and methodology

---

Based on the aforementioned information, global and local regression models were developed using the ESRI ArcGIS spatial modeling module. The dependent variable was Near-surface fog water, while the explanatory variables included Elevation, Aspect, Slope, Curvature, and Hillshade.

The OLS constitutes a widely employed global regression method that endeavors to establish a single equation elucidating the relationship between variables. This method assumes homogeneity, presuming a consistent relationship across the entire study area, thereby assuming stationarity.

The formulation of applied OLS to the research case can be expressed as:

$$\text{Near-surface fog water} = \beta_0 + \beta_1 * \text{explanatory 1} + \beta_2 * \text{explanatory 2} + \dots + \beta_n * \text{explanatory n} + \varepsilon$$

Here's the breakdown:

**Near-surface fog water** is the response variable to be estimated.

$\beta_0$  is the intercept term.

$\beta_1, \beta_2, \dots, \beta_n$  are the regression coefficients that indicate the relationship between each of the explanatory variables and the response variable.

**explanatory 1, explanatory 2, ..., explanatory n** are the variables used to predict the response variable.

$\varepsilon$  is the error term, which represents the variability in the response variable that cannot be explained by the explanatory variables.

In contrast, GWR deviates from global regression techniques by computing distinct equations for individual elements, such as pixels in the context of a dependent variable dataset. This model introduces the estimation of local parameters, diverging from uniform global parameters (Fotheringham et al., 2002). By integrating spatial coordinates into the analysis, the GWR discerns spatial disparities and relationships between environmental determinants and factors through localized estimations (Abdul-Rahim et al., 2022).

The formulation of applied GWR regression model is as follows:

$$\text{Near-surface fog water}_i = \beta_0(u_i, v_i) + \beta_1(u_i, v_i)x_{1i} + \beta_2(u_i, v_i)x_{2i} + \dots + \beta_k(u_i, v_i)x_{ki} + \varepsilon_i$$

Here's the breakdown:

**Near-surface fog water<sub>i</sub>** signifies the observed value of the response variable at the i-th geographic location.

$x_{1i}, x_{2i}, \dots, x_{ki}$  denote observed values of k explanatory variables at the same i-th location.

$\beta_0(u_i, v_i), \beta_1(u_i, v_i), \beta_2(u_i, v_i), \dots, \beta_k(u_i, v_i)$  stand for spatially variable regression coefficients, contingent on the spatial coordinates  $(u_i, v_i)$  of the i-th location.

$\epsilon_i$  represents the error term, capturing unaccounted variability in the response variable at the i-th geographic point.

The GWR model employs an Adaptive Kernel with spatial variability, adjusting the weighting of regression and neighboring observation points based on their geographical proximity. The model proceeds by iteratively estimating parameters using the local-weighted least squares method fortifying the model's capacity to accommodate spatial disparities (Zhou et al., 2019).

The start of data processing for local analysis dates from the start of Near-surface fog measurements network until May 2023.

## 5. Fog variability characterization in the Chilean Atacama Desert: fog climate data quality assurance and multiscale spatio-temporal analysis

### 5.1. Quality control and data processing

#### 5.1.1. Data quality flow

Application of stringent quality control procedures to data series assumes pivotal importance. This chapter delves into the key processes of data validation, aggregation, rectification, and deletion, which together form the bedrock of ensuring data integrity. An effective quality control system requires a comprehensive approach that traces data back to their sources, thereby safeguarding their accuracy and proactively averting the recurrence of errors whenever possible. This data quality control examination is particularly pertinent within the context of the fog climate measurement network situated in the Atacama Desert, where the imperatives of precision and accuracy loom large, serving as the linchpin for gaining deep insights into the intricate dynamics characterizing the fog climate system in hyper arid environments.

To ensure the quality of fog climate data, a proposed a three-part flow of quality control procedures that should be applied after the initial data transmission/entry process as shown in Figure 16. This process starts with the reception of data in the central repository, serving as a fundamental cornerstone for subsequent analyses and research endeavors. The subsequent step, called as "gaps checking," focuses on identifying and addressing any missing data and non-numeric values within the dataset. This data review ensures data completeness and integrity. Following this, the data undergo a filtering process referred to as "outliers checking." This stage differentiates between continuous and non-continuous climatic data, directing each through specific pathway. This categorization allows for identification and tailored data correction based on its climatic variability. These checking processes are supported by carefully designed programming flows, the outliers identified are documented in a database table and graphical representation in every filter, ensuring efficiency and precision throughout its execution. After completing the identification stages, the process advances to the "filling data process". Here, the data intervals are classified into short, middle, and long, each category requiring specialized treatment to ensure coherence and consistency. This process culminates in a

highly reliable dataset. These data represent the outcome of a rigorous flow, thoroughly examined, and curated to serve as the foundation for credible and insightful climatological analyses.

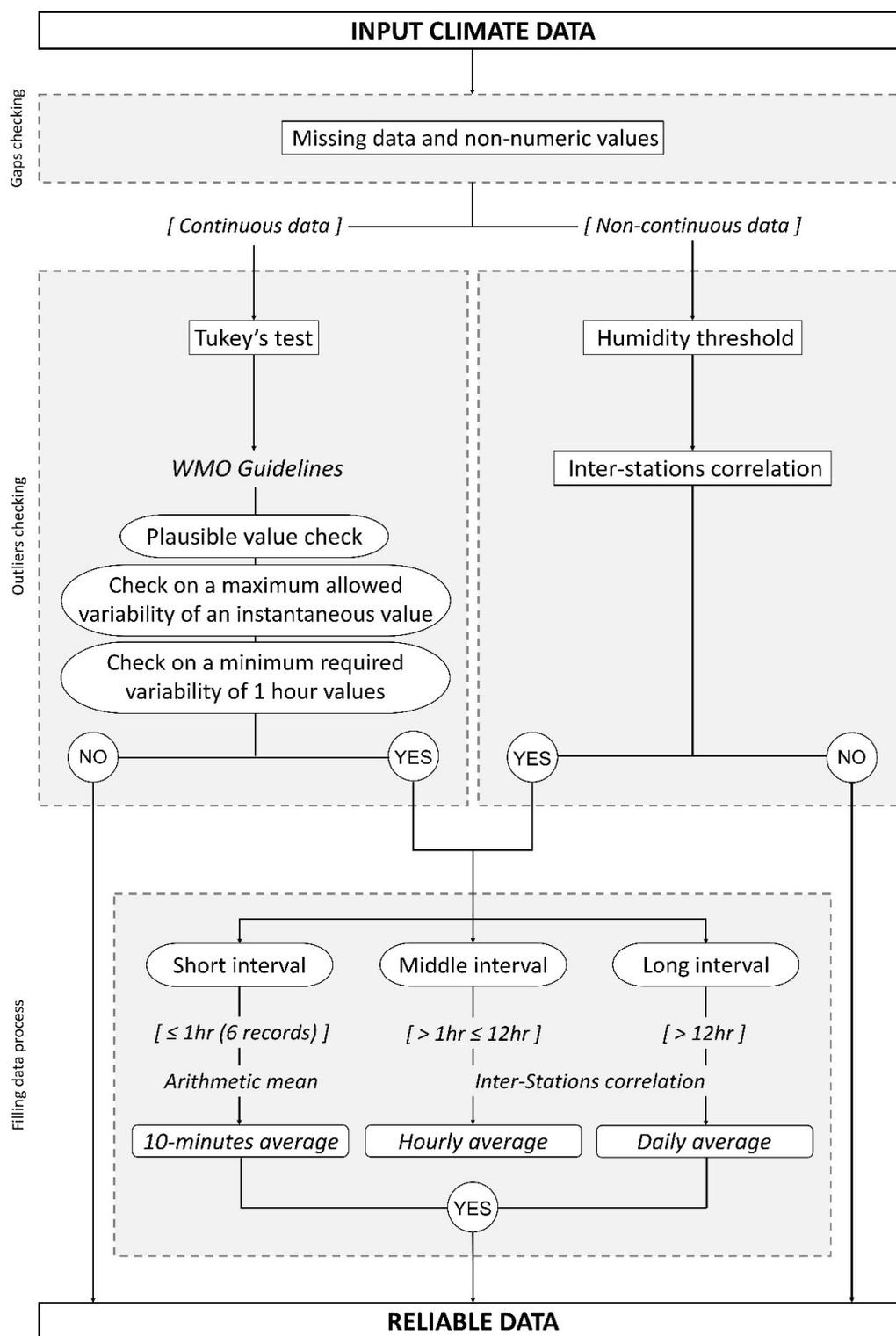


Figure 16: Data quality flow to improve regional-scale fog climate network climate data series. Source: author's elaboration

### 5.1.2. Gaps checking: missing data and non-numeric values

A fundamental prerequisite for the precise execution of a regional climatological analysis resides in the accessibility of comprehensive climate data sets devoid of any gap. However, in practice, missing data is a common issue, originating from a myriad of factors, inclusive of, yet not confined to, instrument malfunction, inaccuracies in data input, and alterations in observational methodologies or instrumentation.

The process of ensuring data quality starts by identifying gaps and non-numeric values in the climatic data series (see Programming flow 1), by climatological station and by variable, which is recorded for subsequent correction or filling, as shown in Figure 17.

---

```
# Codes package import
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
from matplotlib.dates import YearLocator, ConciseDateFormatter, date2num
from datetime import timedelta

# Database selection
oya1 = pd.read_excel('518_ALL.xlsx')

# Formatting DATE and TIME in DataFrame
oya1.loc[oya1['TIME'] == 0, 'TIME'] = '00:00'
oya1['DATE_FULL'] = pd.to_datetime(oya1['DATE']+oya1['TIME'], format='%d.%m.%Y%H:%M')
oya1['DATE'] = pd.to_datetime(oya1['DATE'], format='%d.%m.%Y')

# Define the relevant variables and their names
variable_names = {
    'SFW': 'Standard fog water',
    'FW360': 'Fog water 360°',
    'DW': 'Dew water',
    'AP': 'Air pressure',
    'RH': 'Relative humidity',
    'AT': 'Air temperature',
    'LW': 'Leaf wetness',
    'WS2': 'Wind speed 2m',
    'WD2': 'Wind direction 2m',
    'WS10': 'Wind speed 10m',
    'WD10': 'Wind direction 10m',
    'GR': 'Global radiation',
    'DP': 'Dew',
    'SoT': 'Soil temperature'}
'SoM': 'Soil moisture'}
'ST': 'Surface temperature'}

# Create an Excel writer object
```

## 5. Fog variability characterization in the Chilean Atacama Desert

---

```
output_file_excel = 'Filter_Missing data_OYA_518.xlsx'
writer_excel = pd.ExcelWriter(output_file_excel, engine='xlsxwriter')

# Iterate through each variable specified in variable_names and check for missing data
num_plots = 0
variables_with_missing_data = []
for variable, variable_name in variable_names.items():
    if variable in oya1.columns:
        missing_non_numeric_data = oya1[oya1[variable].notna() & ~pd.to_numeric(oya1[variable],
errors='coerce').notna()]
        gaps_data = oya1[oya1[variable].isnull()]
        if not missing_non_numeric_data.empty or not gaps_data.empty:
            num_plots += 1
            variables_with_missing_data.append(variable)
num_rows = (num_plots // 2) + (num_plots % 2)
num_cols = min(num_plots, 2)
fig, axes = plt.subplots(num_rows, num_cols, figsize=(12, 6))
axes = axes.flatten()
counter = 0

# Iterate through variables with missing data to create plots
for variable in variables_with_missing_data:
    variable_name = variable_names[variable]
    for variable, variable_name in variable_names.items():
        if variable in oya1.columns:
            missing_non_numeric_data = oya1[
                oya1[variable].notna() & ~pd.to_numeric(oya1[variable], errors='coerce').notna()]
            gaps_data = oya1[oya1[variable].isnull()]
            if not missing_non_numeric_data.empty or not gaps_data.empty:
                if not missing_non_numeric_data.empty:
                    axes[counter].scatter(date2num(missing_non_numeric_data['DATE']),
np.ones_like(missing_non_numeric_data.index), c='red', s=1, label='Non-Numeric Data')
                if not gaps_data.empty:
                    axes[counter].scatter(date2num(gaps_data['DATE']), np.zeros_like(gaps_data.index), c='red', s=1,
label='Gaps')

                # Set title with percentages in parentheses and axis
                total_data_points = len(oya1)
                percent_non_numeric = len(missing_non_numeric_data) / total_data_points * 100 if
missing_non_numeric_data.size > 0 else 'ND'
                percent_gaps = len(gaps_data) / total_data_points * 100 if gaps_data.size > 0 else 'ND'
                title_non_num = f'{percent_non_numeric:.2f}%' if isinstance(percent_non_numeric, float) else 'ND'
                title_gaps = f'{percent_gaps:.2f}%' if isinstance(percent_gaps, float) else 'ND'
                acronym = variable
                title = f'{acronym} ({title_non_num}/{title_gaps})'
                axes[counter].set_title(title, fontsize=12.5)
                axes[counter].set_yticks([0, 1])
                axes[counter].set_yticklabels(['Gaps', 'Non-Numeric Data'], fontsize=12.5)

                # Remove the right and top spines
                axes[counter].spines['right'].set_visible(False)
                axes[counter].spines['top'].set_visible(False)

                # Set a common X-axis based on the original 'DATE' column
                axes[counter].set_xlim(oya1['DATE'].min(), oya1['DATE'].max())
                axes[counter].set_ylim(-0.5, 2)
                years = YearLocator()
                axes[counter].xaxis.set_major_locator(years)
                axes[counter].xaxis.set_major_formatter(ConciseDateFormatter(years))
```

## 5. Fog variability characterization in the Chilean Atacama Desert

---

```
# Store data in a dataframe for the variable with missing data and save relevant columns in the Excel file
oya1['ID'] = oya1.index
df_missing_data = pd.concat([
    oya1.loc[oya1['ID'].isin(missing_non_numeric_data.index), ['ID', 'DATE', 'TIME', variable]],
    oya1.loc[oya1['ID'].isin(gaps_data.index), ['ID', 'DATE', 'TIME', variable]])
df_missing_data['Type'] = np.where(df_missing_data['ID'].isin(missing_non_numeric_data.index), 'Non-
Numeric Data', 'Gaps')
df_missing_data['DATE'] = df_missing_data['DATE'].dt.date
df_missing_data['ID'] = df_missing_data['ID'] + 2
df_missing_data[['ID', 'DATE', 'TIME', variable, 'Type']].to_excel(writer_excel, sheet_name=f'{variable}',
index=False)
    counter += 1
writer_excel.close()

# Save all plots in a single JPG file
plt.tight_layout()
plt.savefig('Filter_Missing data_OYA_518.jpg', bbox_inches='tight', dpi=300)

plt.show()

#####

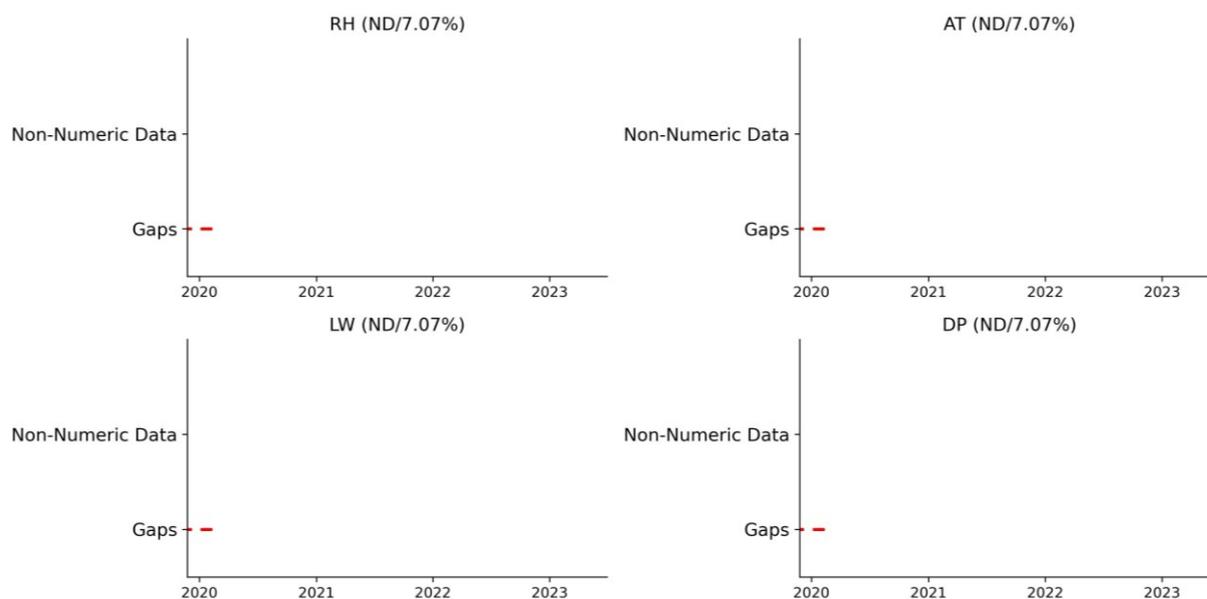
# Converting 'DATE_FULL' column to 'timestamp' column and setting up variables for time intervals
oya1['timestamp'] = (pd.to_datetime(oya1['DATE_FULL'], format='%d:%m:%Y %H:%M') +
pd.Timedelta(hours=3)).astype('int64') // 10**9
dt = 60*10
missed_date = False

# Iterating through the 'timestamp' column to find missing date intervals
for i in range(1,len(oya1['timestamp'])):
    if(oya1['timestamp'][i]-oya1['timestamp'][i-1]!=dt and not missed_date):
        missed_date = True
        i_missed = i - 1
        continue
    if(oya1['timestamp'][i]-oya1['timestamp'][i-1]!=dt and missed_date):continue
    if(oya1['timestamp'][i]-oya1['timestamp'][i-1]==dt and missed_date):
        n_data = int((oya1['timestamp'][i]-oya1['timestamp'][i_missed])/dt) - 2
        if(n_data<5):
            message = "%i minutes"%(n_data*10)
        elif(n_data<144):
            message = "%i hours"%(n_data*10/60)
            if (n_data*10%60!=0):
                message += " %i minutes"%(n_data*10%60)
        else:
            message = "%i days"%(n_data*10/1440)
            if (n_data*10%1440!=0):
                if(n_data*10%1440>=60):message += " %i hours"%((n_data*10%1440)/60)
                if ((n_data*10%1440)%60!=0):
                    message += " %i minutes"%((n_data*10%1440)%60)
        print("Date interval with missing data: n° records %i corresponding to %s between %s and
%s"%(n_data,message,oya1['DATE_FULL'][i_missed],oya1['DATE_FULL'][i-1]))
        missed_date = False

---
```

Programming flow 1: Gaps checking: missing data. Source: author's elaboration

## 5. Fog variability characterization in the Chilean Atacama Desert



\*ND: No data

Figure 17: Standard plots of gaps and non-numeric data by variable in a data series (OYA\_518).  
Source: author's elaboration

Non-numeric values are not recorded in climate data series, both in the dates of data records and in the availability of data by climatic variable. Meanwhile missing data are recorded, these manifest as temporal gaps, disrupting the continuity of the recorded climate data in the continuity of the dates as well as in the availability of data by climatological variable as exemplified the Figure 18.

At OYA\_518, the record for each climatic variable exhibited consistent gaps, with a uniform 7.07% across RH, AT, LW, and DP variables, indicating potential systemic issues during data collection. In contrast, the nearest station, OYA\_780, demonstrated exceptional data integrity, with only 0.009% of gaps across multiple variables, underscoring its highly reliable data acquisition. Since its installation in 2022, OYA\_862 has proven exceptional reliability, boasting a perfect record with zero gaps and non-numeric values across all variables. Similarly, OYA\_1069 has shown high reliability, with minimal gaps of 0.006% in most variables, ensuring consistent and dependable data.

The same reliability can be seen in OYA\_1128, OYA\_1193 and OYA\_1211. These stations exhibited gaps in SFW, accounting for 0.13%, 0.25% and 0.01% respectively, while maintaining minimal gaps in other variables. Meanwhile, OYA\_1354 demonstrated moderate gaps of 0.034% in SFW, suggesting just sporadic data collection issues.

The identification of complete data series (all variables) missing in a range of dates (same day, month, year, time) is determined by the satellite transmission in two climate data packages, the first Chile\_1 which includes OYA\_1211, OYA\_1128, OYA\_1354 and OYA\_1193, and Chile\_2 which includes OYA\_518, OYA\_780, OYA\_1069 and OYA\_862. However, the climatic stations present different date gaps due to the different date of installation and beginning of the record of each climatic station.

Considering the data package and stations nearest coast, OYA\_518 portrays sporadic instances of missing data, distributed across various temporal intervals. Notably, the largest temporal gap spans 424 days and 17 hours, significantly impacting the dataset's continuity between February 7<sup>th</sup>, 2020, and April 6<sup>th</sup>, 2021, and then until May 2<sup>nd</sup>, 2022. Moreover, intermittent short-duration anomalies further disrupt the dataset's consistency, standing out as the climatological station with the greatest amount of missing data in the database, recording 34.60% data availability when considering the start of data transmission and the final record included in this research. OYA\_780 the anomalies identified in this station span diverse temporal extents, with instances recurring across several intervals. Notable durations include anomalies persisting for 502 days and 9 hours, significantly affecting data continuity between November 21<sup>st</sup>, 2018, and April 6<sup>th</sup>, 2021. The OYA\_862 exhibits just 8 missing records (80-minutes records). For OYA\_1069, no missing data were identified within the dataset for this station.

Chile\_1 data package has minor anomalies in its data set. OYA\_1128 records missing data predominantly manifest as shorter-duration gaps, intermittently scattered across the dataset. OYA\_1193 highlights for being the climatic station with the greatest continuity in the data series, recording 96.80% availability of the database. At OYA\_1211 the anomalies identified within this station span minor temporal extents, showcasing interruptions in data continuity across multiple intervals, mainly shorts. OYA\_1354 akin to previous instances, present recurring interruptions at various intervals, thereby impacting the overall temporal continuity of the recorded climate data.

## 5. Fog variability characterization in the Chilean Atacama Desert

---

Date interval with missing data: n° records 5165 corresponding to 35 days 20 hours 50 minutes between 2019-12-04 19:00:00 and 2020-01-09 16:00:00  
Date interval with missing data: n° records 3 corresponding to 30 minutes between 2020-01-09 16:10:00 and 2020-01-09 16:50:00  
Date interval with missing data: n° records 2 corresponding to 20 minutes between 2020-02-05 13:20:00 and 2020-02-05 13:50:00  
Date interval with missing data: n° records 61162 corresponding to 424 days 17 hours 40 minutes between 2020-02-07 00:10:00 and 2021-04-06 18:00:00  
Date interval with missing data: n° records 1 corresponding to 10 minutes between 2021-04-07 15:30:00 and 2021-04-07 15:50:00  
Date interval with missing data: n° records 23 corresponding to 3 hours 50 minutes between 2021-04-13 20:00:00 and 2021-04-14 00:00:00  
Date interval with missing data: n° records 2 corresponding to 20 minutes between 2021-05-06 13:20:00 and 2021-05-06 13:50:00  
Date interval with missing data: n° records 3 corresponding to 30 minutes between 2021-05-14 08:20:00 and 2021-05-14 09:00:00  
Date interval with missing data: n° records 114 corresponding to 19 hours between 2021-05-14 10:30:00 and 2021-05-15 05:40:00  
Date interval with missing data: n° records 7 corresponding to 1 hours 10 minutes between 2021-05-20 02:50:00 and 2021-05-20 04:10:00  
Date interval with missing data: n° records 1 corresponding to 10 minutes between 2021-05-20 16:50:00 and 2021-05-20 17:10:00  
Date interval with missing data: n° records 1 corresponding to 10 minutes between 2021-06-02 14:10:00 and 2021-06-02 14:30:00  
Date interval with missing data: n° records 4 corresponding to 40 minutes between 2021-06-02 14:40:00 and 2021-06-02 15:30:00  
Date interval with missing data: n° records 3 corresponding to 30 minutes between 2021-06-03 18:10:00 and 2021-06-03 18:50:00  
Date interval with missing data: n° records 1 corresponding to 10 minutes between 2021-06-11 14:10:00 and 2021-06-11 14:30:00  
Date interval with missing data: n° records 46819 corresponding to 325 days 3 hours 10 minutes between 2021-06-11 15:30:00 and 2022-05-02 18:50:00  
Date interval with missing data: n° records 2 corresponding to 20 minutes between 2022-05-03 15:40:00 and 2022-05-03 16:10:00  
Date interval with missing data: n° records 1 corresponding to 10 minutes between 2022-05-04 15:10:00 and 2022-05-04 15:30:00  
Date interval with missing data: n° records 3 corresponding to 30 minutes between 2023-01-21 14:20:00 and 2023-01-21 15:00:00  
Date interval with missing data: n° records 3 corresponding to 30 minutes between 2023-01-21 15:40:00 and 2023-01-21 16:20:00  
Date interval with missing data: n° records 2 corresponding to 20 minutes between 2023-01-23 15:30:00 and 2023-01-23 16:00:00  
Date interval with missing data: n° records 2 corresponding to 20 minutes between 2023-01-23 16:20:00 and 2023-01-23 16:50:00  
Date interval with missing data: n° records 3 corresponding to 30 minutes between 2023-01-23 17:00:00 and 2023-01-23 17:40:00  
Date interval with missing data: n° records 3 corresponding to 30 minutes between 2023-01-24 16:30:00 and 2023-01-24 17:10:00  
Date interval with missing data: n° records 2 corresponding to 20 minutes between 2023-01-27 14:40:00 and 2023-01-27 15:10:00

Figure 18: Missing data across complete time intervals in a data series (OYA\_518). Source: author's elaboration

### 5.1.3. Outliers checking

The differentiation of raw climatic data into continuous and non-continuous datasets is essential due to their varied nature and measurements.

Employing the Tukey's test for identifying potential outliers in continuous climatic data provides significant advantages across a range of datasets. However, augmenting this approach with checks for plausible climatic values and acceptable climatological changes, as stipulated within ten-minute recording intervals and hourly variability, following scientifically endorsed guidelines by the WMO adapted to the fog climate research network, establishes a robust foundation for enhancing data quality. On the other hand, the treatment of continuous climatic data, such as the SFW collection, needs a distinct methodology for outliers' identification. Optimal results in detecting outliers are achieved when fog collection is recorded with RH below 90%, following with the correlation of water collection records within a homogeneous atmosphere-biosphere system.

### *Continuous data*

The first filter applied is the Tukey's test (David & Tukey, 1977), as seen in Programming flow 2. Applied to this data series, it aims to identify errors, such as when a data value significantly differs from the preceding or succeeding value within the same time series (Feng et al., 2004). The test employs three internal steps to detect outliers; the filter identifies the interquartile range (IQR) and subsequently calculates the lower and upper extremes, designating values beyond these boundaries as potential outliers as seen, for example, in the standard output plots shown in Figure 19.

Distance from the coast has a significant impact on climatic conditions due to the moderating effect of the ocean. Most of the analyzed variables record outliers in the Tukey's test, with a global average of the data series and variables at 5.45%, fluctuating between 1.26% in OYA\_1069 and 10.07% in OYA\_518 with the measurements of RH as those that record larger outlier records. Measurement records near the ocean experience greater stability in humidity conditions due to the influence of the maritime climate. In contrast, as distance from the coast increases toward the interior, climate variability increases due to the decrease in the influence of the ocean, which can generate a greater proportion of outliers in the measurements. Local factors such as the presence of vegetation or microtopography can increase local RH and therefore influence the identification of outliers. At Oyarbide Site a higher proportion of LW outliers could indicate the influence of local vegetation, which retains moisture and gives rise to atypical measurements.

In the context of RH measurements in the Atacama Desert, outliers may arise due to fog events arrives. Therefore, Tukey's test to measurements in the Atacama Desert, it is essential to consider that some values identified as outliers may genuinely reflect exceptional climatic conditions or natural events, rather than measurement errors. In these cases, it is necessary to contextualize and comprehend the environmental conditions to appropriately interpret the values identified as outliers and determine their relevance or correctness based on the research or monitoring objectives.

---

```
# Codes package import
```

```
import numpy as np
```

```
import matplotlib.pyplot as plt
```

```
import pandas as pd
```

```
# Database selection
```

```
oya1 = pd.read_excel('1128_ALL.xlsx')
```

```
# Define units of measurement for each variable
```

## 5. Fog variability characterization in the Chilean Atacama Desert

---

```
variable_units = {
    'AP': 'hPa',
    'WS10': 'm/s',
    'WD10': '°',
    'RH': '%',
    'AT': '°C',
    'ST10': '°C',
    'GR': 'W/m2',
    'LW': '%',
    'WS2': 'm/s',
    'WD2': '°',
    'ST2': '°C',
    'DP': 'No/Yes'}

# Create a dictionary to store outliers for each variable along with their IDs
outliers_dict = {}
plt.figure(figsize=(15, 30))
pos = 1
for col in oya1.columns:
    if col in variable_units:
        Q1 = oya1[col].quantile(0.25)
        Q3 = oya1[col].quantile(0.75)
        IQR = Q3 - Q1
        filter = (oya1[col] < Q1 - 1.5 * IQR) | (oya1[col] > Q3 + 1.5 * IQR)
        if any(filter):
            outliers = pd.DataFrame({'ID': oya1.index[filter]+ 2, col: oya1[col][filter]})

            # Add the 'DATE', 'TIME', and 'SFW' columns to outliers_data and reorder the columns
            outliers['DATE'] = oya1['DATE'][filter]
            outliers['TIME'] = oya1['TIME'][filter]
            outliers['SFW'] = oya1['SFW'][filter]
            outliers = outliers[['ID', 'DATE', 'TIME', 'SFW', col]]
            outliers = outliers.reset_index(drop=True)
            outliers_dict[col] = outliers
            plt.subplot(7, 3, pos)
            boxprops = dict(linewidth=1.5, color='black')
            medianprops = dict(linestyle='-', linewidth=1.5, color='black')
            flierprops = dict(marker='o', markersize=1, markeredgcolor='red', linestyle='none') # Outliers as points

            # Customize the y-axis label with units of measurement
            y_label = f'{variable_units[col]}'
            oya1.boxplot(column=col, grid=False, showfliers=True, boxprops=boxprops, medianprops=medianprops,
flierprops=flierprops)
            plt.ylabel(f'{variable_units[col]}', fontsize=12.5)

            # Set the title with the percentage of outliers and remove the x-axis label
            total_data = len(oya1[col])
            percentage_outliers = len(oya1[col][filter]) / total_data * 100
            plt.title(f'{col} ({percentage_outliers:.2f}%)', fontsize=12.5)

            # Remove top and right axis lines and x-axis label
            plt.gca().spines['top'].set_visible(False)
            plt.gca().spines['right'].set_visible(False)
            plt.xticks([])
            plt.xticks(fontsize=12.5)
            plt.yticks(fontsize=12.5)
            pos = pos + 1

# Save all plots in a single JPG file and Excel file with outliers per variable
plt.tight_layout()
```

## 5. Fog variability characterization in the Chilean Atacama Desert

```
plt.savefig('Filter_Tukeys_OYA_1128.jpg', bbox_inches='tight', dpi=300)
with pd.ExcelWriter('Filter_Tukeys_OYA_1128.xlsx') as writer:
    for col, outliers in outliers_dict.items():
        outliers.to_excel(writer, sheet_name=col, index=False)
```

```
plt.show()
```

```
--
```

Programming flow 2: Outliers checking by Tukey's test. Source: author's elaboration

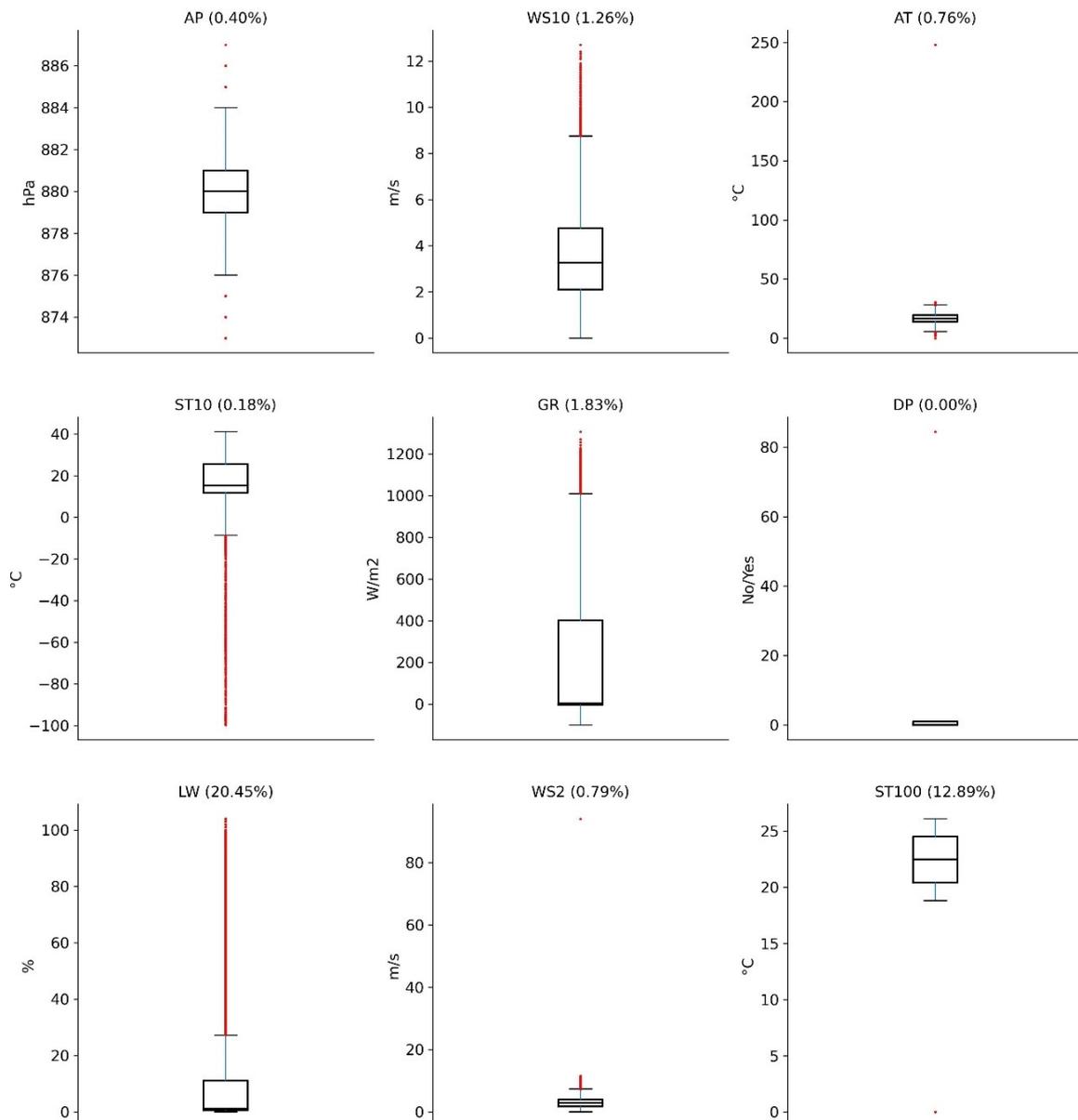


Figure 19: Standard plots of the Tukey's test check by variable in a data series (OYA\_1211). Source: author's elaboration

After Tukey's test the check "plausible value" assessment phase is to validate whether the data falls within the permissible range boundaries defined by measurement range of each climatological sensor (see Programming flow 3) as seen for example in Figure 20. Each sample is subject to a scrutiny to ascertain whether its value aligns with the measurement range specified by the climatological device as specified in Table 1. If a value fails this assessment, it is rejected, and data that successfully passes this stage continue in the data flow with the next stage processes.

At OYA\_518 and OYA\_780 present a notable outlier percentage of 2.99% and 4.28% in WS2 respectively. Moving to OYA\_862, WS2 exhibits a moderate outlier percentage of 1.62% and OYA\_1069 presents a higher 3.25% in WS2. In contrast, OYA\_1128 reports minimal outliers in WS2 (0.09%) and AT (0.008%). However, the most significant disparities arise at OYA\_1211, showcasing remarkable outlier percentages in GR (48.77%), alongside notable outliers in LW (0.10%) and WS2 (0.92%). OYA\_1354 displays minor outliers in WS2 (0.02%).

On average, 3.47% of non-plausible data are observed within the global database. The stations exhibiting the lowest proportion of non-plausible data are those of more recent installation at the "Oyarbide Site," specifically OYA\_1193 (0.001%) and OYA\_1128 (0.06%). Similarly, OYA\_1354 in "Oyarbide Hill" system records only 0.05% of implausible data in its global database. Conversely, the station OYA\_1211 displays the highest count (6) of variables exceeding plausible limits.

---

```
# Codes package import
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt

# Database selection
oya1 = pd.read_excel('1211_ALL.xlsx')

# Convert the 'DATE' column to a datetime data type
oya1['DATE'] = pd.to_datetime(oya1['DATE'], format='%d.%m.%Y')

# Units of measurement for each variable
variable_units = {
    'AP': 'hPa',
    'RH': '%',
    'AT': '°C',
    'LW': '%',
    'WS2': 'm/s',
    'WD2': '°',
    'WS10': 'm/s',
    'WD10': '°',
    'GR': 'W/m2',
    'DP': 'Yes/No',
    'ST10': '°C'}
```

## 5. Fog variability characterization in the Chilean Atacama Desert

---

```
cols = list(variable_units.keys())
lim = [[800, 1060], [0, 100], [-30, 70], [0, 100], [0.3, 50],[0, 360],[0.3, 50],[0, 360],[0, 2000], [0, 1], [-30, 50]]

# Iterate through each column and create a separate plot for each variable
outlier_data = []
num_plots = 0
for col in cols:
    if col in oya1.columns and oya1[col].dtype in [np.int64, np.float64]:
        outliers = oya1[(oya1[col] < lim[cols.index(col)][0]) | (oya1[col] > lim[cols.index(col)][1])]
        total_data_points = len(oya1[col])
        outlier_count = len(outliers)
        outlier_percentage = (outlier_count / total_data_points) * 100
        if not outliers.empty:
            num_plots += 1
num_rows = (num_plots // 3) + (num_plots % 3 > 0)
num_cols = min(num_plots, 3)
fig, axes = plt.subplots(num_rows, num_cols, figsize=(12, 4))
axes = axes.flatten()
counter = 0
cols_outliers = []
for col in cols:
    if col in oya1.columns and oya1[col].dtype in [np.int64, np.float64]:
        outliers = oya1[(oya1[col] < lim[cols.index(col)][0]) | (oya1[col] > lim[cols.index(col)][1])]
        if not outliers.empty:
            if counter < len(axes):
                axes[counter].plot(outliers['DATE'], outliers[col], '.r', markersize=2)

                # Set labels and title with the percentage and the Y-axis label with the unit of measurement
                axes[counter].set_title(f'{col} ({outlier_percentage:.2f}%)', fontsize=12.5)
                axes[counter].set_xlabel('Date', fontsize=12.5)
                axes[counter].set_ylabel(variable_units[col], fontsize=12.5)

                # Remove right and top borders
                axes[counter].spines['right'].set_visible(False)
                axes[counter].spines['top'].set_visible(False)
                axes[counter].tick_params(axis='x', rotation=45)

                # Collect outlier data
                counter += 1
                cols_outliers.append(col)
                for i, value in outliers.iterrows():
                    new_id = i + 2
                    outlier_data.append([i,
                                        value['DATE'],
                                        value['TIME'],
                                        variable_units[col],
                                        value[col]])

# Save all plots in a single JPG file
plt.tight_layout()
fig.savefig('Filter_Plausible values_OYA_1211.jpg', bbox_inches='tight', dpi=300)

# Create a DataFrame for the outlier data
outlier_df = pd.DataFrame(outlier_data, columns=['ID', 'DATE', 'TIME', 'Variable', 'Value'])
outlier_df['ID'] = outlier_df['ID'] + 2
outlier_df['DATE'] = pd.to_datetime(outlier_df['DATE']).dt.date

# Create an Excel file and iterate through the variables and save the data in separate sheets
fileoutxlsx = 'Filter_Plausible values_OYA_1211.xlsx'
writer = pd.ExcelWriter(fileoutxlsx, engine='xlsxwriter')
```

## 5. Fog variability characterization in the Chilean Atacama Desert

```
for col in cols_outliers:
    variable_data = outlier_df[outlier_df['Variable'] == variable_units[col]]
    if not variable_data.empty:
        variable_data = variable_data.drop(['Variable'], axis=1)
        variable_data.to_excel(writer, sheet_name=col, index=False)

# Save the Excel file
writer.save()

plt.show()

---
```

Programming flow 3: Outliers check by plausible value. Source: author's elaboration

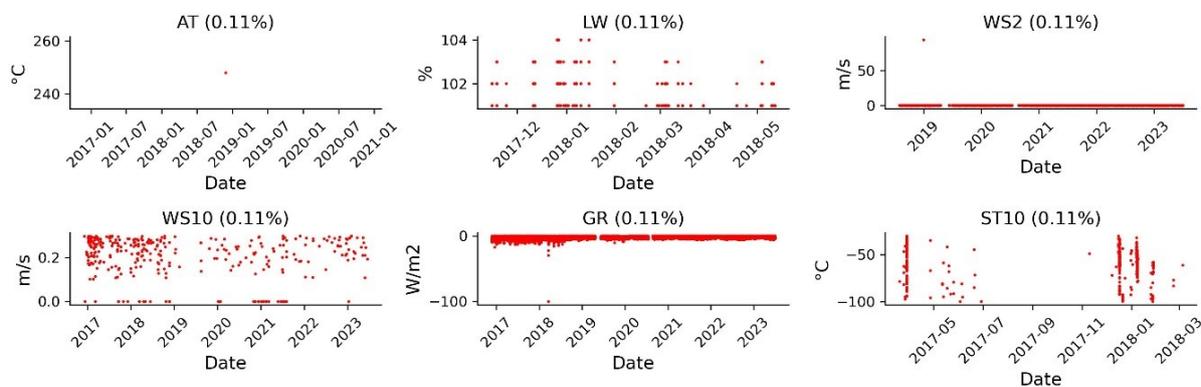


Figure 20: Standard plots of the plausible value check by variable in a data series (OYA\_1211). Source: author's elaboration

A calculation of standard deviation for the process of outlier detection in climate data is strongly recommended, as it serves as a valuable tool for identifying outliers, as well blocked sensors and addressing long-term sensor drift as was stipulated as a WMO Guidelines by Zahumensky (2004). The objective of the time consistency checks on maximum allowed variability of an instantaneous value is to examine the rate of change in instantaneous 10-minute data. Following each data measurement, the sample undergoes a comparative analysis with its predecessor. In instances where the disparity between these two samples surpasses the predetermined threshold of 10 standard deviations, the current sample is classified as, at least questionable. Consequently, it is duly recorded both in a comprehensive data table (see Programming flow 4) and in a graphical representation (see Figure 21).

OYA\_518 presents outlier occurrences across several variables. RH (2.72%), AT (1.85%), GR (1.26%), LW (3.86%), WS2 (11.12%), and WD2 (10.97%). Notably, the year 2022 emerged prominently,

contributing significantly to heightened outlier percentages across these variables. OYA\_780 displayed minimal outliers in AP (0.01%, 13 data points) but notable deviations in RH (6.07%), AT (3.72%), LW (3.33%), WS2 (11.57%), and WD2 (13.98%). Distinct high outlier patterns were observed notably in 2018 and 2019, significantly impacting these variables. OYA\_862 presented substantial outliers in RH (8.03%), AT (4.45%), LW (6.00%), WS2 (11.76%), and WD2 (13.28%). Notably, the year 2022 demonstrated heightened deviations across multiple variables, marking a distinct outlier trend. OYA\_1069 showcased outliers in RH (8.56%), AT (5.14%), GR (1.42%), WS2 (11.79%), and WD2 (15.54%), prominently influenced by the year 2023. Similarly, OYA\_1128 portrayed considerable outliers in RH (8.36%), AT (4.28%), LW (4.19%), WS2 (12.30%), and WD2 (13.55%), significantly influenced by the anomalies observed in the year 2021. OYA\_1193 exhibited remarkable outliers in RH (9.09%), AT (3.84%), LW (3.29%), WS2 (12.25%), and WD2 (12.39%). The year 2021 emerged as a significant outlier's year, showcasing distinct outlier trends across various variables. At climate station OYA\_1211, standout outlier occurrences were observed in WD10 (11.90%) during 2022 and RH (8.61%) in 2018. Lastly, OYA\_1354 revealed substantial outliers in RH (10.21%), AT (3.59%), LW (3.99%), WS2 (11.93%), and WD2 (11.86%). Notably, the year 2022 marked significant outlier trends across multiple variables.

---

```
# Codes package import
import numpy as np
import matplotlib.pyplot as plt
import pandas as pd
from datetime import datetime
from time import time

# Database selection
oya1 = pd.read_excel('780_ALL.xlsx')

# Define units of measurement for each variable
variable_units = {
    'AP': 'hPa',
    'WS10': 'm/s',
    'WD10': '°',
    'RH': '%',
    'AT': '°C',
    'ST10': '°C',
    'GR': 'W/m2',
    'LW': '%',
    'WS2': 'm/s',
    'WD2': '°',
    'ST100': '°C',
    'DP': 'No/Yes'}

# Convert the columns to floats for all variables and the 'DATE' column to datetime format
for var in variable_units.keys():
```

## 5. Fog variability characterization in the Chilean Atacama Desert

---

```
if var in oya1.columns:
    oya1[var] = pd.to_numeric(oya1[var], errors='coerce')
oya1['DATE'] = pd.to_datetime(oya1['DATE'], format='%d.%m.%Y').dt.date

# Initialize arrays and detect outliers for all variables
outliers = {var: {'data': np.zeros((6, len(oya1))), 'count': 0} for var in variable_units.keys()}
dates_outliers = {}
for var in variable_units.keys():
    list_dates_outliers = []
    if var in oya1.columns and var != "DP":
        for i in range(2, len(oya1) - 1):
            std_dev = np.std([oya1[var][i-1], oya1[var][i-2]])
            mean_value = np.mean([oya1[var][i-1], oya1[var][i-2]])
            factor_std = 10
            difference_abs = abs(oya1[var][i] - mean_value)
            if difference_abs >= factor_std * std_dev and std_dev > 0:
                outliers[var]['data'][0, outliers[var]['count']] = mean_value
                outliers[var]['data'][1, outliers[var]['count']] = oya1[var][i]
                outliers[var]['data'][2, outliers[var]['count']] = i + 2
                outliers[var]['data'][3, outliers[var]['count']] = factor_std * std_dev
                outliers[var]['data'][4, outliers[var]['count']] = difference_abs
                outliers[var]['count'] += 1
            list_dates_outliers.append(oya1['DATE'][i])
        dates_outliers[var] = list_dates_outliers
variables_with_outliers = [var for var in variable_units.keys() if var in outliers and outliers[var]['count'] > 0]

# Calculate the number of needed rows to organize the plots in columns by 3 to down
num_filas = len(variables_with_outliers) // 3
if len(variables_with_outliers) % 3 != 0:
    num_filas += 1

# Create a single figure for all plots
plt.figure(figsize=(15, 5 * num_filas))
for idx, var in enumerate(variables_with_outliers):
    plt.subplot(num_filas, 3, idx + 1)
    outliers_data = outliers[var]['data']
    plt.plot(dates_outliers[var], outliers_data[1, :len(dates_outliers[var])], 'ro', markersize=1)

# Set the Y-axis label with the unit of measurement and the title of the plot with the percentage of data plotted
plt.ylabel(f'{variable_units[var]}', fontsize=12.5)
plt.tick_params(axis='x', rotation=45)
total_points = len(oya1)
percentage_plotted = (outliers[var]['count'] / total_points) * 100
plt.title(f'{var} ({percentage_plotted:.2f}%)', fontsize=12.5)

# Remove spines from top and right sides
plt.gca().spines['top'].set_visible(False)
plt.gca().spines['right'].set_visible(False)

# Generation of data list with limit for outliers values
with pd.ExcelWriter('Filter_Maximum instantaneous change_OYA_780.xlsx', engine='xlsxwriter') as writer:
    for var in variable_units.keys():
        if var in outliers and outliers[var]['count'] > 0:
            df = pd.DataFrame({
                'ID': outliers[var]["data"][2, :],
                'Previous mean': outliers[var]["data"][0, :],
                'Value Exceeded': outliers[var]["data"][1, :],
                f'{factor_std} * std': outliers[var]["data"][3, :],
                'Difference abs': outliers[var]["data"][4, :]}
            )
            df = df[df['ID'] != 0]
```

## 5. Fog variability characterization in the Chilean Atacama Desert

```
# Save the data as a Excel file
df.to_excel(writer, sheet_name=var, index=False)

# Save all plots in a single JPG file
plt.tight_layout()
plt.savefig('Filter_Maximum instantaneous change_OYA_780.jpg', bbox_inches='tight', dpi=300)

plt.show()

---
```

Programming flow 4: Outliers check by maximum allowed variability of an instantaneous value.  
Source: author's elaboration

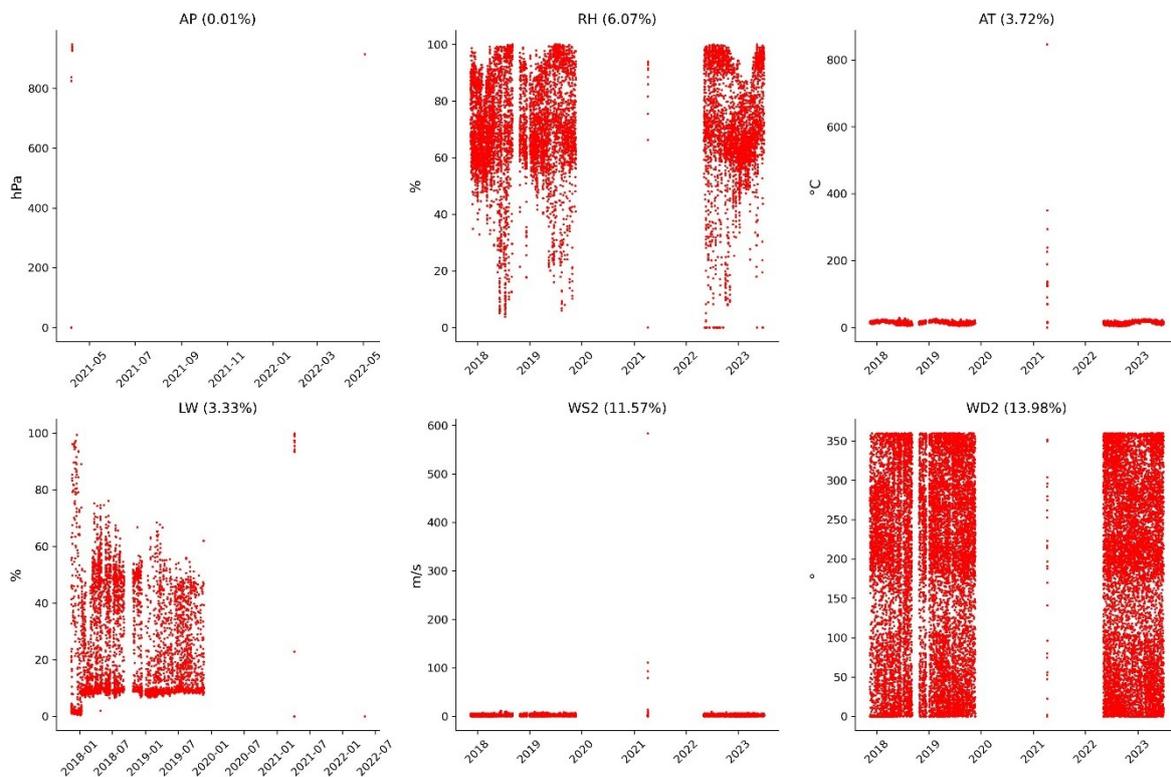


Figure 21: Standard plots of the check on a maximum allowed variability of an instantaneous value by variable in a data series (OYA\_780). Source: author's elaboration

The aim of this checking on a minimum required variability of instantaneous values is to identify outliers based on the degree of variability in instantaneous data over 60 minutes after the measurement of a given parameter as exemplified the Figure 22. Within a 6-records timeframe

(persistence test), data values failing to exhibit variation beyond a threshold of 0.2 standard deviations are deemed, at a minimum, questionable (see Programming flow 5).

The station OYA\_518 consistently exhibited significant outliers in RH (2.35%), AT (3.69%), GR (8.11%), LW (2.34%), WS2 (6.70%), and WD2 (6.80%) throughout the entire observation period. Notably, the year 2022 displayed the most pronounced variations across multiple variables, contributing substantially to the cumulative outlier percentage. Transitioning to OYA\_780 revealed persistent outlier occurrences in RH (10.05%) and AT (10.92%), primarily influenced by the anomalies recorded in the year 2022, leading to substantial deviations in these specific variables. OYA\_862 consistently presented diverse outliers in RH (4.50%), AT (4.89%), LW (2.69%), WS2 (8.39%), and WD2 (8.72%) over the entire duration under consideration, prominently driven by the anomalies observed in the year 2022. OYA\_1069 displayed outliers in RH (5.26%), AT (5.65%), GR (7.01%), WS2 (8.06%), and WD2 (7.61%) across the complete timeframe, notably prominent in the year 2023. In a similar vein, OYA\_1128 showcased notable outliers in RH (4.50%), AT (5.14%), LW (2.37%), WS2 (7.91%), and WD2 (8.23%), notably influenced by the anomalies in the year 2021. OYA\_1193 recorded moderate outlier occurrences in RH (3.85%), AT (3.95%), LW (1.67%), WS2 (5.60%), and WD2 (6.43%), particularly influenced by the anomalies observed in the year 2022. OYA\_1211, displayed varied outlier occurrences ranging from 0.02% to 7.52%, mostly recorded in 2021 and 2017. Finally, OYA\_1354 consistently revealed presence of outliers in RH (5.72%), AT (5.51%), LW (2.95%), WS2 (7.20%), and WD2 (6.49%) throughout the entire observation duration.

---

```
# Codes package import
import numpy as np
import matplotlib.pyplot as plt
import pandas as pd
from datetime import datetime
from time import time

# Database selection
oya1 = pd.read_excel('780_ALL.xlsx')

# Define units of measurement for each variable
variable_units = {
    'AP': 'hPa',
    'WS10': 'm/s',
    'WD10': '°',
    'RH': '%',
    'AT': '°C',
    'ST10': '°C',
    'GR': 'W/m2',
    'LW': '%',
    'WS2': 'm/s',
```

## 5. Fog variability characterization in the Chilean Atacama Desert

---

```
'WD2': '°',
'ST100': '°C',
'DP': 'No/Yes'}

# Convert the columns to floats for all variables and the 'DATE' column to datetime format
for var in variable_units.keys():
    if var in oya1.columns:
        oya1[var] = pd.to_numeric(oya1[var], errors='coerce')
    oya1['DATE'] = pd.to_datetime(oya1['DATE'], format='%d.%m.%Y').dt.date

# Initialize arrays and detect outliers for all variables
outliers = {var: {'data': np.zeros((5, len(oya1))), 'count': 0} for var in variable_units.keys()}
dates_outliers = {}
for var in variable_units.keys():
    list_dates_outliers = []
    if var in oya1.columns and var != "DP":
        for i in range(6, len(oya1)):
            std_dev = np.std(oya1[var][i-6:i-1])
            mean_value = np.mean(oya1[var][i-6:i-1])
            factor_std = 0.20
            difference_abs = abs(oya1[var][i] - mean_value)
            if difference_abs <= factor_std * std_dev and std_dev != 0:
                outliers[var]['data'][0, outliers[var]['count']] = mean_value
                outliers[var]['data'][1, outliers[var]['count']] = oya1[var][i]
                outliers[var]['data'][2, outliers[var]['count']] = i + 2
                outliers[var]['data'][3, outliers[var]['count']] = factor_std * std_dev
                outliers[var]['data'][4, outliers[var]['count']] = difference_abs
                outliers[var]['count'] += 1
            list_dates_outliers.append(oya1['DATE'][i])
        dates_outliers[var] = list_dates_outliers
variables_with_outliers = [var for var in variable_units.keys() if var in outliers and outliers[var]['count'] > 0]

# Calculate the number of needed rows to organize the plots in columns by 3 to down
num_filas = len(variables_with_outliers) // 3
if len(variables_with_outliers) % 3 != 0:
    num_filas += 1

# Create a single figure for all plots
plt.figure(figsize=(15, 5 * num_filas))
for idx, var in enumerate(variables_with_outliers):
    plt.subplot(num_filas, 3, idx + 1)
    outliers_data = outliers[var]['data']
    plt.plot(dates_outliers[var], outliers_data[1, :len(dates_outliers[var])], 'ro', markersize=1)

# Set the Y-axis label with the unit of measurement and the title of the plot with the percentage of data plotted
plt.ylabel(f'{variable_units[var]}', fontsize=12.5)
plt.tick_params(axis='x', rotation=45)
total_points = len(oya1)
percentage_plotted = (outliers[var]['count'] / total_points) * 100
plt.title(f'{var} ({percentage_plotted:.2f}%)', fontsize=12.5)

# Remove spines from top and right sides
plt.gca().spines['top'].set_visible(False)
plt.gca().spines['right'].set_visible(False)

# Generation of data list with limit for outliers
with pd.ExcelWriter('Filter_Minimum change_OYA_780.xlsx', engine='xlsxwriter') as writer:
    for var in variable_units.keys():
        if var in outliers and outliers[var]['count'] > 0:
            df = pd.DataFrame({
```

## 5. Fog variability characterization in the Chilean Atacama Desert

```

'D': outliers[var]["data"][2, :],
'Previous mean': outliers[var]["data"][0, :],
'Value Exceeded': outliers[var]["data"][1, :],
f'{factor_std} * std': outliers[var]["data"][3, :],
'Difference abs': outliers[var]["data"][4, :])
df = df[df['ID'] != 0]

# Save the data as a Excel file
df.to_excel(writer, sheet_name=var, index=False)

# Save all plots in a single JPG file
plt.tight_layout()
plt.savefig('Filter_Minimum change_OYA_780.jpg', bbox_inches='tight', dpi=300)

plt.show()

---
```

Programming flow 5: Outliers check by minimum required variability of 1 hr values. Source: author's elaboration

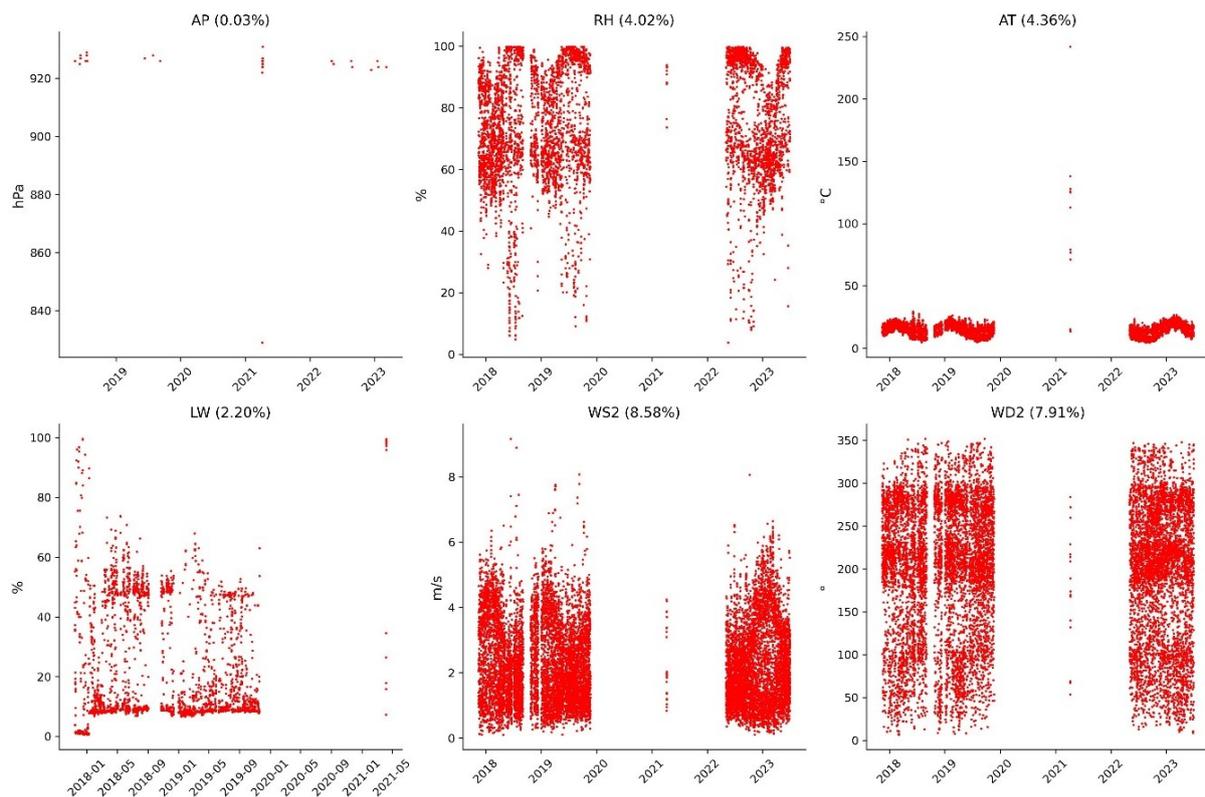


Figure 22: Standard plots of the check on a minimum required variability of 1 hr values by variable in a data series (OYA\_780). Source: author's elaboration

### *Non-continuous data*

The approach employed to identify outliers in non-continuous data involves a combination of two methods was implemented, relying on a minimum threshold of RH when SFW is collected, and subsequently with an inter-station correlation within the similar hydrological system.

Within the dataset of fog climate network, 0.82% of data records exhibits the presence of SFW collected under conditions of RH equal to or below the threshold 90% (Liu et al., 2018) as exemplified the Figure 23 (see Programming flow 6). Among the climate stations, 62.5% of climate stations records less than 1% records of SFW collected data below the stipulated RH threshold, OYA\_518 and OYA\_1211 records the lowest incidence with a 0.13% and 0.19%, respectively. Conversely, OYA\_862 stands out as the station recording by far the largest numbers of SFW collected records falling beneath the prescribed RH threshold, recording at 2.28%.

---

```
# Codes package import
import numpy as np
import matplotlib.pyplot as plt
import pandas as pd

# Database selection
oya1 = pd.read_excel('1193_ALL.xlsx')

# Filtering the table
threshold = 90
oya1_filter_incorrect = oya1.loc[(oya1['RH'] <= threshold) & (oya1['SFW'] > 0)]

# Create a new column 'DATETIME' by merging 'DATE' and 'TIME'
oya1_filter_incorrect['DATETIME'] = pd.to_datetime(oya1_filter_incorrect['DATE'] + ' ' +
oya1_filter_incorrect['TIME'], format='%d.%m.%Y %H:%M')

# Calculate the percentage of data points below the threshold within the original data
total_data_points_original = len(oya1)
total_data_points_filtered = len(oya1_filter_incorrect)
percentage_below_threshold_filtered = (total_data_points_filtered / total_data_points_original) * 100

# Create the plot
fig1 = plt.figure(figsize=(15, 13))
ax1 = fig1.add_subplot(111)
oya1_filter_incorrect.plot('DATETIME', 'RH', use_index=True, kind='scatter', marker='.', color='red', ax=ax1, s=15)
ax1.set_xlabel('Date', fontsize=12.5)

# Set y-axis label with the unit of measurement
ax1.set_ylabel('%', fontsize=12.5)

# Create a dataframe from the filtered data and add a new column "ID"
df = oya1_filter_incorrect[['DATE', 'TIME', 'RH', 'SFW']]
```

## 5. Fog variability characterization in the Chilean Atacama Desert

---

```
df['ID'] = df.index + 2

# Save the dataframe to an excel file
df.to_excel('Filter_Humidity threshold_OYA_1193.xlsx', sheet_name='tab1', index=False, columns=['ID', 'DATE',
'TIME', 'RH', 'SFW'])

# Remove the top and right lines from the plot
ax1.spines['top'].set_visible(False)
ax1.spines['right'].set_visible(False)

# Set the title with the percentage in parentheses and increased font size
ax1.set_title(f'Relative humidity ({percentage_below_threshold_filtered:.2f}%)', fontsize=12.5)

# Save all plots in a single JPG file
plt.tight_layout()
plt.savefig('Filter_Humidity threshold_OYA_1193.jpg', bbox_inches='tight', dpi=300)

plt.show()

---
```

Programming flow 6: Outliers checking by RH threshold. Source: author's elaboration

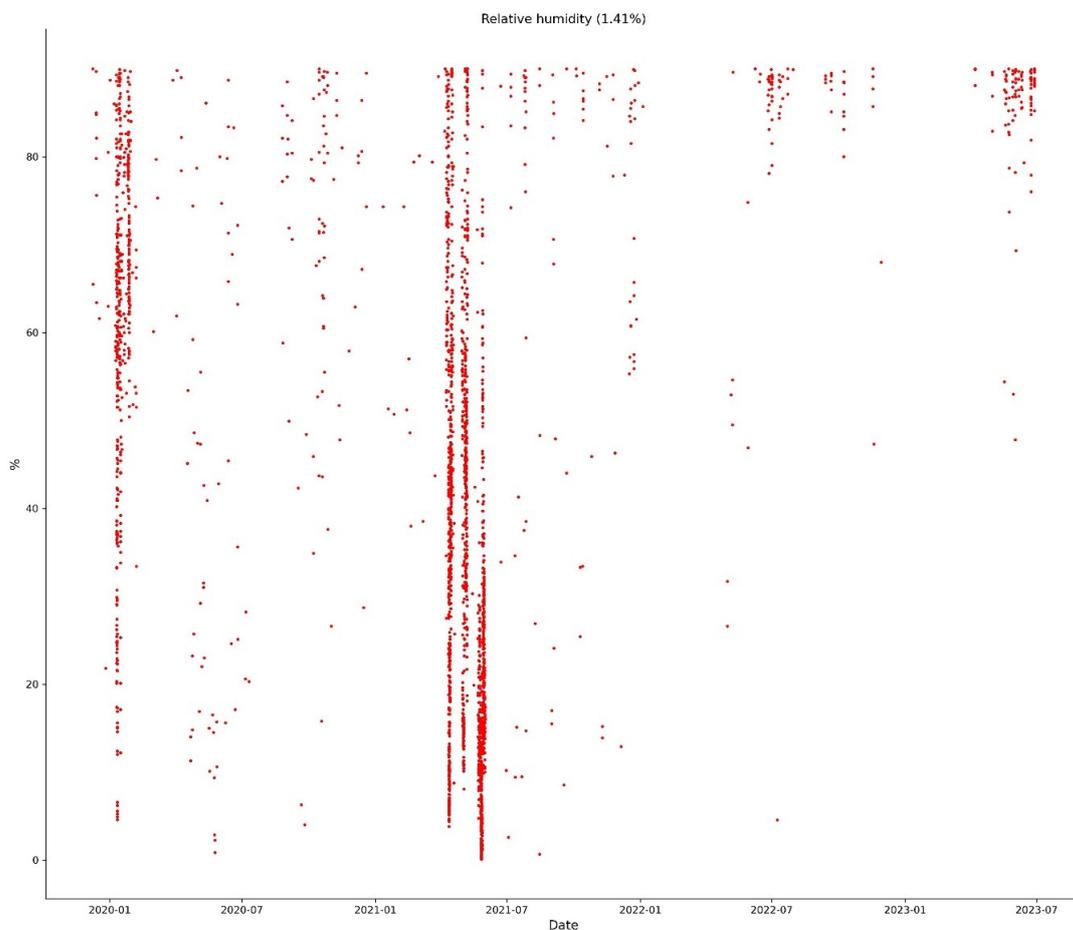


Figure 23: Standard plot of the check on the RH threshold in a data series (OYA\_1193). Source: author's elaboration

## 5. Fog variability characterization in the Chilean Atacama Desert

In the dataset under analysis, stations were defined for each homogeneous fog climatic system (see Figure 24), based on topographic and altitudinal criteria. The study site "Oyarbide" constitutes a homogeneous fog climatic group, as indicated by the  $r$  between the site's stations, where the collection of SFW shows values greater than 0.7 among them. Taking these considerations into account, three groups are obtained: group 1 is composed of stations OYA\_518, OYA\_780, OYA\_862, and OYA\_1069, all located below 1,069 masl and 10 km from the coast; group 2 (Oyarbide Site) includes OYA\_1128, OYA\_1193, and OYA\_1211; and group 3 consists of station OYA\_1354, which is the only one in the network that seasonally overpassed the thermal inversion layer. Due to the impossibility of making an interrelation in system 3, the process is not applied as exemplified the Programming flow 7 and Figure 25.

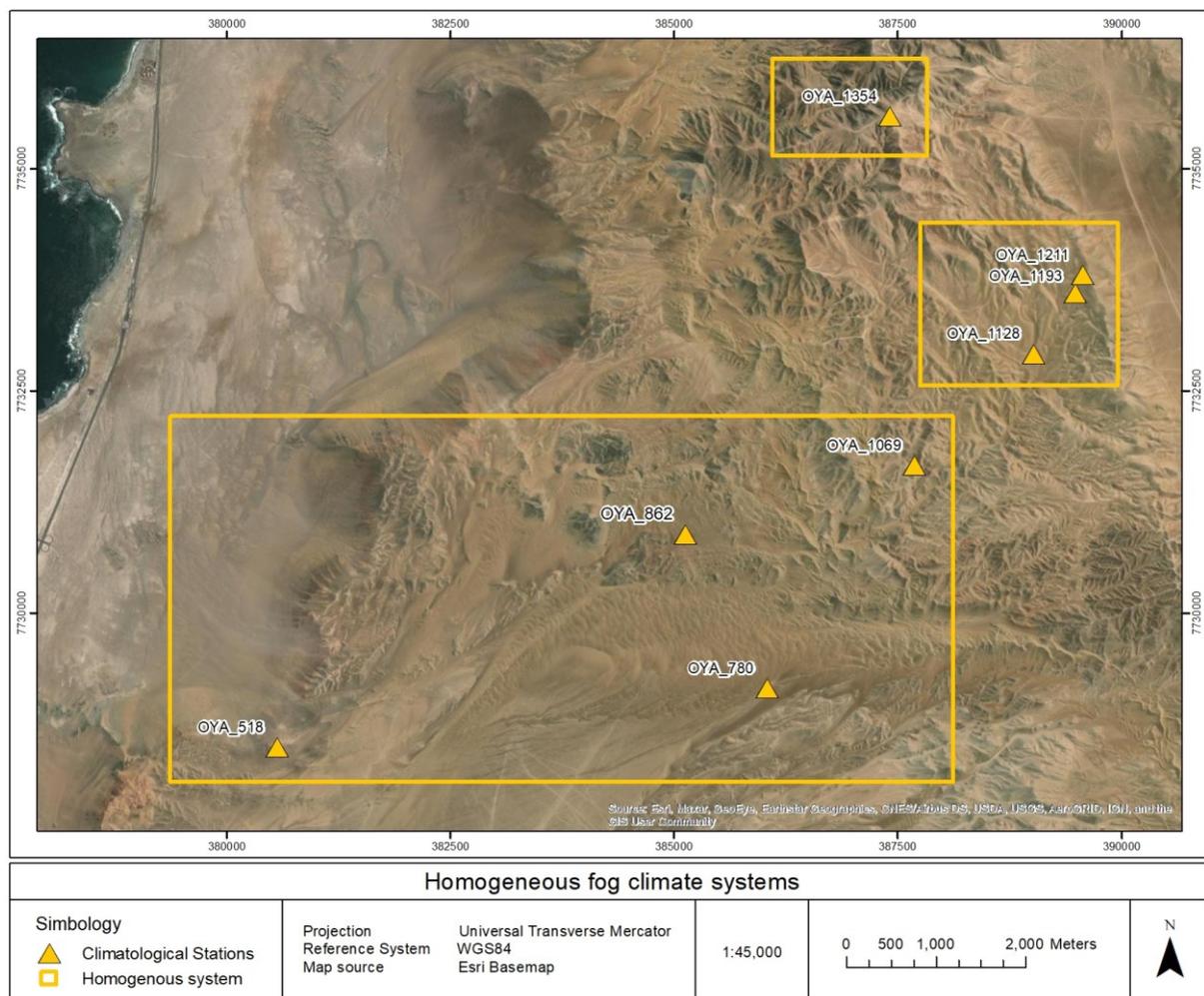


Figure 24: Homogeneous fog climate measurement systems at regional scale. Source: author's elaboration

At Oyarbide Site, the higher percentages of non-correlation observed in stations OYA\_1128 (3.91%), OYA\_1193 (4.52%), and OYA\_1211 (4.00%) may be attributed to local micro-topographic and climatic influences affecting the SFW collection process. Conversely, the group located furthest to the southwest exhibits a lower average interstation non-correlation (0.97%). Nevertheless, the absence of correlation in individual stations OYA\_518 (0.60%), OYA\_780 (0.23%), OYA\_862 (2.57%), and OYA\_1069 (0.49%), suggests a more homogeneous SFW collection environment with reduced influences that could disrupt the process at each location under consideration.

---

```
# Codes package import
import numpy as np
import matplotlib.pyplot as plt
import pandas as pd
from datetime import datetime
from matplotlib import pyplot as plt

# Database selection
oya1 = pd.read_excel('OYA_1128 09-2022.xlsx')
oya2 = pd.read_excel('OYA_1193 09-2022.xlsx')
oya3 = pd.read_excel('OYA_1211 09-2022.xlsx')

# Data visualization
n_date = 0
data = []
for i in range(len(oya1)):
    if(oya1.DATE[i]==oya2.DATE[i] and oya1.DATE[i]==oya3.DATE[i]):
        if(oya1.Collected[i]==oya2.Collected[i] and oya1.Collected[i]==oya3.Collected[i]):
            continue
        else:
            n_date += 1
            if(list([oya1.Collected[i],oya2.Collected[i],oya3.Collected[i]])==[1,0,0]):
                data.append([oya1.index[i]+1, oya1['DATE'][i], oya1['TIME'][i],"OYA_1128"])
                continue
            if(list([oya1.Collected[i],oya2.Collected[i],oya3.Collected[i]])==[0,1,0]):
                data.append([oya1.index[i]+1, oya1['DATE'][i], oya1['TIME'][i],"OYA_1193"])
                continue
            if(list([oya1.Collected[i],oya2.Collected[i],oya3.Collected[i]])==[0,0,1]):
                data.append([oya1.index[i]+1, oya1['DATE'][i], oya1['TIME'][i],"OYA_1211"])
                continue
            if(list([oya1.Collected[i],oya2.Collected[i],oya3.Collected[i]])==[1,1,0]):
                data.append([oya1.index[i]+1, oya1['DATE'][i], oya1['TIME'][i],"OYA_1211"])
                continue
            if(list([oya1.Collected[i],oya2.Collected[i],oya3.Collected[i]])==[1,0,1]):
                data.append([oya1.index[i]+1, oya1['DATE'][i], oya1['TIME'][i],"OYA_1193"])
                continue
            if(list([oya1.Collected[i],oya2.Collected[i],oya3.Collected[i]])==[0,1,1]):
                data.append([oya1.index[i]+1, oya1['DATE'][i], oya1['TIME'][i],"OYA_1128"])
                continue
        else:
            break
df = pd.DataFrame (data, columns = ['ID',"DATE","TIME","Table'])
df.to_excel('Filter intertable_Oyarbide Site.xlsx', sheet_name='tab1',index=False)
```

## 5. Fog variability characterization in the Chilean Atacama Desert

---

```
i=0
cols= ['DATE', 'Collected']
oya1['YEAR'][0]
params = {'legend.fontsize': 'x-large',
          'figure.figsize': (15, 11),
          'axes.labelsize': 'x-large',
          'axes.titlesize': 'x-large',
          'xtick.labelsize': 'medium',
          'ytick.labelsize': 'medium'}
plt.rcParams.update(params)
fig, ax = plt.subplots(1, 1)

# Data plotting
oya1_filtered_indices = oya1.index[(oya1.Collected==1)&(oya2.Collected==0)&(oya3.Collected==0)]
oya2_filtered_indices = oya2.index[(oya1.Collected==0)&(oya2.Collected==1)&(oya3.Collected==0)]
oya3_filtered_indices = oya3.index[(oya1.Collected==0)&(oya2.Collected==0)&(oya3.Collected==1)]
plt.plot(oya1_filtered_indices, np.arange(len(oya1_filtered_indices))*0+1, ".r", markersize=5)
plt.plot(oya2_filtered_indices, np.arange(len(oya2_filtered_indices))*0+2, ".r", markersize=5)
plt.plot(oya3_filtered_indices, np.arange(len(oya3_filtered_indices))*0+3, ".r", markersize=5)
ax = plt.gca()

# X and Y axis labeling
dates = []
for i in range(0, len(oya1), 145):
    dates.append(oya1.DATE[i])
ax.set_xticks(np.arange(0, len(oya1), 145))
ax.set_xticklabels(dates, fontsize=12.5)

# Percentage calculation and define axis labels and its percentages
total_data_oya1 = len(oya1)
total_data_oya2 = len(oya2)
total_data_oya3 = len(oya3)
percentage_oya1 = (len(oya1_filtered_indices) / total_data_oya1) * 100
percentage_oya2 = (len(oya2_filtered_indices) / total_data_oya2) * 100
percentage_oya3 = (len(oya3_filtered_indices) / total_data_oya3) * 100
labels_y = [f'OYA_1128 ({percentage_oya1:.2f}%)', f'OYA_1193 ({percentage_oya2:.2f}%)', f'OYA_1211
({percentage_oya3:.2f}%)']
ax.set_yticks([1, 2, 3])
ax.set_yticklabels(labels_y, fontsize=12.5)
ax.set_ylim(0.5, 3.5)
plt.xticks(rotation=45)
ax.set_xlabel('Date', fontsize=12.5)

# Remove the top and right line from the plot
ax.spines['top'].set_visible(False)
ax.spines['right'].set_visible(False)

# Save all plots in a single JPG file
plt.tight_layout()
plt.savefig('Filter intertable_Oyarbide Site.jpg', bbox_inches='tight', dpi=300)

plt.show()

---
```

Programming flow 7: Outliers identification by inter-stations correlation. Source: author's elaboration

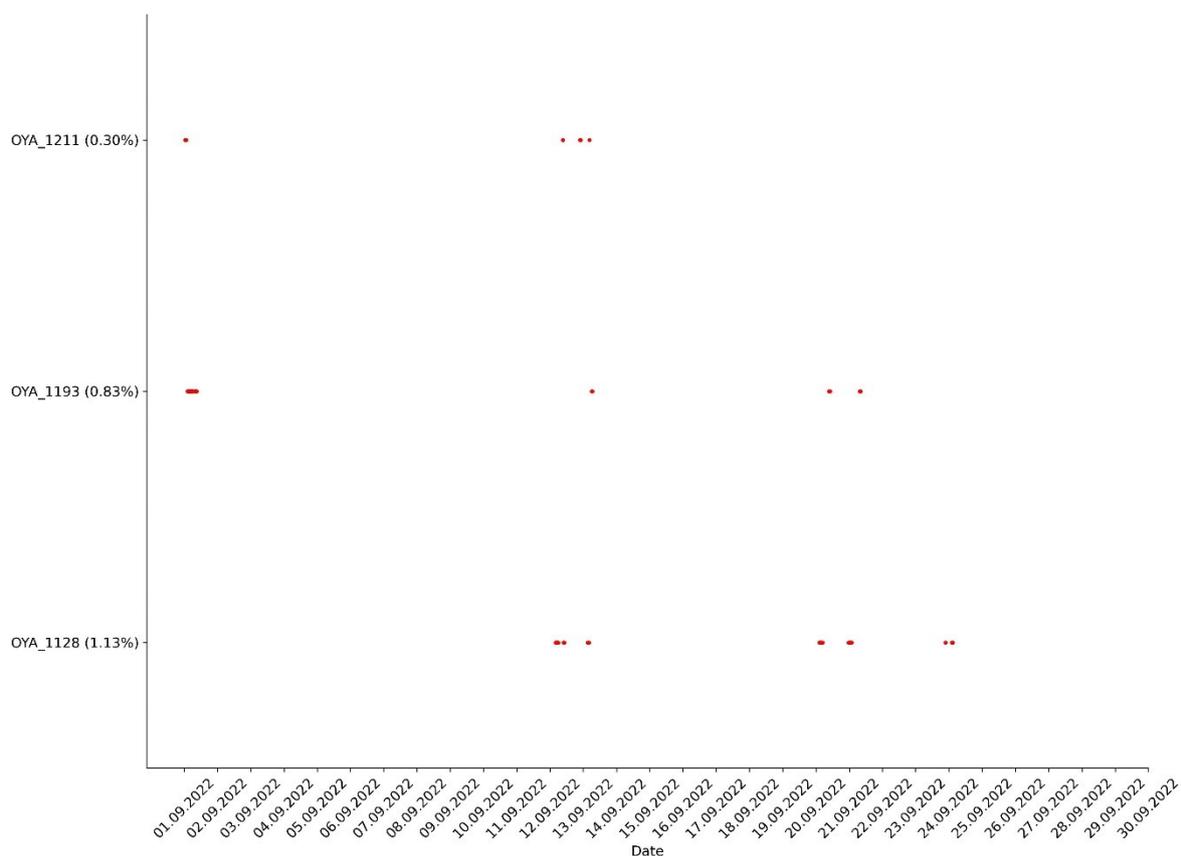


Figure 25: Standard plot of the check on the inter-stations correlation in a data series (System 1).  
Source: author's elaboration

#### 5.1.4. Filling data process

Differing from the prior stages of automated programming flows, the non-automatic filling or correction data process is conducted on a case-by-case basis. Its primary objective is to ascertain the effectiveness of the automated tests and the appropriateness of interval selection. Furthermore, potentially it serves as a mechanism for identifying novel error types (Cerlini et al., 2020).

After identifying the intervals data from filters in data quality flow, the intervals were categorized as short (less than or equal to 6 records), medium (more than one hour up to 12 hr), and long (more than 12 hr) (see Table 2). The objective is to correct or fill in these intervals using methods relevant based on their temporal resolution. For short intervals, a simple linear interpolation is applied to the respective climatic station. This involves calculating the average between the last recorded data and the first recorded data after the interval. For medium and long intervals, an inter-station correlation is employed with stations that constitute the same homogeneous fog climate measurement system.

This calculation involves determining the hourly average for medium intervals and the daily average for long intervals.

For Missing data and non-numeric values, short intervals show substantial proportions in OYA\_518 (100%) and OYA\_780 (90.6%), potentially reflecting the influence of specific climatic events or underlying technical challenges. Middle intervals present nuanced relationships through climatological stations, with OYA\_1128 (9.4%) suggesting a complex interplay of climatic influences and suddenly defecting data recording mechanisms. Long intervals reveal spatial nuances, particularly in OYA\_1211 (3.40%).

Moving to Continuous data analysis, Tukey's test and plausible value checks correlate with interval distributions. OYA\_780 demonstrates heightened anomalies in short intervals (70.8%), indicating potential transient, non-frequent, climatic events influencing specific data subsets. The OYA\_1211 dataset displays a significant increase in implausibility within middle intervals (33.4%), suggesting a complex interplay of technical challenges or unique climatic conditions influencing data recording.

In Non-continuous data examination, correlations emerge in Humidity thresholds, with OYA\_1069 showcasing robust compliance in short intervals (100%), reflecting reliability to climatic variations or proper instrument calibration of the fog climate network. OYA\_780's lower compliance suggests susceptibility to climatic fluctuations within specific data subsets.

The inclusion of Inter-stations correlation adds a new dimension, highlighting the importance of every homogeneous fog climate measurement system. In the short interval, the considerable 92.30% lack of correlation in System 1 and the 85.4% in System 2 suggest isolated instances where at least one station within each system did not collect fog while others did. This discrepancy could be attributed to localized variations or transient climatic conditions impacting the consistency of fog collection. Transitioning to the middle interval, the lower proportions of 7.70% and 14.6% absence of correlation in System 1 and System 2, respectively, align with the earlier hypothesis derived from the short intervals. The complete absence of correlation (0%) in the long interval might be attributed to the fact that the occurrence of not collecting fog in one station while the rest do is a topographically local and brief phenomenon, conditioned by local climatic phenomena that affect the non-correlated measurement. OYA\_1354 (System 3), is exempt from this correlation analysis, due its distinct nature that does not allow for correlation with other measurements within the same system.

## 5. Fog variability characterization in the Chilean Atacama Desert

Data filters	Interval	Climatological stations (OYA)							
		518	780	862	1069	1128	1193	1211	1354
Missing data and non-numeric values	Short	*	100%	*	*	90.6%	91.9%	88.2%	88.9%
	Middle	*	0%	*	*	9.4%	8.1%	8.4%	11.1%
	Long	*	0%	*	*	0%	0%	3.40%	0%
Tukey's test	Short	70.8%	63.8%	64.0%	66.9%	41.6%	44.8%	50.9%	58.3%
	Middle	27.4%	36.0%	36.0%	33.1%	58.2%	55.1%	49.0%	41.4%
	Long	1.7%	0.3%	0%	0%	0.1%	0.1%	0.2%	0.3%
<i>Continuous data</i> Plausible value check	Short	96.8%	95.7%	100%	96.7%	90.0%	_	66.6%	98.3%
	Middle	3.2%	4.3%	0%	1.0%	10.0%	_	33.4%	1.7%
	Long	0%	0%	0%	2.4%	0%	_	0%	0%
Check on a max. allowed variability of an instantaneous value	Short	100%	100%	100%	100%	100%	100%	100%	100%
	Middle	0%	0%	0%	0%	0%	0%	0%	0%
	Long	0%	0%	0%	0%	0%	0%	0%	0%
Check on a min. required variability of 1 hour values	Short	100%	100%	100%	100%	100%	100%	100%	100%
	Middle	0%	0%	0%	0%	0%	0%	0%	0%
	Long	0%	0%	0%	0%	0%	0%	0%	0%
<i>Non-continuous data</i> Humidity threshold	Short	82.6%	99.1%	90.3%	100%	97.8%	96.5%	99.8%	95.8%
	Middle	17.4%	0.9%	9.7%	0%	2.2%	3.5%	0.2%	4.2%
	Long	0%	0%	0%	0%	0%	0%	0%	0%
Inter-stations correlation	Short	85.4%	85.4%	85.4%	85.4%	92.3%	92.3%	92.3%	§
	Middle	14.6%	14.6%	14.6%	14.6%	7.70%	7.70%	7.70%	§
	Long	0%	0%	0%	0%	0%	0%	0%	§

\* There are no missing values in the available climate data series; however, some complete time intervals are entirely unavailable.

\_ No data recorded.

§ OYA\_1354 represents a fog measurement climate system by itself, therefore this data filter does not apply as it cannot be correlated with other measurements within the same system.

Table 2: Percentage of filtered data by climate station and data filter. Source: author's elaboration

## 5.2. Regional scale - Temporal analysis

### 5.2.1. Altitudinal and coast-inland variability of fog climate

Overall, the annual variability, range of values, and geographical influence on SFW collection stand out among the various climatological stations. March and April tend to show the greatest increases, whereas May and June tend to have lower accumulation and slower increases.

May records the highest accumulation of SFW for locations from the coast up to 5 km inland (862 masl), whereas at higher altitudes, the highest accumulation is observed in October (see Figure 26). The stations nearest to the coast exhibit relatively lower average SFW collection compared to other climatological stations. The standard deviation suggests a relatively low variability in SFW collection. The most significant increases occur from January to February, with increments below 0.2 L. The records remain relatively stable during the months of May and June (Austral Autumn), with lower accumulation and slower increases.

In the interior of the desert, the annual variability in SFW accumulation fluctuates, with averages ranging between 82.36 and 88.10 L. The most significant increases occur from March to April, with increments above 4.00 L and standard deviations ranging between 0.36 and 27.35. Stability is maintained in May and June, characterized by lower accumulation and slower increases.

Significant annual variability exists in SFW collection over the TIL, with averages ranging from 0.03 L to 3.67 L, and accumulation averages ranging between 34.21 L and 40.83 L. Standard deviations range between 2.19 L and 29.77 L. The data displays a wide range of values, with minimums of 0.0 and maximums of 500 L. The most significant accumulation increase occurs from March to April, with an increment of 4.98 L. Measurements indicate relative stability in May and June, with lower accumulation and slower increase.

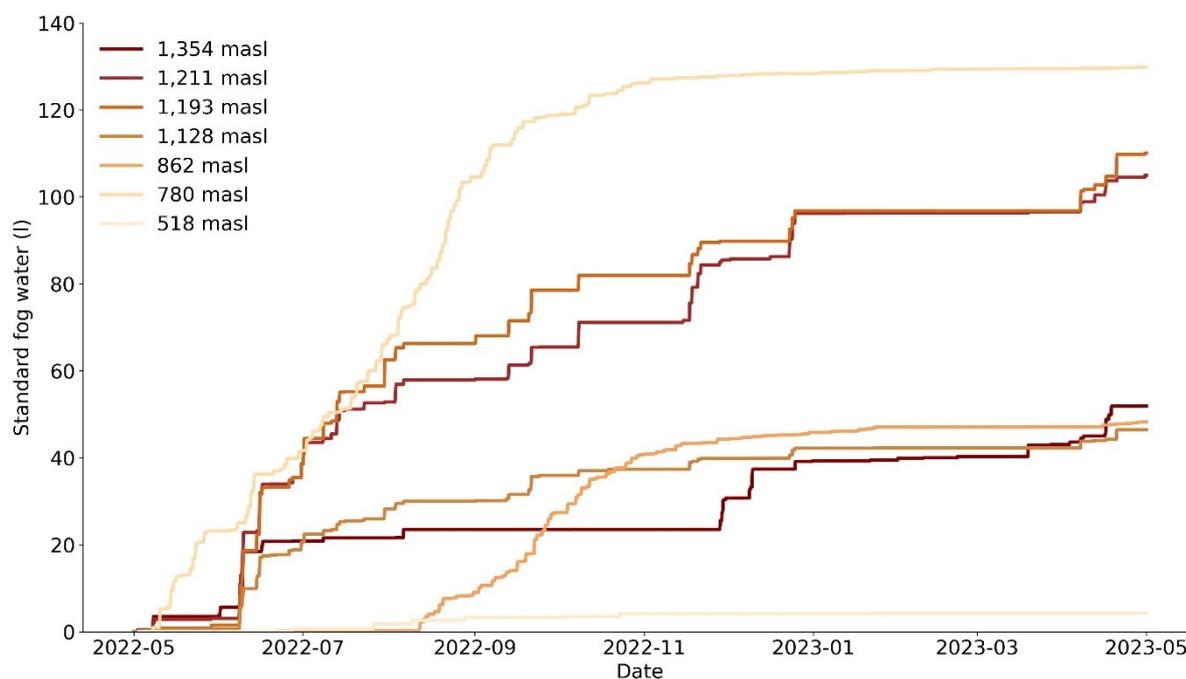


Figure 26: Annual cumulative SFW from coast to inland desert. Source: author's elaboration

In coastal areas, the presence of relatively stable AT and moisture conditions significantly contributes to a more consistent and concentrated collection of SFW. This phenomenon is illustrated in Figure 27, where a tighter distribution of SFW collected is observed. Overall, a concentration of lower values is observed (25% and 50%) and a wider dispersion towards higher values (75%). This suggests that in most climatological stations, the amount of SFW tends to be lower, but there are some cases where significantly higher quantities are recorded.

Kernel density estimation reveals that SFW collection is highly concentrated around the median of fog data recorded along the coast, ranging from as low as 30 ml at 1,128 masl. In contrast, the IQR inland desert exhibits lower values, ranging from 112 ml at 1,211 masl. The highest values for SFW collection are observed inland desert, with measurements reaching 524 ml at 1,193 masl. Similarly, the median SFW collection is higher at 46 ml at 1,211 masl, and the average stands at 79.2 ml at 1,193 masl. Above the TIL, a similar dispersion pattern to the Oyarbide variability is observed, albeit with a slightly lower average SFW collection of 67.9 ml.

When considering the different altitudes, it is observing that there is no clear linear relationship between altitude and the amount of SFW. The SFW records at 862 masl has a lower average value

than the station at 780 masl. This suggests that other factors, such as local topography, may influence the amount of recorded SFW, in addition to altitude.

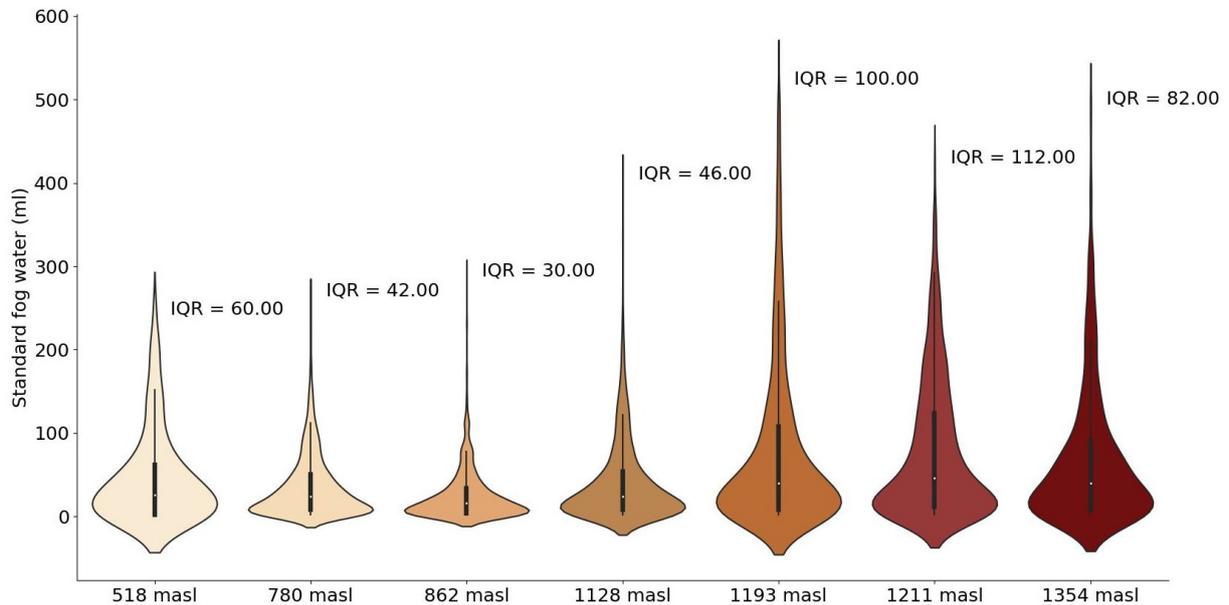


Figure 27: Variability of the collected SFW from coast to inland desert. Source: author's elaboration

Wind patterns during fog events exhibit temporal variability. The wind speed along the coastal to inland desert transect of climatic records remains relatively constant but shows an increase during the afternoon and night, possibly due to heightened atmospheric turbulence. The wind direction, typically originating from the southwest (SW) from the Pacific Ocean, exhibits an inverse pattern during fog events (see Figure 28).

Near the coast (518 masl) during the early time hours (2:00 - 6:00 hr), the average wind direction shows values around  $336.56^\circ$  and  $337.97^\circ$ . This indicates a prevailing trend of north-northwest (NNW) winds during these periods with fog records, with average wind speeds ranging from 3.75 to 5.58 m/s. From 6:00 to 16:00 hr, the average wind direction varies between  $103.51^\circ$  and  $180.00^\circ$ , and the average wind speed ranges from 2.69 to 7.97 m/s. This suggests that during the morning and midday, wind direction is more variable compared to the early hours, predominantly from the southeast (SE), with significant fluctuations in wind speed.

Going into the desert at 862 masl during the early time ranges (0:00 - 6:00 hr), the average wind direction is consistently between  $31.69^\circ$  and  $42.09^\circ$ . This indicates that in that specific context of the

Atacama Desert, ca. 5 km inland from the coast, there is a prevailing NE wind direction during the early hours of the morning. From 6:00 to 22:00 hr, the average wind direction varies widely between  $1.06^\circ$  and  $305.46^\circ$ . This suggests significant changes in wind direction during the day within a broad range. Between 22:00 and 24:00, the average wind direction returns to early morning values, around  $313.27^\circ$  (NNW). The average wind speed remains relatively constant throughout the day, with values ranging from 1.47 to 3.04 m/s. This indicates that in the Atacama Desert, ca. 5 km inland, the wind speed remains at low to moderate levels throughout the day.

The records furthest from the coast, 11 km inland at 1,211 masl, show an average wind direction between  $4.64^\circ$  and  $36.36^\circ$  during the early hours (00:00 - 06:00). This suggests a prevailing NE wind direction during the early hours of the morning. From 06:00 to 10:00 hr, the average wind direction varies between  $33.5^\circ$  and  $36.0^\circ$ . This suggests minor variations in wind direction during the morning, generally within a narrow range. From 10:00 to 22:00 hr, the average wind direction fluctuates widely between  $32.13^\circ$  and  $242.20^\circ$ . This indicates greater variability in wind direction during the day, within a broad northwest (NW) to southwest (SW) range. Between 22:00 - 24:00 hr, the average wind direction returns to values close to the early time ranges, around  $249.82^\circ$  (WSW). The average wind speed remains moderate throughout the day, ranging from 2.28 to 5.43 m/s. This indicates that inland in the desert, the wind speed can be relatively constant during the day, with occasional higher peaks in the afternoon. During the morning, the wind direction remains within a narrower range, while during the day, it becomes more variable.

## 5. Fog variability characterization in the Chilean Atacama Desert

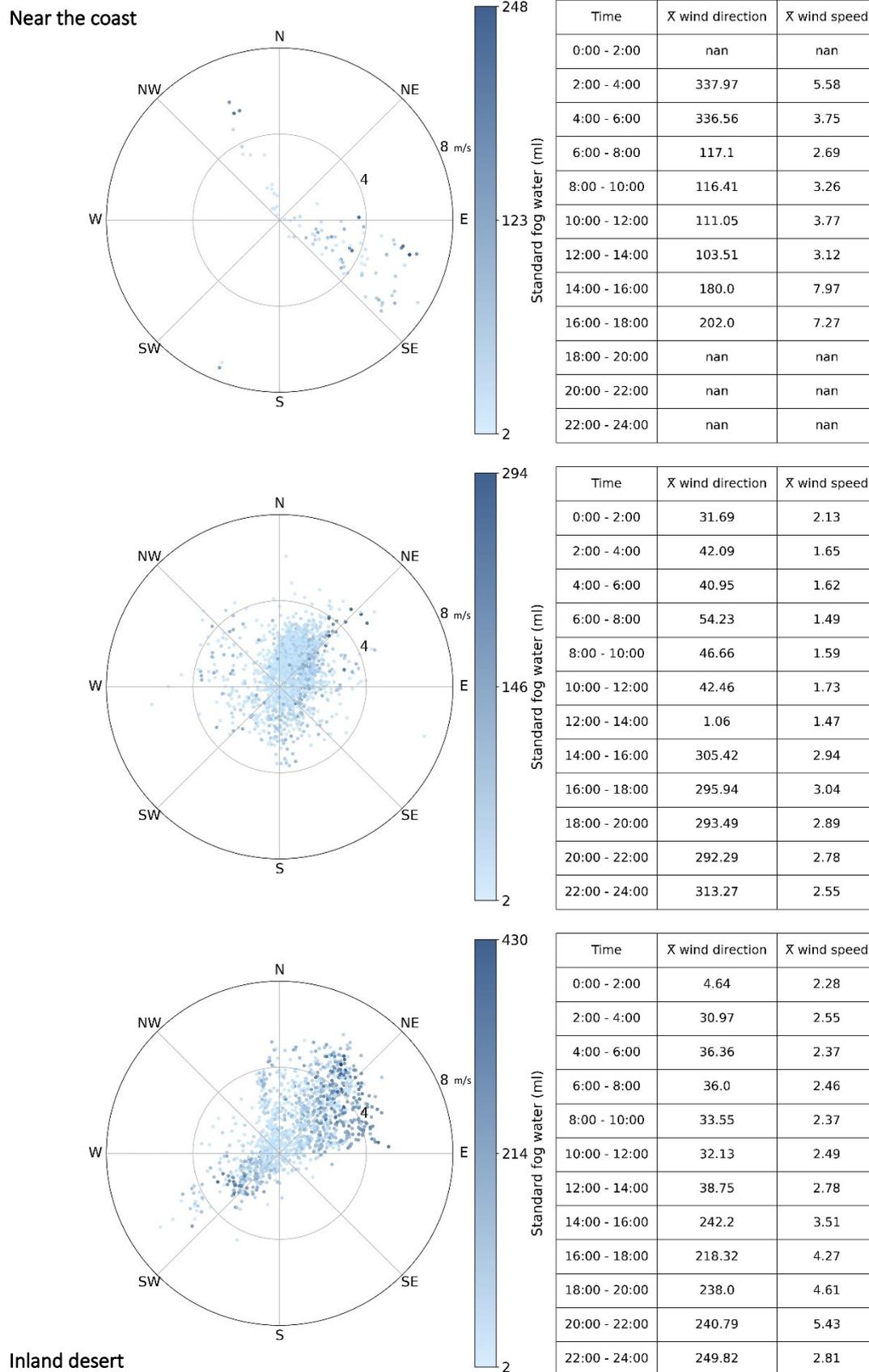


Figure 28: Wind patterns gradients from coast (518 masl) to inland desert (862 and 1,211 masl).

Source: author's elaboration

The northern direction exhibits the most significant influence, represent 21% of the frequency of records. This implies that when the wind originates from N, the probability of fog is notably higher. Significant correlations are also recorded for SW wind patterns (see Figure 29). Conversely, in fog-free conditions, the correlations between wind direction and fog probability are less pronounced. Scenarios with N and NE wind patterns also display high correlations. In this scenario, the influence of wind direction on fog probability is less significant. In these cases, wind direction continues to impact fog probability, albeit to a lesser extent than in foggy conditions.

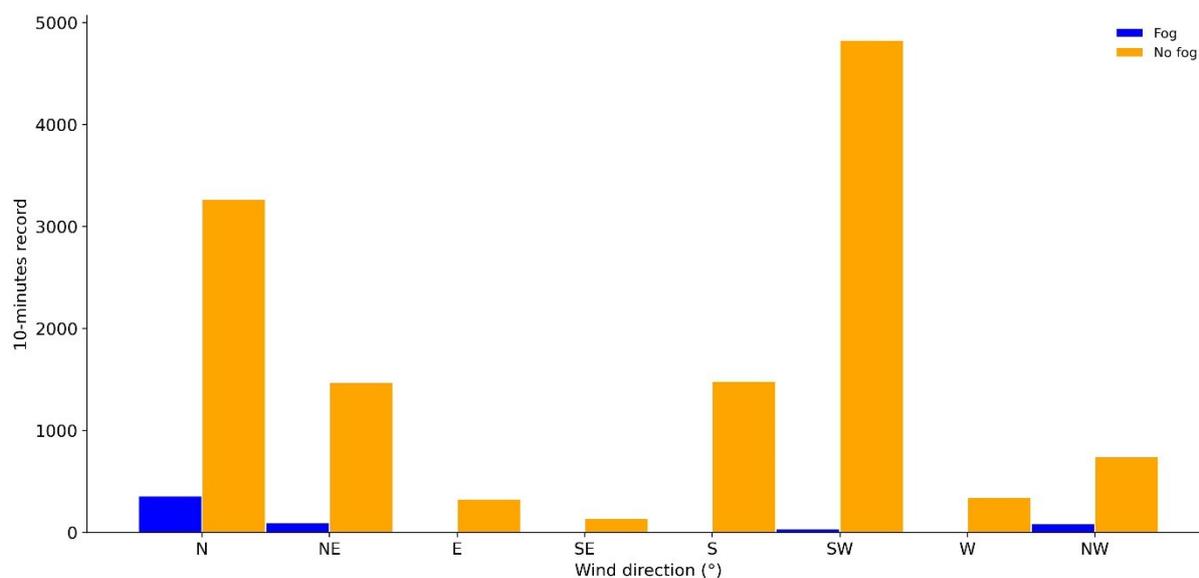


Figure 29: Relationship between the amounts of SFW collected and wind patterns (1,211 masl).  
Source: author's elaboration

A consistent variability is observed in the records of dew point (DP), fog collection record, and SFW amounts across the months inland desert (see Figure 30). The variables exhibit a positive trend, with the proportion of records reaching the DP displaying greater variability compared to the fog collection and SFW amounts.

The highest annual correlation between SFW collected and DP records occurs during the months of June, July, and August, with coefficients of 0.44, 0.39, and 0.44, respectively. Conversely, the lowest correlation is observed during the Austral summer months (January, February, and March), with values approaching 0. The average DP values are higher in January and February (Austral Summer) and lower in May and August. This indicate a seasonality in the conditions under which the DP is

reached, associated with the movement of the TIL, which reaches its highest point during summer with ca. 1,230 masl (Del Río, 2019) overpassing the climatological stations of the recording.

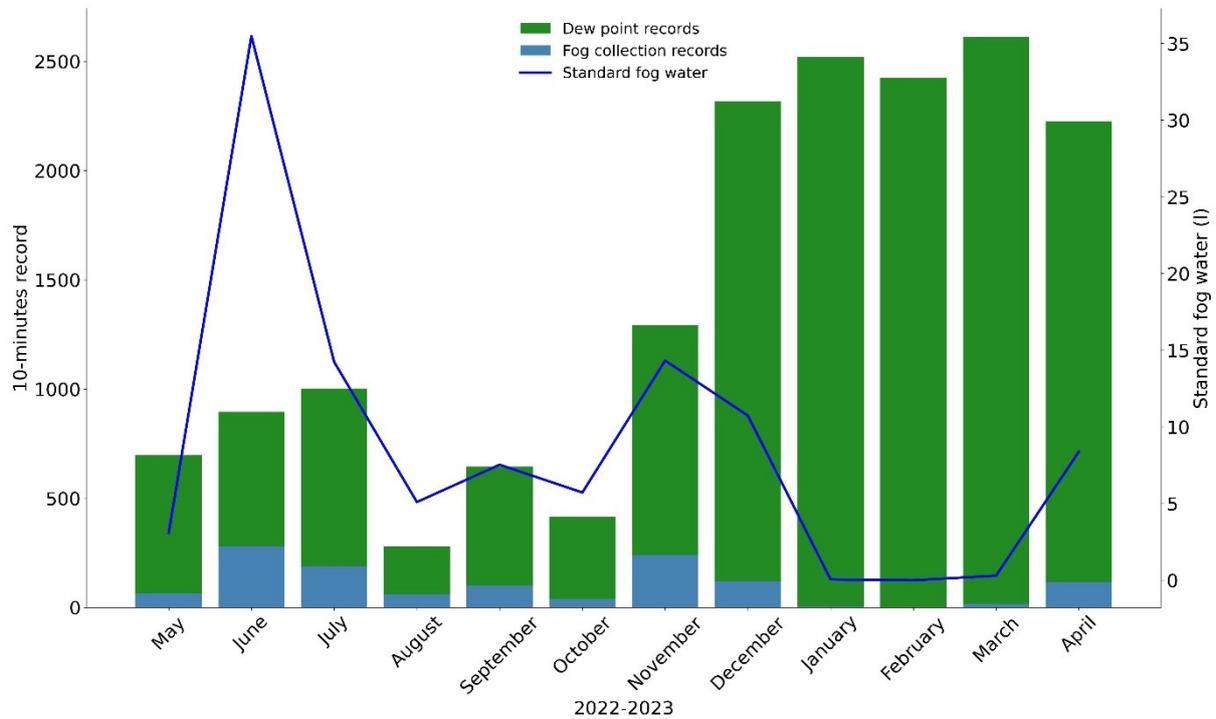


Figure 30: Fog annual variability and its relationship with SFW collection and DP records. Source: author's elaboration

In December, January, and February, the records reaching the DP exhibit significantly higher values, indicating consistently high RH in the air. A considerable high records of fog collection and SFW amounts is also recorded. These patterns suggest that during the Austral Summer, there is a higher probability of fog formation and humidity in the Atacama Desert. This is related to the influence of cold coastal winds interacting with the hot and dry desert air, creating favorable conditions for fog formation.

Throughout March, April, and May, the records reaching the DP continue to exhibit a high frequency compared to other months. Although the fog collection recording and SFW amounts decrease, a significant presence is still observed. This indicates that during the Austral Autumn, there is still be some RH and potential for fog formation in the Desert, although to a lesser extent than in summer.

During June, July, and August, the DP records decrease considerably, indicating a lower presence of humidity in the air. The records of SFW collected and the amount of SFW also diminish significantly.

## 5. Fog variability characterization in the Chilean Atacama Desert

---

These patterns suggest that during the Austral Winter the air becomes drier, and the probability of fog occurrence is reduced, then the conditions for RH and fog formation are much less frequent.

During the months of September, October, and November, the DP records, SFW collected recording, and SFW amounts show an increase compared to Winter, although they do not reach the levels observed in Summer and Autumn. This suggests that during the Austral Spring, RH and fog formation gradually begin to increase in the Desert, preparing for the wetter conditions of Summer.

The records of LW also provide valuable data to understand the seasonal changes in moisture levels and their connection to the TIL variability (see Figure 31), which reaches its maximum height in Austral Summer-Autumn, and lowest in Austral Winter-Spring reaching lower altitudes and less penetration into the desert, affecting the distribution of moisture and humidity in the region.

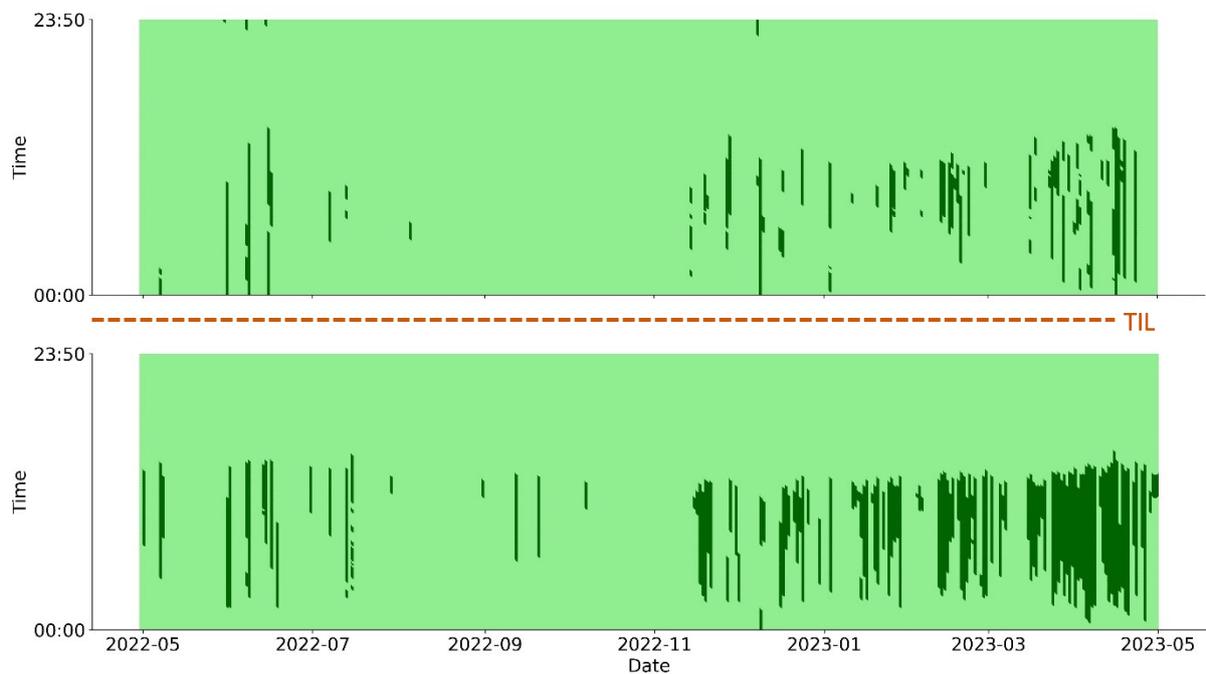


Figure 31: LW over 70% recording fog above and below TIL (1,354 masl and 1,211 masl). Source: author's elaboration

### 5.2.2. Monthly variability of fog climate

The Figure 32 presents a detailed overview of the monthly accumulation of SFW inland Atacama Desert (ca. 11 km). The Austral Autumn stands out as the period with the highest accumulation of SFW, representing 54.81% of the annual total. Within this period, April exhibits the highest SFW accumulation, reaching 55.41 L. On the other hand, the Austral Summer experiences the lowest accumulation of SFW annually, representing only 5.35% of the total. February emerges as the month with the least SFW accumulation, with less than 1 L (0.72 L) of accumulated SFW.

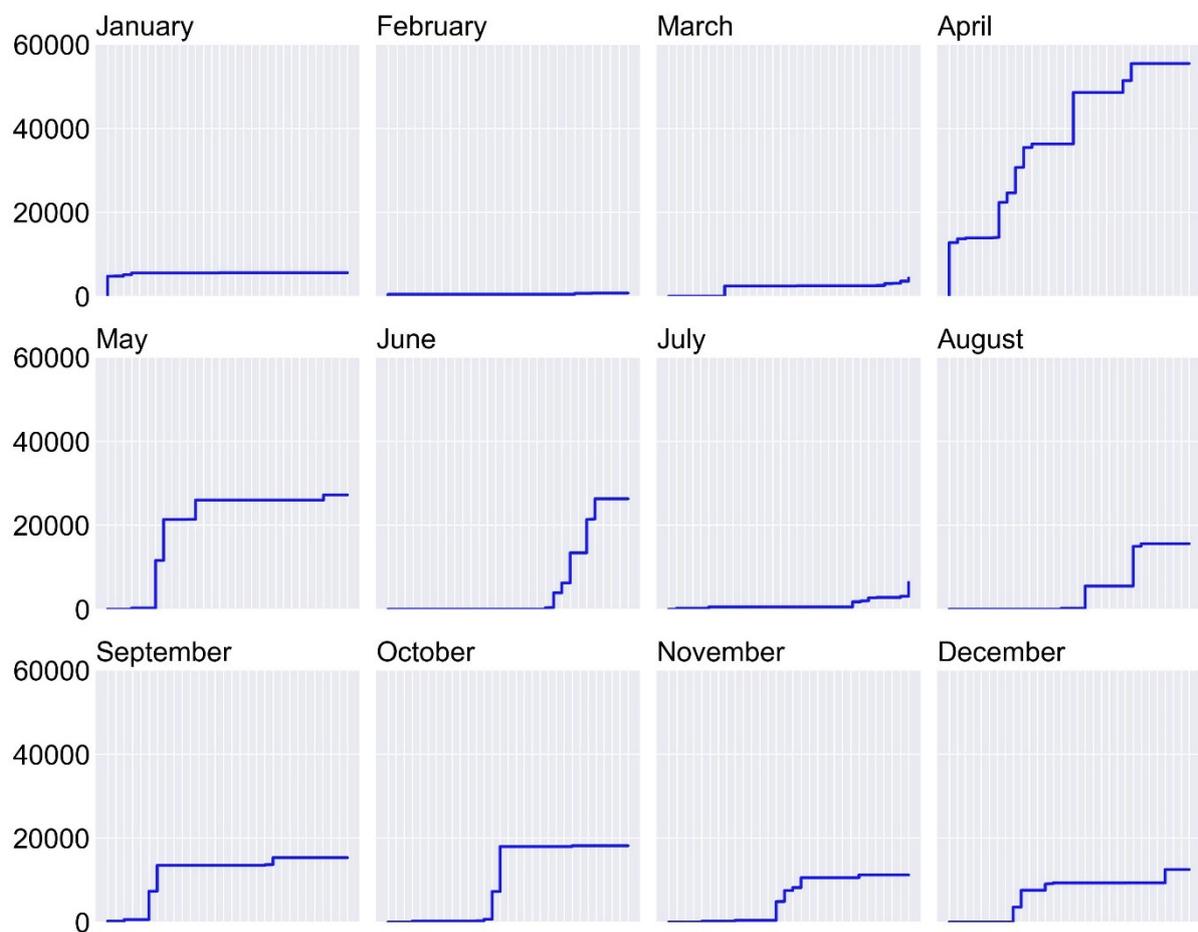


Figure 32: Monthly cumulative (ml) SFW collected inland desert (year 2020). Source: author's elaboration

The fog variability events in the Atacama Desert, defined for this research as 3 continuous 10-minute records of collected water, are influenced by seasonal and local factors. The Figure 33 shedding light on the interaction between fog events variability in size and duration and seasonal patterns.

There is a growing trend of fog events during the Austral Winter inland desert. Among the months analyzed, July stands out with the highest number of fog events, totaling 19 occurrences. However, June (35.07 L) and November (14.30 L) exhibit the greatest amounts of SFW. No fog events are recorded in January and February (2023), while March only records a single event. This annual variability suggests the influence of seasonal and local climatic factors on fog formation. The longest-lasting fog events are observed in June, with an average maximum duration of 4.97 hr, and in July, with maximum durations of 14.83 hr and 6.33 hr, respectively. These months exhibit greater atmospheric stability and lower temperatures, which facilitate the formation and persistence of fog over more extended periods.

A correlation coefficient ( $r$ ) of 0.60 exists between the number of fog events and the amount of collected SFW. Winter months, particularly June and July, showcase the highest quantity and duration of fog events due to the prevailing lower temperatures during these months. In June, the largest quantities of SFW are collected, reaching a maximum of 9.88 L. Additionally, June also exhibits the longest-lasting events, with a maximum duration of 14.8 hr. On the contrary, July records the smallest fog collection events (0.008 L) and a significant proportion of shorter events, with a minimum duration of 0.5 hr. In October, a perfect correlation (1) is observed between the amount of SFW collected and the duration of the two recorded events. August and September also display high correlations, with values of 0.97 and 0.98, respectively. However, November exhibits the lowest annual correlation between the amount of SFW and the duration of events, with a value of 0.53.

## 5. Fog variability characterization in the Chilean Atacama Desert

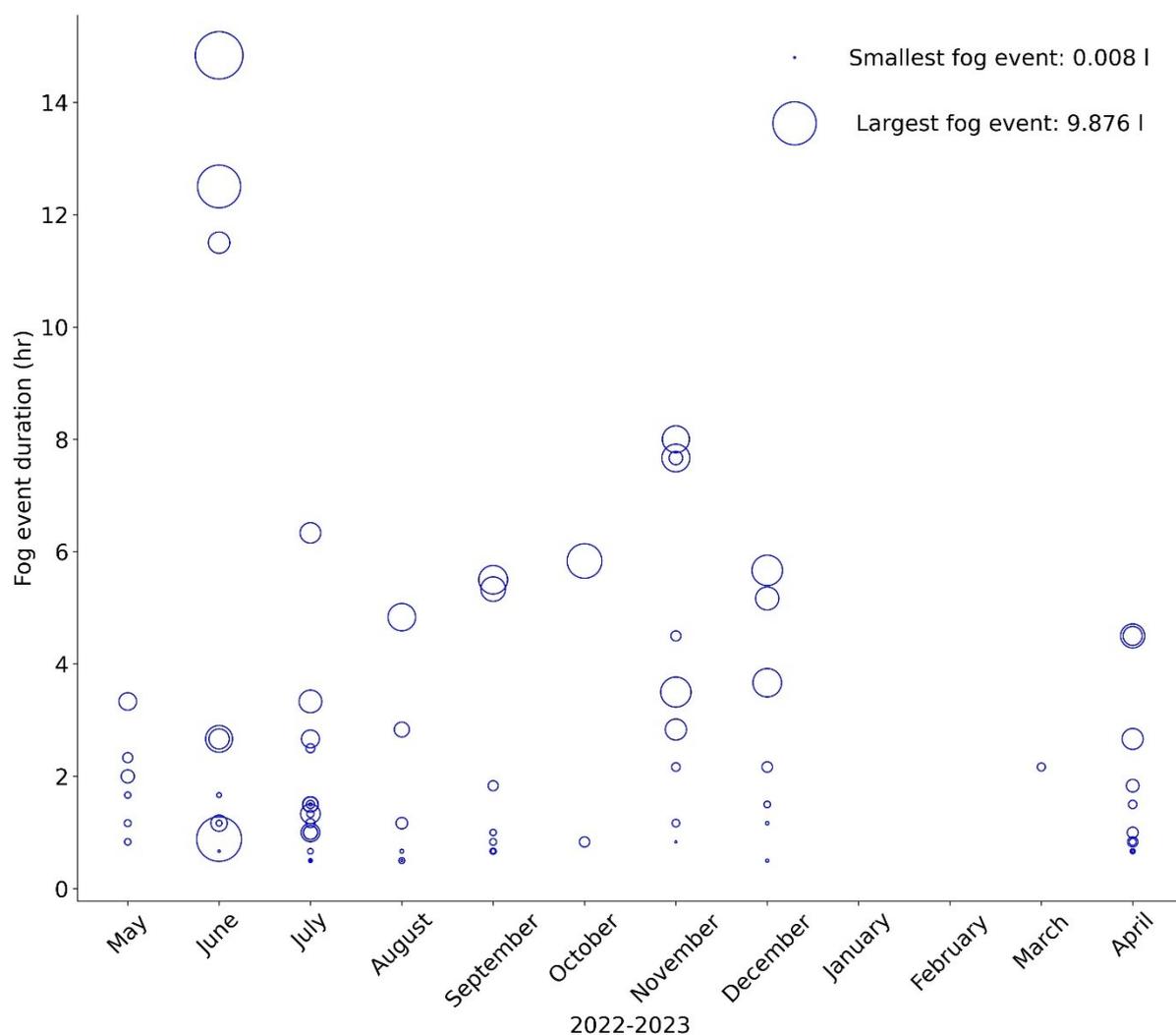


Figure 33: Annual gradient of fog events inland desert. Source: author's elaboration

The variability of RH under foggy conditions demonstrates that high RH is a key factor in fog formation. Coastal areas exhibit elevated levels of RH due to their proximity to the ocean, while inland desert regions are characterized by aridity (see Figure 34). Indeed, more than 97% of fog occurrences both along the coast and inland desert are associated with RH levels exceeding 80%. The disparities between coastal and inland desert regions are pronounced.

The increased humidity in coastal areas facilitates fog formation and enhances the potential for SFW collection. Near the coast (518 masl), most records fall within the RH ranges of 90 to 100%, representing 95.71% of the total records. This indicates that fog formation is more likely under high RH conditions.

The 60 to 70% and 70 to 80% ranges display standard deviations of 29.03 and 51.86, respectively, suggesting higher variability in RH when fog occurs within these ranges. From 90 to 100% RH range

accounts for 95.71% of the total records, confirming the greater importance of high RH in fog formation near the coast compared to the inland desert. Along the coast, fog collection takes place at an average RH with a declining trend throughout the year. However, over 95% of fog collection and fog quantity are recorded at RH levels exceeding 90%. On the other hand, in the inland desert, there is an annual increasing trend in RH during fog collection events, with more than 90% of fog occurrences and over 87% of the collected fog quantity observed at RH levels above 90%. During the summer, the collection of fog near the coast represents the period of lowest RH, reaching a minimum of 64.93% in March. Throughout more than half of the year, the average RH during foggy conditions remains above 90%, with peaks in October (96.78%). In the absence of fog, the annual average drops from 86.22% to 75.79%, with the least humid months occurring during summer, particularly in January (67.58%). Conversely, the winter months exhibit the highest average RH, with September marking the peak at 84.61%.

Inland desert (1,211 masl), months with fog collection display average RH levels surpassing 90%, with a maximum of 99% in April. The distribution of RH under foggy conditions shows that most records fall within the RH ranges of 90 to 100%. This range represents 90.05% of the total records, suggesting that fog formation is highly likely under high RH conditions. Additionally, a greater dispersion of data is observed in the 80 to 90% range, indicating a more pronounced variability in RH compared to other ranges. January exhibits the lowest average RH at 75.68%, excluding February, which does not record fog collection (2023). This leads to an annual "anomaly" as the winter and southern spring months record the lowest average RH levels with fog occurrences. In the absence of fog, the annual average declines from 93.68% to 47.76%, characterized by pronounced seasonal variation in RH during summer and the first half of autumn, with a peak in April at 75.44%. Furthermore, the winter and early spring experience low average RH, reaching a minimum in August at 20.67%.

In general, considering the fog records along the coastal to inland desert transect, the 90 to 100% RH ranges are the most prevalent during foggy conditions. Additionally, a greater dispersion of data is observed in the 60 to 70% and 70 to 80% ranges, indicating a more pronounced variability in RH compared to other ranges.

## 5. Fog variability characterization in the Chilean Atacama Desert



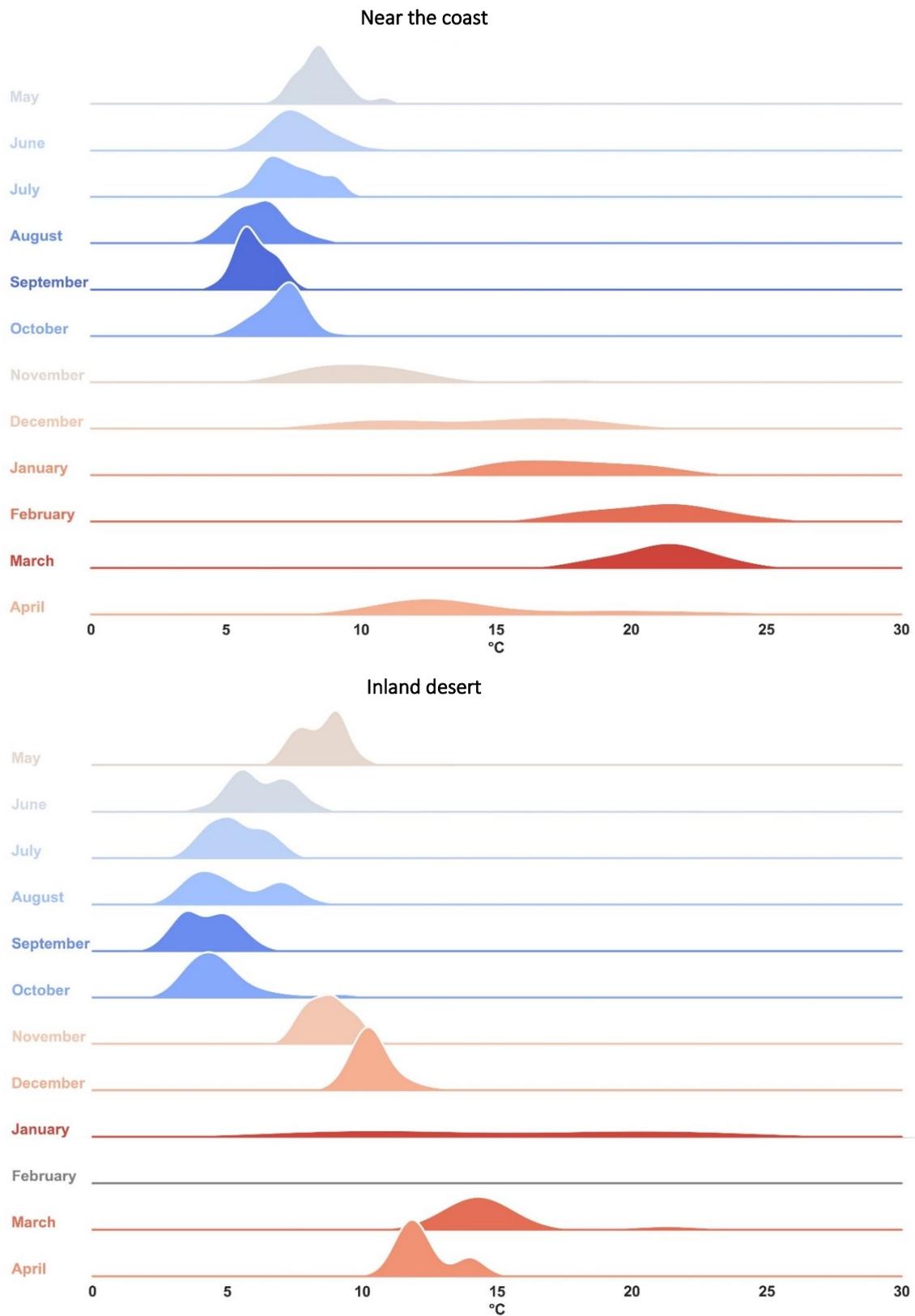
Figure 34: RH gradient near coast (518 masl) and inland desert (1,211 masl). Source: author's elaboration

The gradient of AT during SFW collection exhibit significant variability throughout the year. The Figure 35 presents the data on the annual temperature range in both the coastal and inland regions of the Atacama Desert. Near the coast, over 95% of fog occurrences are associated with AT below 10°C, while this percentage decreases to slightly above 86% in the desert's interior.

Close to the coast, the dispersion of temperature data around the mean varies between 0.93°C and 3.90°C (annual standard deviation of 1.90°C). There is relatively high data dispersion in months with average temperatures above 14°C. In the inland desert, the average temperature range during fog events varies between 1.04°C and 2.05°C (annual standard deviation of 1.38°C) with relatively high data dispersion in months with average temperatures above 10°C. Fog collection is accompanied by an average AT that exhibits a decreasing trend throughout the year. More than 98% of fog collection events and the corresponding amount of fog collected occur at temperatures below 15°C. Similarly, in the inland desert, there is an annual decreasing trend in AT during fog collection, with over 99% of fog occurrences and the quantity of fog collected recording temperatures below 15°C.

A pronounced bimodal temperature trend is observed, with a temperature difference of 15.03°C between the lowest and highest monthly averages. The period with low temperatures during fog collection corresponds to the winter months and the first half of spring, with a minimum in September (6.14°C), which also exhibits the lowest thermal variability (0.93°C standard deviation). On the other hand, the highest temperature records occur in summer, with a maximum average of 21.17°C in March. The difference between the highest and lowest monthly temperature averages is only 6.31°C. Additionally, oscillations are present that disrupt the thermal trend during certain periods. The monthly temperature average shows a maximum decrease of 5.96°C in October, accompanied by low temperatures throughout the first half of the period, and an increase in summer up to 12.25°C in January. Thermal oscillations are pronounced during the summer period, with a maximum of 2.05°C in December.

## 5. Fog variability characterization in the Chilean Atacama Desert



\*No fog collection in February 2022 inland desert (OYA\_1211)

Figure 35: AT gradient near coast (518 masl) and inland desert (1,211 masl). Source: author's elaboration

Fog has a dual effect on both diurnal and nocturnal temperatures, while also contributing to increased RH levels. When fog settles in this region, it acts as an insulating layer that decreases heat during the night and mitigates daytime warming (see Figure 36). Consequently, a discernible decrease in both minimum and maximum temperatures is observed during fog events, accompanied by an evident rise in RH. This climatic phenomenon profoundly influences the thermal regulation of the desert, creating a comparatively cooler and more humid environment in comparison to the usual conditions.

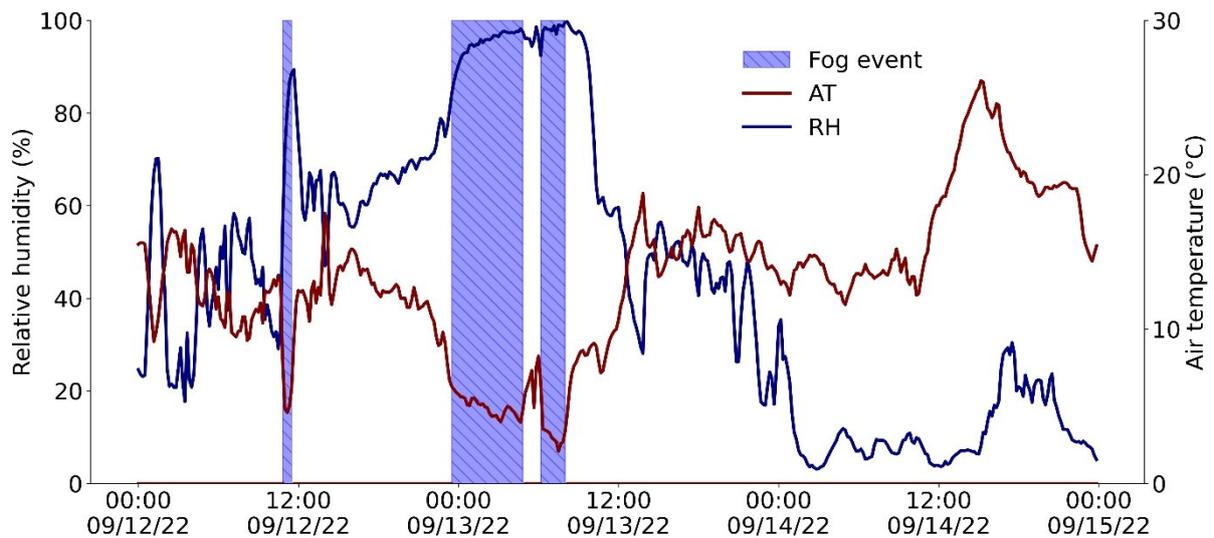


Figure 36: Day and night fog events and its relationship with AT and RH. Source: author's elaboration

The correlation between fog records and AT is negative, while the correlation with RH is positive. Since the beginning of the fog climate network records, the correlation in both the coastal (780 masl) and inland desert (1,211 masl) areas exceeds +/- 0.3 for both variables, with a higher correlation observed for AT (-0.38 and -0.37, respectively) compared to RH (0.32 for both measuring stations).

### 5.2.3. Daily variability of fog

The dynamics of fog collection throughout the day provides valuable insights into its occurrence and variability. Significant temporal variability in SFW collection is observed by the presence of the TIL and its seasonal variability (see Figure 37). Throughout the year, distinct seasonal variations in fog occurrence were observed. In the spring months, such as April and May, fog was present during the early morning hours, with an average occurrence of approximately 80% of days. As the day progressed, fog gradually dissipated, and by midday, the occurrence reduced to around 50% of days. In contrast, during the summer months, like June, July, and August, fog occurrence exhibited a different pattern. It was relatively less frequent during the early morning hours, with an average occurrence of around 33% of days. However, as the day advanced, fog occurrence gradually increased, reaching a peak of approximately 60% of days in the late afternoon or evening.

There is lower or negligible fog formation between 14:00 and 22:00 hr, accounting for only 1.96% of the annual hourly fog formation. In contrast, the highest levels of fog formation are recorded between 6:00 and 10:00 hr encompassing 42.50% of the annual hourly formation. This daily fog variability aligns with previous findings on fog frequency presence (Del Río, 2019), which indicates that its peak occurrence is during the night and dawn (~40%).

During the Austral Winter, when the TIL is lower, there is greater fog formation throughout the day. On an average year, August stands out as a month with fog formation present in most hours, except in the 20:00 hr range. This suggests that the lower height of the TIL during the season allows for greater moisture retention in the atmosphere and, consequently, increased fog formation during the day. In contrast, during the Austral Summer, when the TIL reaches its maximum height, fog formation is primarily concentrated in the early hours of the morning, with minimal presence during the day. This hourly distribution is reflected in the period with the lowest fog formation of the year, with February recording the lowest value of only 0.7 L.

When examining the specific hourly variability above the TIL (1,354 masl), no clear decreasing trend in fog formation at specific times is evident. The lowest values are generally observed between 18:00 and 20:00 hr, while there is no distinct pattern of high fog formation at a particular time. However, peak values are observed at 10:00 hr, followed by significant values in the early evening hours.

## 5. Fog variability characterization in the Chilean Atacama Desert

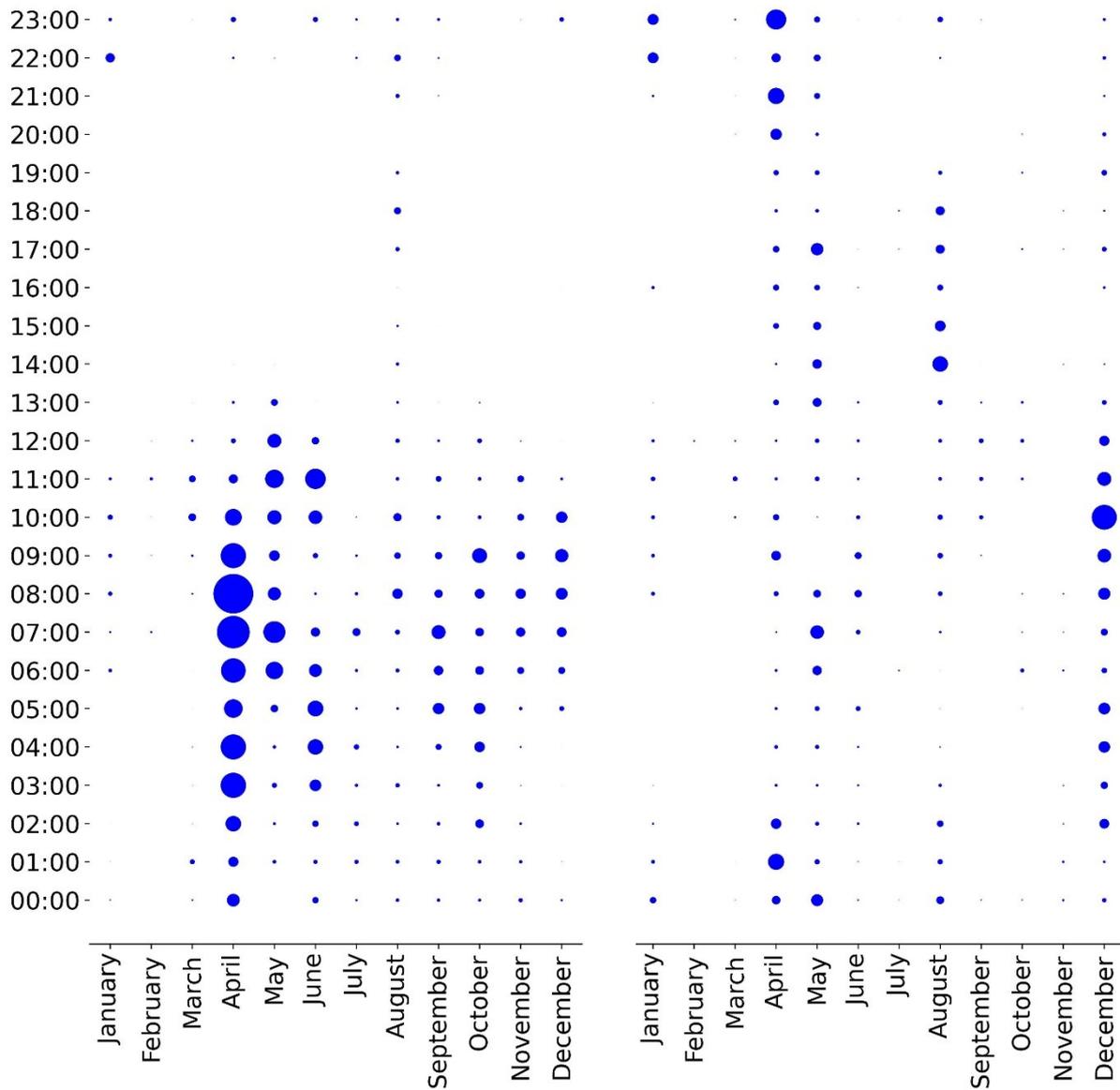


Figure 37: Hourly distribution of SFW collection below (1,211 masl) and above TIL (1,354 masl).  
Source: author's elaboration

In terms of monthly distribution, April, and May exhibit fog formation in almost all hourly records. These months coincide with the second and third consecutive months with the highest annual fog formation, slightly lower than December, which stands out as the month with the highest annual fog formation. Months with fewer hourly ranges of fog collection do not exhibit a clear trend, with February being the only month that records a single fog formation range between 12:00 and 13:00 hr, coinciding with the month of lowest formation (0.12 L).

Under the TIL, the hourly fog collection pattern shows that the month with the highest SFW concentration is in Autumn, with highest concentration in April, with 55.41 L. The monthly average SFW concentration is highest in April at 4.26 L and lowest in February at 0.06 L. Above the TIL, the month with the highest total SFW concentration is December, with a value of 30.93 L. The monthly average SFW concentration varies across different months, being highest in December at 23.80 L and lowest in October at 0.18 L.

The analysis of fog occurrence and its relationship with hourly temperature variability reveals distinct seasonal patterns. In periods of largest records of fog collection, such as during the spring and summer months, average daily temperatures tend to exhibit smaller fluctuations compared to periods with lower fog occurrence. For instance, during foggy days in April, temperature variations range from 10.50°C to 12.80 °C, while on clear days, the range extends from 9.11°C to 11.70°C. This suggests that fog acts as a temperature regulator, preventing excessive heating during the day and maintaining relatively stable temperatures (see Figure 38). The influence of fog on daily temperature patterns can be attributed to its cooling effect. During foggy mornings, the presence of water droplets in the air reduces solar radiation reaching the surface, resulting in lower temperatures. As the fog dissipates and clears during the day, more solar radiation can penetrate, leading to increased daytime temperatures.

The variability in fog records and its relationship with temperature can also vary within each season. Further analysis reveals that in April, fog occurrence ranges from approximately 80% to 90% of days, while in May, it ranges from 70% to 80% of days. Fog acts as a temperature regulator, reducing daily temperature fluctuations and influencing local climate dynamics. Fog tends to be present during the morning hours in spring, dissipating as the day progresses representing approximately 80% of days. In contrast, during the summer months, fog occurrence increases throughout the day, peaking in the late afternoon or evening, representing approximately 33% to 60% of days.

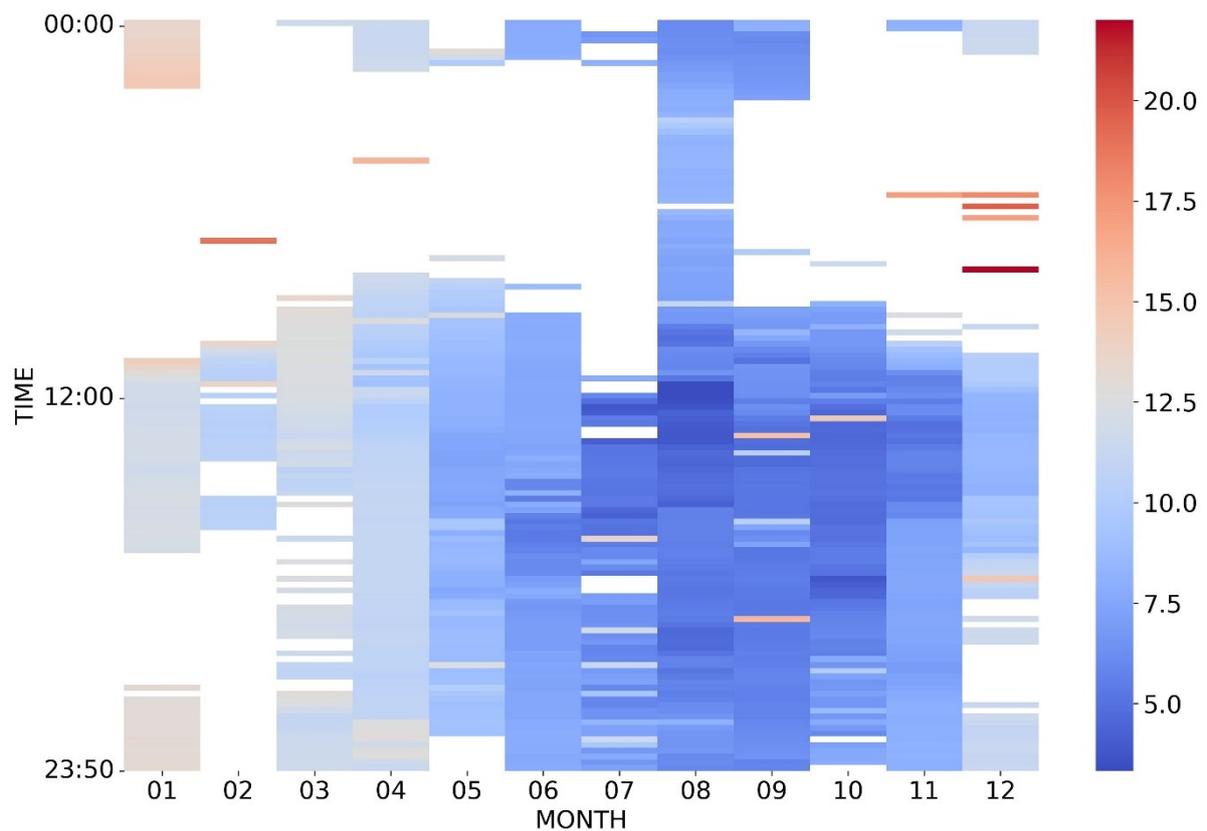


Figure 38: 10-minute average AT and SFW collection records at 1,211 masl. Source: author's elaboration

### 5.3. Local scale - Spatial analysis

#### 5.3.1. Near-surface fog water measurements and explanatory variables

An overview of Near-surface fog water data across distinct altitudinal tiers delineates discernible fog collection patterns. Higher altitudes exhibit elevated levels of fog water, whereas intermediate and lower altitudes exhibit diminished concentrations. The variability in fog water collection presents a spatial distribution pattern that concentrates large amounts of fog water collected in the southeast sector, especially above 1,180 masl. Conversely, there is a negative trend of collection below this threshold, spatially concentrated in a north-southwest strip of Oyarbide (see Figure 39).

At higher elevations (1,213, 1,208, 1,185 masl), there is a consistent recording of larger quantities of fog water. This pattern presents higher mean values and wider ranges between minimum and maximum readings at these altitudes, indicative of favorable topographic conditions conducive to fog formation and accumulation. The correlations between these Mini FCs show high and consistent positive values.

Measurements collected at intermediate elevations (1,191, 1,184, 1,165 masl) display a moderate collection of Near-surface fog water. Despite exhibiting lower average values compared to higher altitudes, appreciable levels of fog persist. These Fog Collectors show positive correlations, although with slightly lower values than those observed at higher altitudes.

Lower elevations (1,206, 1,163, 1,162, 1,135 masl) exhibit diminished Near-surface fog water content compared to higher altitudes. The notably lower averages and narrower ranges between minimum and maximum values at these elevations imply reduced fog accumulation. Feasible features include drier climatic conditions on this area or reduced topographic exposure to fog-generating meteorological phenomena at these lower altitudes. The Mini FCs at lower altitudes present variable correlations, with lower values compared to those at higher and intermediate elevations.

In whole measurements records of Near-surface fog water, the largest collections are recorded during the May-June and September-October periods, and the lowest collection period corresponds to the Austral Summer (November-January).

## 5. Fog variability characterization in the Chilean Atacama Desert

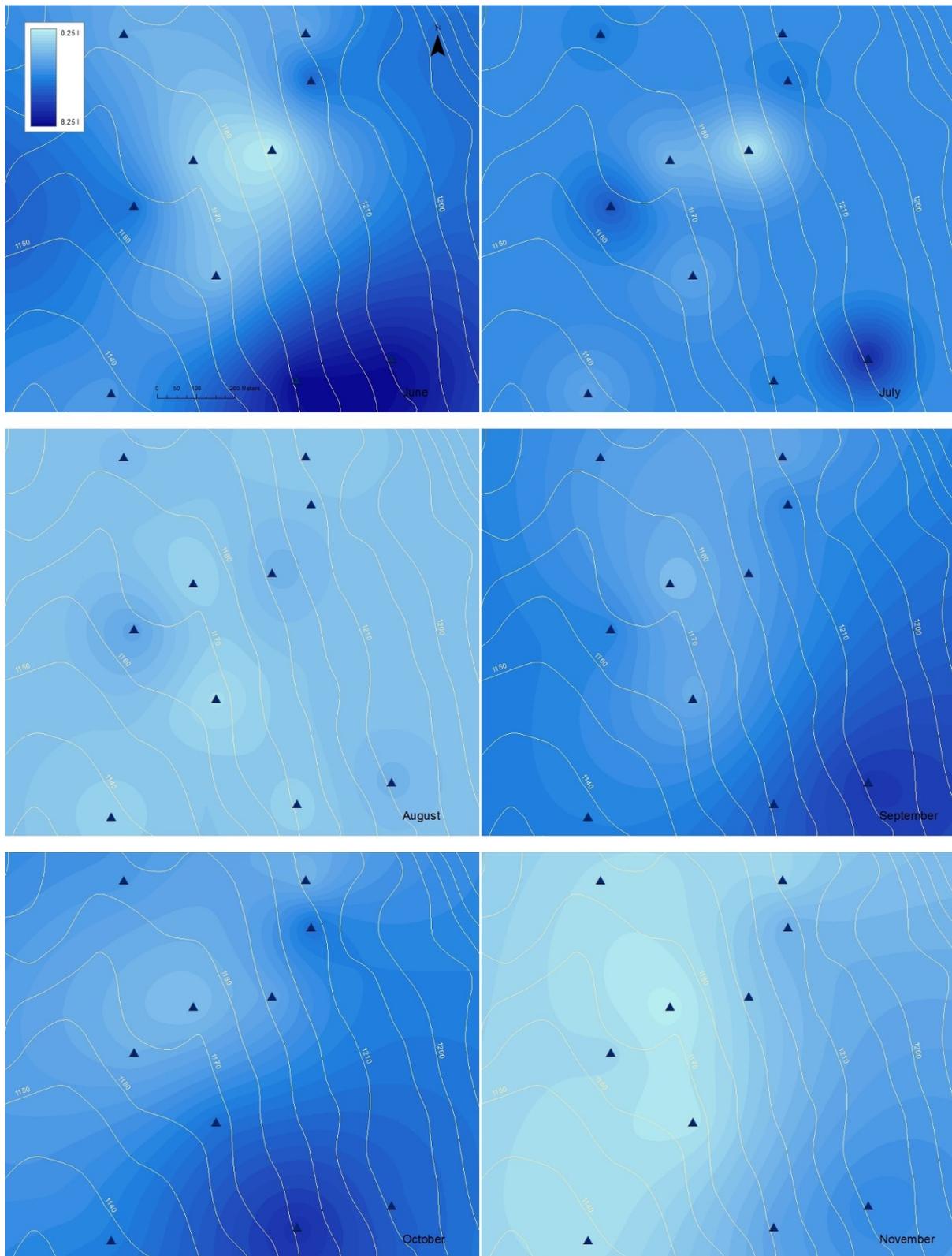


Figure 39: Interpolation of Near-surface fog water variability from June to November 2019. Source: author's elaboration

Five potential topographical predictors of Near-surface fog water were analyzed in Oyarbide exploring global pattern and spatial variability of associations between Near-surface fog water variability and the selected potential topographical parameters based on the results from the processing of ca. 6,800 pictures captured during the flight campaigns November 2018 in the Atacama Desert by a consumer digital SLR Camera DJI Zenmuse X5S Gimbal stabilized mounted on a DJI Matrice 200 UAV/Drone, with a height set to 90 m and draped 1.42 cm/pixel resolution.

The study site's altitude map was generated through interpolation and contouring of the topographical data into a DEM, providing a detailed representation of the altitude gradient with a range of 102 meters. Additional explanatory variables derived from the DEM using ArcGIS spatial analyst tools include Slope, Aspect, Curvature, and Hillshade (see Figure 40).

Approximately half of the study site exhibits slopes of less than  $11^\circ$  (48.7%), with 1% characterized by flat surfaces, indicative of a terrain with moderate slopes. In contrast, only 1.8% of the land displays slopes exceeding 45%. Most of the terrain faces southwest (17.2%) and west (15.3%), with lesser proportions oriented towards the north (6.8%) and northwest (9.6%). Hillshade allows visualization of the terrain based on a light source and the slope and orientation of the elevation surface. Consequently, only 10.3% of the study site experiences shading. Finally, the Curvature of the terrain reveals that 50% of Oyarbide's surface features a convex terrain, while 48.8% is concave, and a mere 1.3% corresponds to flat surfaces.

## 5. Fog variability characterization in the Chilean Atacama Desert

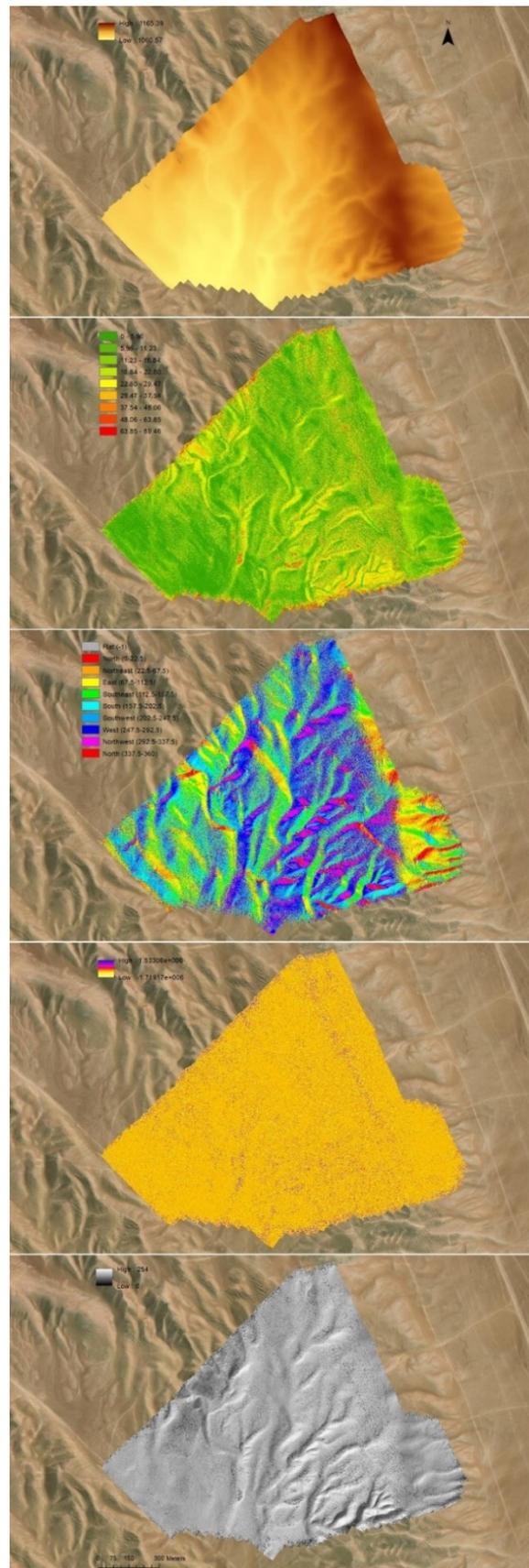


Figure 40: Topographic explanatory variables in Oyarbide extracted from the DEM. Source: author's elaboration

Before applying regression techniques to model Near-surface fog variability, a bivariate Pearson correlation ( $r$ ) analysis was performed on the set of candidate explanatory variables (see Figure 41).

The correlation observed between Near-surface fog water and Elevation is significant, with an  $r$  value of 0.57. This signifies a moderate positive relationship between fog presence and elevation, suggesting that higher elevation areas tend to experience increased occurrences of Near-surface fog. Additionally, significant albeit slightly weaker correlations have been identified between Near-surface fog water and Curvature with an  $r$  of about 0.55. Slightly weaker correlations have been identified between Near-surface fog water and Aspect, Hillshade, and Slope, with coefficients ranging between 0.27 and 0.25.

Highly correlated values were excluded from subsequent analysis to address potential issues arising from multicollinearity whose range use the threshold ranging from 0.6 to 0.8 (Tay, 2017). As shown in the plot, closest correlation to the proposed range was observed for variable Elevation.

5. Fog variability characterization in the Chilean Atacama Desert

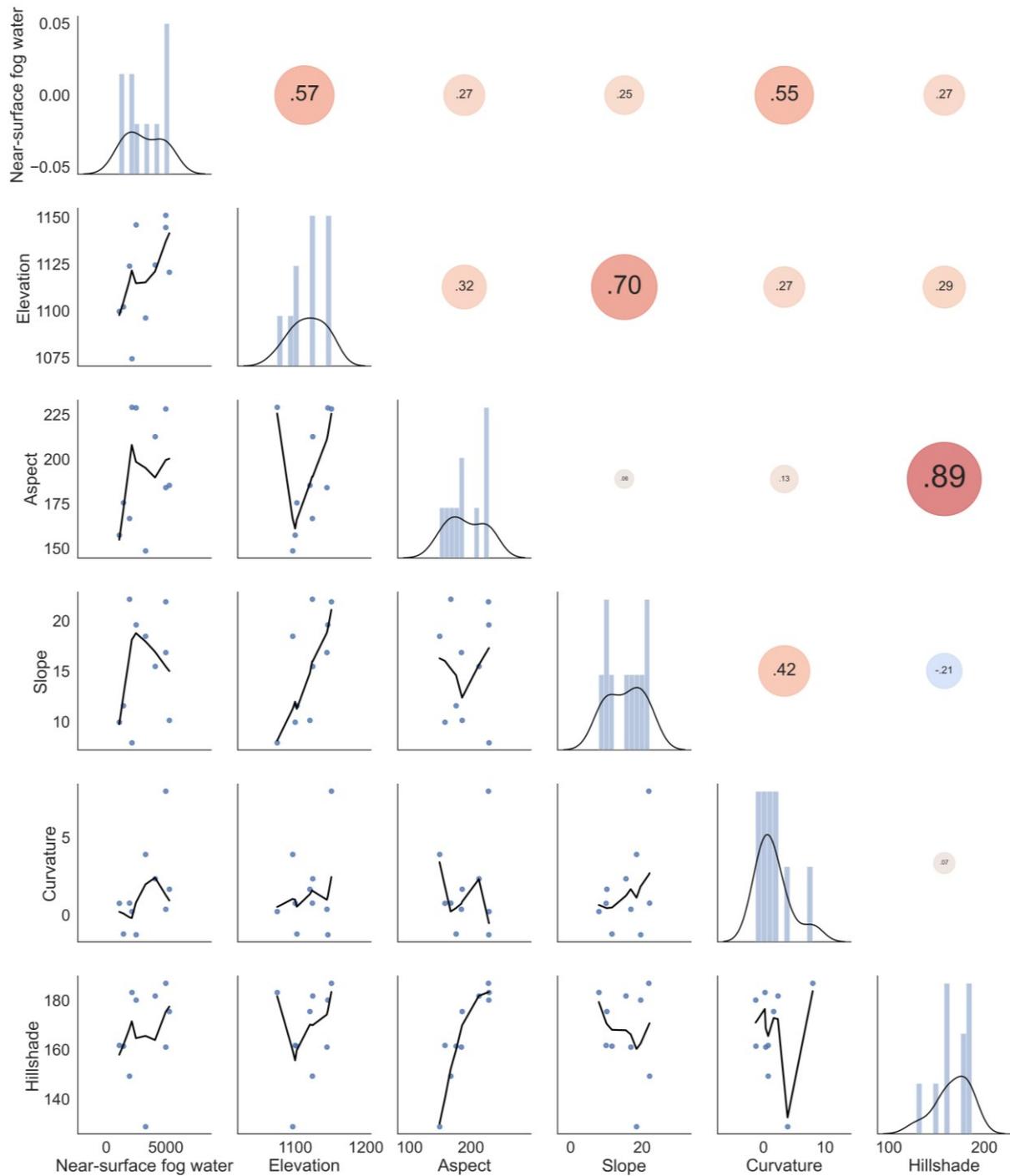


Figure 41: Bivariate correlation between dependent and explanatory variables in Oyarbide. Source: author's elaboration

### 5.3.2. Spatial analysis validation

In the field of climatological research, the intricate relationship between Near-surface fog water and the topographical features of Oyarbide suggests for a sophisticated analytical framework. The "Global Moran's I" and "High/Low clustering – Getis Ord General G" methods, both important tools in spatial analysis, are skillfully employed to assess spatial autocorrelation and spatial clustering, respectively. Implemented prior to engaging in exploratory regression, these methods provide insights into the spatial structure of the data, revealing significant spatial patterns. Through this approach, it becomes possible to identify spatial dynamics prior incorporating spatial variables into the regression model.

#### *Global Moran's I & High/Low clustering – Getis Ord General G*

The use of the Global Moran's I and High/Low clustering – Getis Ord General G indices represents an optimal approach to acquire a comprehensive understanding of the spatial patterns and clustering tendencies exhibited by concentrations of Near-surface fog water concerning topographic attributes. These indices enable the identification of overarching autocorrelation trends and localized clustering patterns. Consequently, they serve as indicators of the topography's influence on the distribution patterns of Near-surface fog water in Oyarbide.

Global Moran's I (Fotheringham et al., 2002) was employed to analyze the variability of Near-surface fog water, aiming to discern whether the pattern exhibited by the dependent variable at the study site is clustered, dispersed, or random. The calculated Moran's I Index show a value of 0.1060 (see Figure 42), indicating a statistically significant spatial autocorrelation. The associated z-score of 0.9403 (regression coefficient divided by standard error) and p-value of 0.3470 (indicating the degree of data compatibility with the null hypothesis) suggest that the null hypothesis cannot be rejected. Consequently, there is insufficient evidence to conclude that the spatial arrangement of Near-surface fog water variability values tends towards randomness. Further analysis is warranted to accurately characterize the spatial pattern of Near-surface fog water distribution.

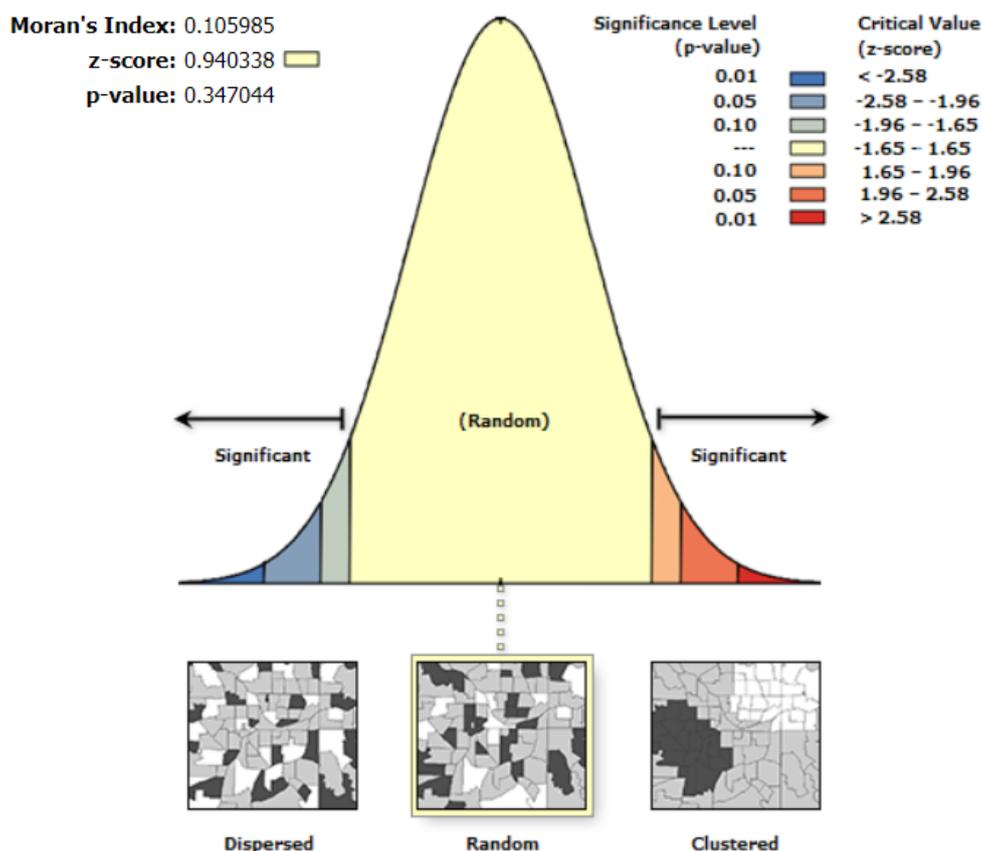


Figure 42: Spatial autocorrelation report of Near-surface fog water collection. Source: author's elaboration

The analysis of explanatory variables revealed statistically significant spatial autocorrelation with a clustered distribution, as indicated by Moran's I (see Table 3). Consequently, rejection of the null hypothesis was supported by both the z-score, and the p-value, signifying that entity values in explanatory variables were non-randomly distributed across the study site.

Variable	Moran's I	Pattern	Z-score*	P-value°
Aspect	0.371336	Clustered	87.814551	0.000000
Slope	0.281953	Clustered	66.685833	0.000000
Curvature	0.001876	Random	0.565123	0.571990
Hillshade	0.059245	Clustered	17.509775	0.000000

\*Measure of standard deviation

°Represents the pseudo-significance level, which is computed using a randomization algorithm (Anselin, 1995). The probability of obtaining an observed pattern  $P < 0.05$

Table 3: Spatial autocorrelation for the explanatory variables. Source: author's elaboration

## 5. Fog variability characterization in the Chilean Atacama Desert

The High-Low Clustering Report for Near-surface fog water collection provides a comprehensive insight into spatial patterns (see Figure 43). The observed General G value of 0.001217 reveals a subtle yet non-significant deviation from spatial randomness. This z-score of 0.148800 and a p-value of 0.881712 suggest that the distribution of the dependent variable does not exhibit significant spatial clustering.

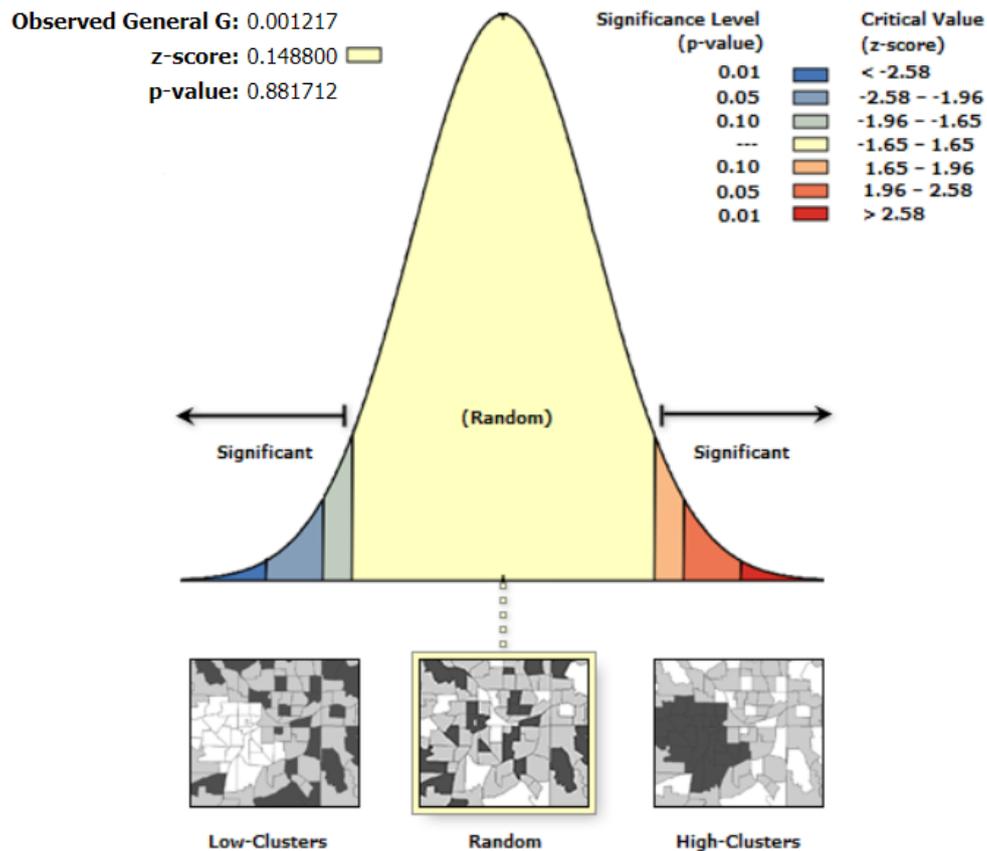


Figure 43: High-low clustering report of Near-surface fog water collection. Source: author's elaboration

### *Exploratory regression*

The exploratory regression approach systematically analyses various potential explanatory variables to identify optimal OLS models. These models effectively elucidate the dependent variable, adhering to local scale-specified criteria.

The provided summaries in Table 4 offer insights into the highest adjusted R-squared results and passing models derived from the exploratory regression analysis and the term "maximum VIF", who denotes the maximum value of variance inflation factor (VIF) serves as a metric indicating the extent to which the variance of a regression coefficient inflates due to multicollinearity among predictor variables. Each summary corresponds to a different combination of predictor variables based as part of a process of exhaustive exploration of different models.

Summary 1: The model with the highest adjusted  $R^2$  (0.02) features the variable Slope as the sole predictor. It demonstrates a low AICc value and no significant Jarque-Bera (JB) or K-Bowman tests (K(BP)) results, indicating a good fit. Additionally, the VIF value of 1.00 suggests no issues with multicollinearity. This model passes all criteria and is denoted by "\*\*\*\*".

Summary 2: Two models yield the highest adjusted  $R^2$  (0.02). Both models include the variable Slope, with additional variables Hillshade and Aspect, respectively. They exhibit low AICc values and no significant JB or K(BP) results, implying adequate fit. The VIF values are below 1.01, indicating no multicollinearity issues. Both models pass all criteria, with one marked as "\*" and the other as "".

Summary 3: The highest adjusted  $R^2$  (0.02) model includes the variables Hillshade, Slope, and Aspect. It features a relatively higher VIF value of 1.05 for Hillshade but remains below the acceptable threshold. This model passes all criteria and includes "\*\*\*\*" for significance.

Summary 4: The model with the highest adjusted  $R^2$  (0.02) incorporates all variables: Hillshade, Slope, Curvature, and Aspect. While Hillshade has a VIF value slightly above 1.05, it remains within an acceptable range. This model passes all criteria and is represented without additional significance markers.

Overall, these summaries highlight the models with the highest explanatory power, predominantly featuring Slope as a significant predictor. So far, multicollinearity concerns appear minimal, and the chosen models adequately fit the data, meeting the specified criteria.

**Choose 1 of 4 Summary**

Highest Adjusted R-Squared Results

<i>AdjR2</i>	<i>AICc</i>	<i>JB</i>	<i>K(BP)</i>	<i>VIF</i>	<i>SA</i>	<i>Model</i>
0.02	464553.83	0.00	0.00	1.00	0.00	Slope***
0.00	465189.25	0.00	0.00	1.00	0.00	Hillshade
0.00	465190.33	0.00	0.00	1.00	0.00	Aspect***

Passing Models

**Choose 2 of 4 Summary**

Highest Adjusted R-Squared Results

<i>AdjR2</i>	<i>AICc</i>	<i>JB</i>	<i>K(BP)</i>	<i>VIF</i>	<i>SA</i>	<i>Model</i>
0.02	464501.29	0.00	0.00	1.01	0.00	Hillshade + Slope***
0.02	464549.78	0.00	0.00	1.01	0.00	Slope*** + Aspect**
0.02	464555.72	0.00	0.00	1.00	0.00	Slope*** + Curvature

Passing Models

**Choose 3 of 4 Summary**

Highest Adjusted R-Squared Results

<i>AdjR2</i>	<i>AICc</i>	<i>JB</i>	<i>K(BP)</i>	<i>VIF</i>	<i>SA</i>	<i>Model</i>
0.02	464502.31	0.00	0.00	1.05	0.00	Hillshade + Slope*** + Aspect
0.02	464503.18	0.00	0.00	1.01	0.00	Hillshade + Slope*** + Curvature
0.02	464551.67	0.00	0.00	1.01	0.00	Slope*** + Curvature + Aspect**

Passing Models

**Choose 4 of 4 Summary**

Highest Adjusted R-Squared Results

<i>AdjR2</i>	<i>AICc</i>	<i>JB</i>	<i>K(BP)</i>	<i>VIF</i>	<i>SA</i>	<i>Model</i>
0.02	464504.20	0.00	0.00	1.05	0.00	Hillshade + Slope*** + Curvature + Aspect

Passing Models

Model variable significance: \* = 0.10, \*\* = 0.05, \*\*\* = 0.01

Table 4: Exploratory regression overview. Source: author's elaboration

The established criterion dictates that the maximum VIF value should remain below 7.50 for the model to be deemed acceptable. The successful outcome of all tests signifies that this criterion has not been breached in any instance, as evidenced by the 100% (% Passed) passing rate (see Table 5). Consequently, it suggests an absence of problematic multicollinearity within the exploratory regression model.

Percentage of Search Criteria Passed					
Search	Criterion	Cutoff	Trials #	Passed	% Passed
Min Adjusted R-Squared	R-Squared	> 0.5	15	0	0.00
Max Coefficient	p-value	< 0.05	15	3	20.00
Max VIF	Value	< 7.5	15	15	100.00
Min Jarque-Bera	p-value	> 0.1	15	0	0.00
Min Spatial Autocorrelatic	p-value	> 0.1	11	0	0.00

Table 5: Exploratory regression multicollinearity testing. Source: author's elaboration

The Table 6 shows a summary of explanatory variable significance and directional impact, within the exploratory regression model. The variable Slope exhibits 100% significance, indicating a strong positive relationship with Near-surface fog water. No negative relationship is observed. The variable Aspect exhibits 75% significance, also shows a predominantly positive relationship with the dependent variable, with no recorded negative relationship. The variable Hillshade does not show any significant relationship with the dependent variable, as indicated by the 0% significance. The variable Curvature, similarly, no significant relationship is found with the dependent variable, with 50% of the instances displaying a positive relationship and the other 50% showing a negative relationship.

Overall, the results presented indicate that variables Slope and Aspect are significant predictors of the dependent variable Near-surface fog water.

<b>Summary of Variable Significance</b>			
<b>Variable</b>	<b>% Significant</b>	<b>% Negative</b>	<b>% Positive</b>
Aspect	75.00	0.00	100.00
Slope	100.00	0.00	100.00
Curvature	0.00	50.00	50.00
Hillshade	0.00	0.00	100.00

Table 6: Exploratory regression assessment of variable significance. Source: author's elaboration

The coefficients from the variables Slope and Aspect consistently demonstrate statistical significance and exhibit positive effects as resilient prognosticators of Near-surface fog water. Conversely, the remaining variables exhibit a lack of statistical significance and display an absence of consistent sign patterns. This renders them inappropriate for inclusion within the models and erodes their efficacy as predictors of Near-surface fog water. Consequently, the definitive course of action will be further analyzed by OLS regression in the next step incorporating the variables Slope and Aspect into the model, thereby ensuring precise and reliable explanatory of the variability in Near-surface fog water.

### 5.3.3. Geo-statistical modelling and analysis

A diverse set of models was iteratively developed by leveraging a selection of candidate variables, ultimately leading to the identification of a model with superior explanatory power. This elected model integrates the quartet of explanatory variables that hold pertinence from the vantage of Near-surface fog water dynamics. The regression analysis encompassed an examination of the variability in Near-surface fog water, treated as the dependent variable, employing two distinct analytical frameworks, the OLS as detailed in the study by Fotheringham et al. (2019), and the GWR method introduced by Fotheringham et al. (2001).

*Ordinary Least Squares (OLS)*

At a 95% confidence level, the OLS model summary (see Table 7) reveals that two of the four explanatory variables, Aspect, and Slope, are significant (p-value) and exhibit expected coefficients. Based on previous results in explanatory regression the remaining variable that are insignificant in the global model therefore are discarded for the local model as well. The VIF values, all below 7.5, indicate no multicollinearity issues among the explanatory variables (Niedzielski & Migala, 2018).

<b>Variable</b>	<b>Coefficient</b>	<b>SE</b>	<b>P-value</b>	<b>VIF</b>
Intercept	3676.607001	1535.108622	0.047663*	–
Aspect	-4.689027	6.987808	0.523583	1.078652
Slope	37.177801	46.922218	0.454102	1.078652

SE: standard error; VIF: variance inflation factor, indicator of redundancy between the explanatory variables, \*P<0.05

Table 7: OLS model summary. Source: author's elaboration

Several key statistical indicators have been employed to assess the efficacy of the global regression model and elucidate the underlying dynamics of Near-surface fog water variability at this analysis level. Each metric shown in Table 8 offers unique insights into different dimensions of the regression model, ranging from its overall fit to assumptions regarding the distribution of residuals and spatial variability.

The negative Adjusted R<sup>2</sup> value, as an indicative of the proportion of variability in the dependent variable explained by the independent variables, suggests a considerable deficiency in the model's ability to elucidate the variance in Near-surface fog water collection based solely on Slope and Aspect.

The AIC serves as a quantitative measure of model parsimony, encapsulating the trade-off between model complexity and goodness of fit. The relatively high value of the AIC value implies potential inadequacies in model specification or the presence of unaccounted factors exerting influence on Near-surface fog water variability.

The Jarque-Bera statistic (Jarque & Bera, 1987) and the Breusch-Pagan (BP) statistic (Koenker, 1981), are a key diagnostic to assess the normality and spatial variability assumptions underlying the regression model. A higher Jarque-Bera statistic, coupled with a p-value exceeding the significance

threshold commonly set at 0.05 implies a lack of evidence to reject the null hypothesis of normality in the residuals. The statistic proximity to unity suggests a reasonable adherence to the normality assumption, indicating that the residuals from the global regression model may indeed follow a normal distribution. This value suggests confidence in the regression analysis and enhances the credibility of the estimated coefficients for Slope and Aspect in explaining Near-surface fog water variability. The BP statistic, with a p-value of 0.055224, slightly exceeds the customary threshold of 0.05 for statistical significance (Koenker, 1981). The high p-value suggests the possibility of non-stationarity in the relationship between the dependent and explanatory variables. This observation suggests nuanced variations across distinct spatial zones within Oyarbide, underscoring the intricate spatial dynamics exerting influence on Near-surface fog water variability.

Statistics	Diagnostic	P
Adjusted R <sup>2</sup> <sup>o</sup>	-0.146039	–
AIC <sup>#</sup>	189.419661	–
Jarque-Bera statistic <sup>§</sup>	0.818878	0.664023
BP statistic <sup>^</sup>	5.792721	0.055224

<sup>o</sup>Coefficient indicating the relative goodness of fit of the regression; <sup>#</sup>measure of the model value of adaption – the lower the value, the higher the model's performance; <sup>§</sup>approach to establish if the residuals of a regression model are normally distributed; <sup>^</sup>approach testing the spatial variability of the variables.

Table 8: OLS diagnostic model. Source: author's elaboration

Spatial regression is justified when there is an enhancement in global fit or when clusters are detected in the residual distribution, necessitating correction. After considering these factors thoroughly and examining the available evidence, it is apparent that opting for GWR is the most supports the adoption of GWR, making it the optimal choice for addressing the research question at hand.

### *Geographical Weighted Regression (GWR)*

The model's local performance was evaluated with various key metrics (see Table 9).

The bandwidth of 6 indicates the spatial extent over which neighboring data points influence the estimation of coefficients. A bandwidth of 6 suggests that the model is considering data points within a relatively proximity.

The RSS (1103474.94625) represents the unexplained variability in the dependent variable after accounting for the explanatory variables. A lower residual compared to the global model, indicates a better fit of the model to the data.

The effective number (9.280045), a measure of the degrees of freedom in the model, considering the spatially varying nature of the relationships between the variables, suggest that each observation contributes as if it were part of a dataset with approximately 9 observations. This implies that the spatial autocorrelation presents in the data results in a reduction of the effective sample size compared to the actual sample size.

Sigma (1238.022649), a standard deviation of the dependent variable, indicating the variability of Near-surface fog water across the study site. A moderate to high sigma value like this one suggests greater variability in Near-surface fog water.

The AICc value of -162.506711 serves as a measure of the model's goodness of fit while penalizing for the number of parameters used. A lower AICc (compared with OLS) suggest a fitting model.

The  $R^2$  indicates that ca. 95.06% of the variance in Near-surface fog water variability can be explained by Slope and Aspect in the model. This suggests a strong overall relationship between the variables.

However, the lower adjusted  $R^2$  suggests that around 38.24% of the variance can be explained by the independent variables after adjusting for the model's degrees of freedom.

<b>Bandwidth</b>	6
<b>Residual squares</b>	1103474.94625
<b>Effective number</b>	9.280045
<b>Sigma</b>	1238.022649
<b>AICc</b>	-162.506711
<b><math>R^2</math></b>	0.950595
<b>Adjusted <math>R^2</math></b>	0.382398

Table 9: Summary of results of the GWR model analysis. Source: author's elaboration

The spatial analysis of localized adaptations, as delineated by the Local  $R^2$  coefficients, offers a nuanced understanding of the model's explanatory power. The range of Local  $R^2$  coefficients spans from 0.20971 to 0.92618 (see Figure 44), reflecting a diverse spectrum of explanatory abilities across different elevations.

## 5. Fog variability characterization in the Chilean Atacama Desert

Areas with maximal  $R^2$  coefficients ( $> 0.6151$ ), representing heightened explanatory prowess, are concentrated within discrete clusters. Specifically, these clusters align along the western north-south Oyarbide strip and in the southeastern highest area of Oyarbide, accounting 50% of registration sensors. Furthermore, there exists a discernible positive correlation between the Local  $R^2$  values and observed Near-surface fog water collection, particularly evident within statistically significant zones. This correlation underscores the heightened influence of the Near-surface fog water variability in response to escalated Aspect and Slope of Oyarbide. Conversely, the model's adequacy diminishes significantly ( $< 0.4304$ ) across 30% of the study site, concentrated in a central patch.

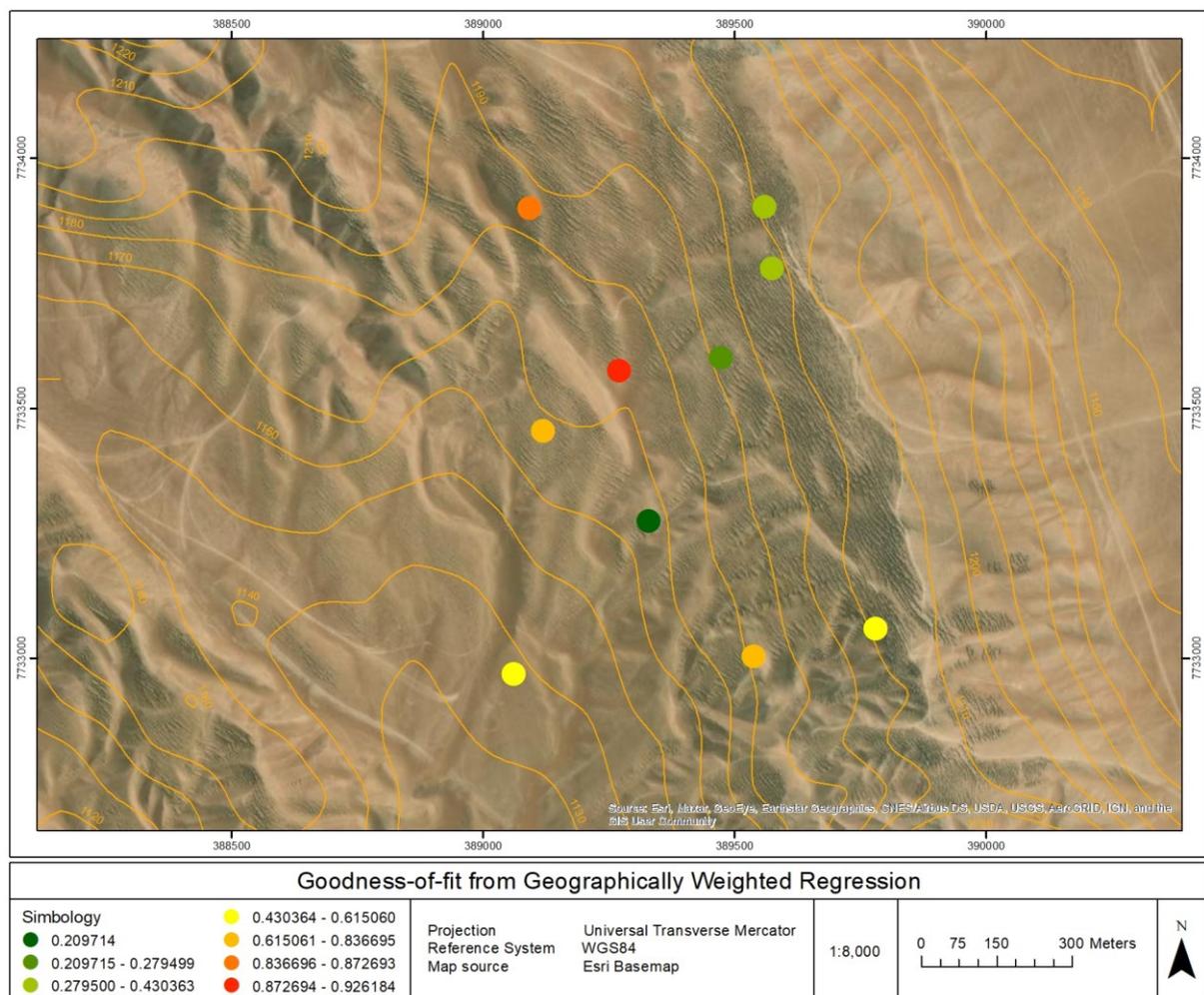


Figure 44: GWR Local  $R^2$ . Source: author's elaboration

The t-values, serving as quantifiers of the disparity between an observed data and its predicted Near-surface fog water data in units of standard error, and rigorously assessed at a confidence level of 95%

is shown for the explanatory variables in Figure 45. It is imperative to note that t-values surpassing the critical threshold of 2.31 are delineated as indicators of pronounced statistical significance.

Mini FCs measurements with high t-values, indicate that those measurements areas have a greater incidence of Aspect and Slope in the Near-surface fog water collection. Conversely, observation with a below threshold t-value suggests a significant negative relationship between explanatory variables and Near-surface fog collection.

The results obtained from the localized modelling offer valuable insights into the discernment of diverse levels of adaptation generating spatialized coefficients for the explanatory variables (elasticities), and providing insight into how these variables interact locally to achieve a specific adjustment in a given area (Fotheringham et al., 2002) (see Figure 46). Higher coefficients of explanatory variables denote a stronger relationship with dependent variable within specific areas, on the other hand, lower coefficients suggest a weaker association.

## 5. Fog variability characterization in the Chilean Atacama Desert

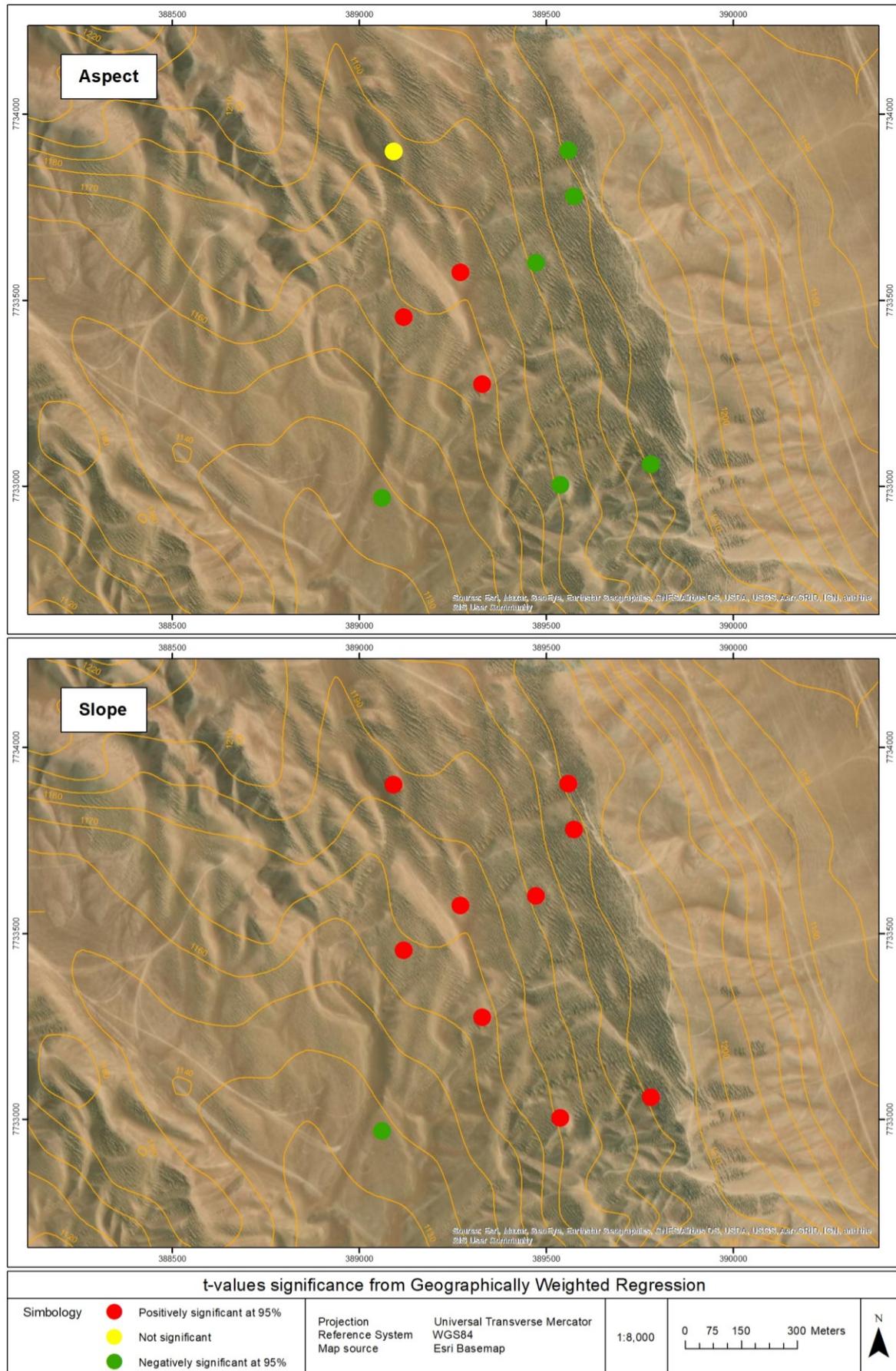


Figure 45: GWR t-value for each explanatory variable. Source: author's elaboration

## 5. Fog variability characterization in the Chilean Atacama Desert

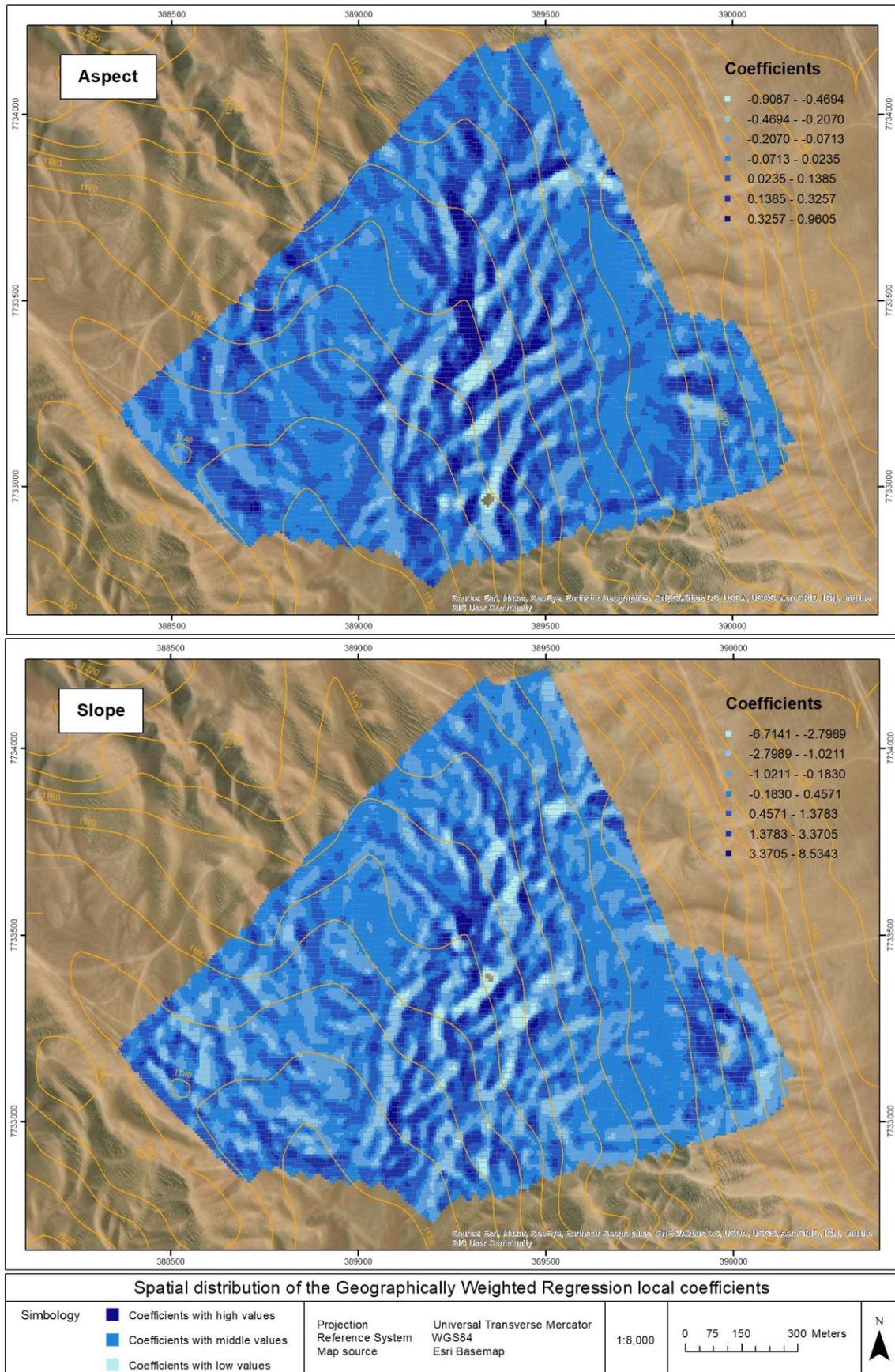


Figure 46: Local influence of explanatory variables on Near-surface fog water variability. Source: author's elaboration

Within the higher coefficients recorded, the Aspect intervals coefficients suggest a weak positive correlation between Aspect and observed Near-surface fog water. As the terrain Aspect tends towards facing towards the sun, observed Near-surface fog water shows a slight increase. The average observed Near-surface fog water (3145.93 ml) closely resembles the predicted Near-surface fog water (3145.85 ml), indicating local model accuracy. The average Slope ( $11.23^\circ$ ) and an average Aspect of  $176.58^\circ$  (SSE), may suggests moderated terrain inclinations, potentially influencing fog accumulation.

At the high Slope intervals coefficients, reflects a strong positive correlation between Slope and observed Near-surface fog water. Steeper slopes correlate with increased observed Near-surface fog collection. The average observed Near-surface fog water (3074.02 ml) aligns closely with predicted Near-surface fog water (3074.21 ml), indicating model reliability. The same lower average Slope and Aspect suggests steeper terrain, potentially facilitating fog collection. At both explanatory variable areas with high coefficients the lower average elevation (1,100 masl) in conjunction with precedent features mentioned may implies terrain conducive to localized Near-surface fog collection.

Conversely, lower coefficients in some areas of Oyarbide indicate a less pronounced relationship between the explanatory variables and the dependent variable.

Within the lower coefficients, the Aspect intervals coefficients indicate a moderate negative correlation between Aspect and observed Near-surface fog water. As the terrain Aspect becomes more facing away from the sun, observed Near-surface fog water tends to increase. The average observed (3279.58 ml) and predicted (3278.75 ml) Near-surface fog water aligns closely, suggesting accurate model predictions. The average Slope angle ( $13.14^\circ$ ) implies moderately inclined terrain, and an average Aspect of  $196.41^\circ$  (SSW), might contribute to moisture dissipation. The Slope intervals coefficients indicate a moderate negative correlation between Slope and observed Near-surface fog water. As slope steepness increases, observed Near-surface fog water tends to decrease. Notably, the average observed (3120.77 ml) and predicted (3120.72 ml) Near-surface fog aligns closely, suggesting reliable model performance also for this variable. The average Slope ( $12.46^\circ$ ) and same average Aspect implies potentially facilitating lower levels of Near-surface fog water collection. Additionally, the average low elevation (1,104 masl) complements explanatory variables characteristics.

In summary, the GWR modelling with its statistics associated offers a superior modelling of the interplay between the dependent variable and its two explanatory variables by delineating nuanced local fluctuations in this local relationship. By discerning and assigning optimal model weights tailored to each geographic location throughout Oyarbide, it facilitates the creation of a proper spatial depiction of these weights, thus enhancing the understanding of the spatial dynamics underlying the variables involved.

## 5. Fog variability characterization in the Chilean Atacama Desert

Figure 47 provides a visual representation of the synthesized impact of Slope and Aspect on Near-surface fog variability. Through the normalization and weighting of the values, these coefficients have been merged to yield a single, representative metric elucidating their collective influence on fog variability across the study site. Low fluctuations in Slope and Aspect values are correlated with lower trends in Near-surface fog water variability. Conversely, when the normalized Slope and Aspect values are higher, the Near-surface fog water variability tends to be higher.

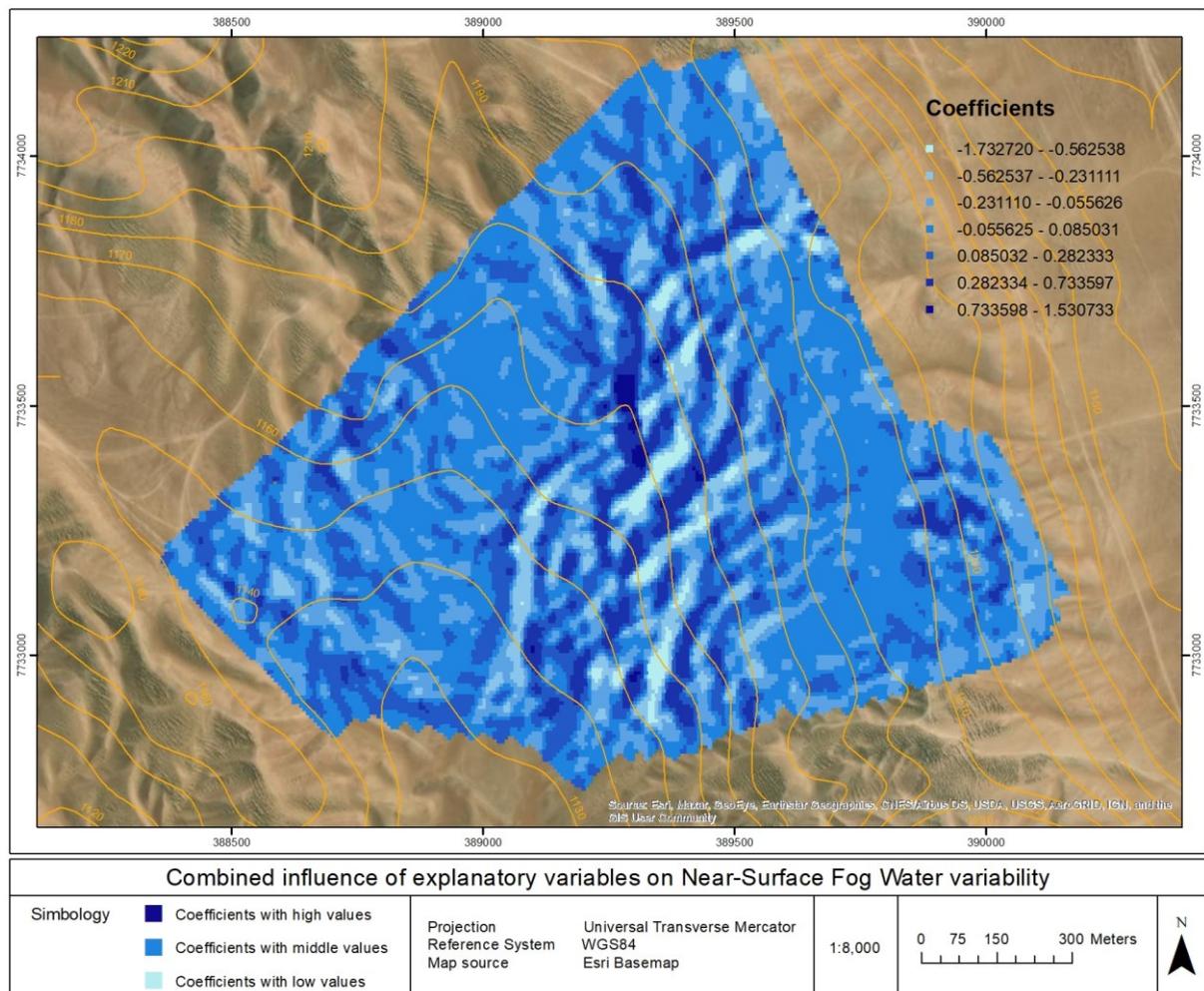


Figure 47: Coefficients of combined explanatory variables and their relationship with Near-surface fog water variability. Source: author's elaboration

Although Elevation was excluded from both the global and local models due to multicollinearity, its relationship with predicted Near-surface fog water values remains significant as an independent explanatory variable. Analyzing this association outside the constraints of the regression models offers a complementary perspective on altitudinal effects. The Figure 48 illustrates the relationship between

Elevation and the predicted Near-surface fog water values, highlighting trends that align with the altitudinal tiers initially discussed.

This complementary assessment provides further evidence of the altitudinal influence on fog water collection, yet underscores the importance of using multivariate models, such as local, to capture the spatial heterogeneity of the interactions between the dependent variable and other topographical predictors. Consequently, the comparative analysis between global and local regression models sheds further light on the efficacy of each approach in accounting for the variability of Near-surface fog water across Oyarbide.

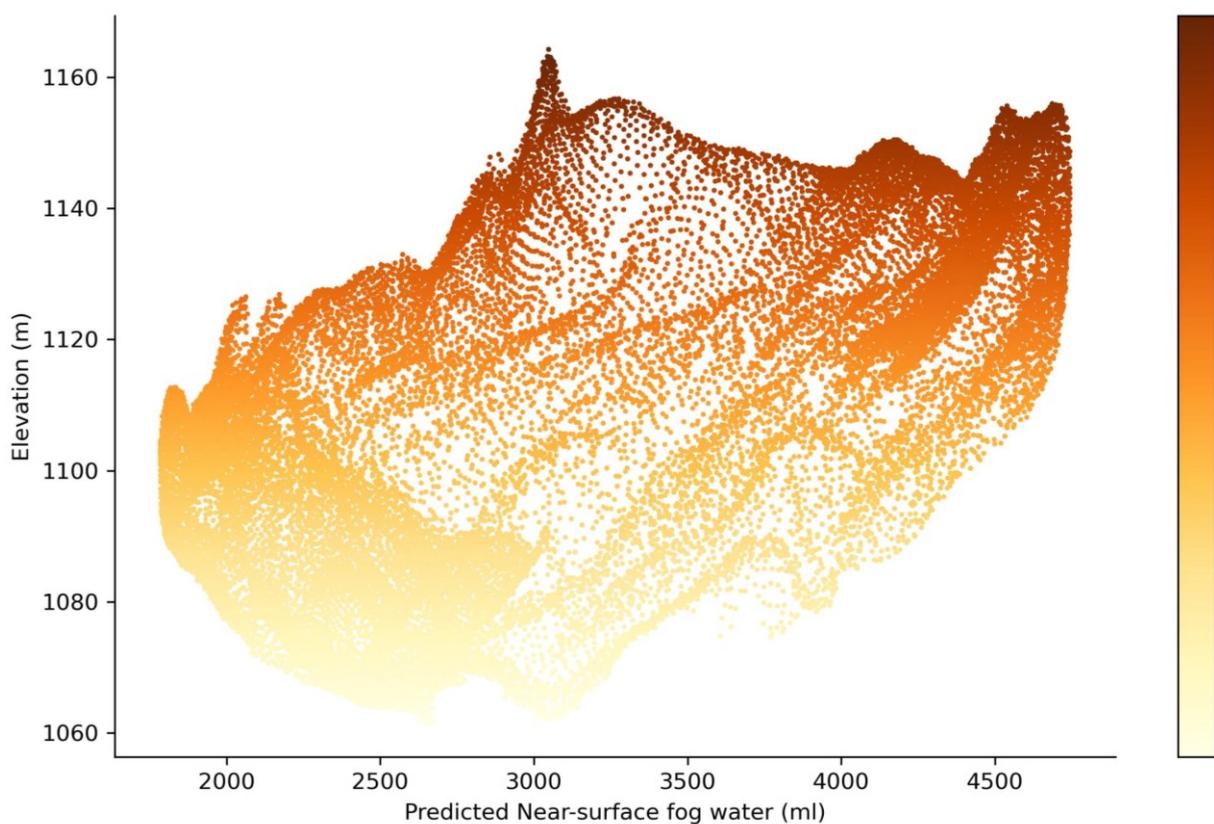


Figure 48: Relationship between Elevation and predicted Near-surface fog water. Source: author's elaboration

*Global (OLS) versus Local (GWR) regression model*

The comparison between these regression models shows the efficacy of each approach in capturing the spatial variability of Near-surface fog water, with widely observed differences in key performance metrics in each one (see Table 10).

<b>Value</b>	<b>OLS</b>	<b>GWR</b>
Residual sum of square	22706518.768	1103474.946
Classic AIC	189.419661	-162.506711
R <sup>2</sup>	0.108636	0.950595
Adjusted R <sup>2</sup>	-0.146039	0.382398

Table 10: Comparison between global and local regression modelling. Source: author's elaboration

The RSS, a key measure of model fit, demonstrates considerable variation between OLS and GWR, with GWR exhibiting substantially lower values. This suggests that GWR better accounts for the localized variability in Near-surface fog water, reflecting its ability to capture spatial heterogeneity more effectively compared to the global modelling.

Furthermore, the AIC underscores the advantages of local modelling, as evidenced by its negative value, indicating a higher level of explanatory power compared to global model. This suggests that GWR offers a better fit for the local data (Hurvich et al., 1998; Niedzielski & Migąła, 2018) and therefore a more efficient framework for modelling Near-surface fog water variability in Oyarbide.

The R<sup>2</sup> further reinforce the superiority of GWR, showing its ability to explain a significant proportion of the spatial variance in Near-surface fog water occurrence. This contrasts with the comparatively lower explanatory capacity of the OLS. Despite the inherent complexity of GWR due to its localized nature, the higher Adjusted R<sup>2</sup> values, providing insight into the balance between model complexity and explanatory power, indicate that GWR achieves a favorable balance, offering a more nuanced understanding of the relationship between Aspect and Slope, and Near-surface fog water variability in the study site.

Assessment of the spatial autocorrelation of residuals between OLS and GWR, allow to assess how each model addresses and captures the spatial structure of prediction errors, thereby aiding in

determining which model is more suitable for describing and predicting Near-surface fog water variability in Oyarbide (see Table 11).

<b>Score</b>	<b>OLS</b>	<b>GWR</b>
Moran's I	0.209222	-0.204591
Expected index	-0.111111	-0.111111
Variance	0.052469	0.044919
z-score	1.398458	-0.441061
p-value	0.161976	0.659169
Pattern	Random	Random

Table 11: Moran's I in the residuals OLS and GWR. Source: author's elaboration

In the OLS model, a significant positive Moran's Index in residuals suggests pronounced clustering, indicating that areas with similar Slope and Aspect tend to exhibit analogous Near-surface fog water collection. Conversely, the GWR model reveals a significant negative spatial autocorrelation, implying spatial heterogeneity, where neighboring areas manifest dissimilar Near-surface fog water variability despite comparable Slope and Aspect characteristics of the terrain.

Both models exhibit equal expected index values, indicating that the observed spatial autocorrelation aligns with what would be expected under spatial randomness. The OLS model shows a slightly higher variance of residuals compared to the GWR model, suggesting greater variability around the mean prediction in the OLS framework.

The OLS model presents a positive z-score and relatively low p-value, affirming significant positive spatial autocorrelation in global approach. In contrast, the GWR model displays a negative z-score and higher p-value, indicating a lack of significant spatial autocorrelation.

Both models show a random spatial pattern, indicating that while there is a spatial association between explanatory and dependent variable, the distribution lacks discernible spatial trends beyond random variation.

## 6. Discussions of multiscale findings on fog variability in the Chilean Atacama Desert

The quality and calibration of instruments used to measure climatic variables may vary across climatological stations, potentially leading to less precise measurements and, consequently, can result in an increased occurrence of outliers at specific climatological stations. Therefore, instrumentation stands as a pivotal requisite in procuring dependable climate data.

Several researchers advocate for the utilization of non-geostatistical methods (Borges et al., 2016; Goovaerts, 2000; Mair & Fares, 2010), which disregard spatial dependence in spatial data dynamics, to interpolate missing data or outlier series in non-continuous variables, typically employing linear regressions and inverse distance weighting, alongside geostatistical methods such as kriging. Nonetheless, meticulous attention must be accorded to selecting and applying statistically robust methods tailored to the dataset's temporal interval when addressing gaps or outliers.

The data quality flow proposed for the fog climate measurements network in Atacama, although yielding favorable outcomes, requires acknowledging that its application may not consistently depict actual variability accurately in climatological data. This is especially true when temporal trends or substantial differences between data intervals exist. While averaging and correlation techniques by different data intervals may offer practical adjustments, they might fall short in providing a comprehensive evaluation of the data's variability. Moreover, employing these techniques to estimate missing data or outliers can significantly impact the overall dataset, this is particularly relevant in situations where the data is highly susceptible to the influence of outliers or temporal trends.

Additionally, as mentioned by Schween et al. (2020), the Atacama Desert suffers from a lack of continuous and representative meteorological observations covering the entire region. Long-term data is only available from meteorological stations located at the airports of Antofagasta, Iquique, Arica, and Calama. These data are often extrapolated to other areas, which poses a challenge due to the localized nature of low cloud and fog distribution (Del Río et al., 2021). Covering this gap, the detailed regional-local fog climate network data and its improvement process described in this research serves as a contribution to address this issue, characterizing the variability of fog climate with high spatiotemporal resolution and its relationship with climatic gradients.

After passing the quality flow, the analysis and evaluation of the climatic data have delivered interesting results recorded in the wind patterns with foggy conditions inland desert (see Figure 49).

The comparative analysis between fog collection and without collection reveals distinctive patterns in the distribution of wind speeds across different wind direction intervals. During fog collection, wind direction patterns with the highest average speeds are centered around northeasterly to northerly directions, notably ENE (2.66 m/s), NE (3.29 m/s), NNE (3.33 m/s), and N (2.40 m/s). These plots suggest a tendency towards more efficient fog collection when the wind blows from these directions. Conversely, wind direction patterns with lower average speeds during fog collection include SSE (0.49 m/s), S (0.85 m/s), SE (0.81 m/s), and W (1.27 m/s), suggesting a potential contribution, but not a major one, to the persistence of fog collection when the wind originates from southerly and westerly directions, where speeds are notably lower.

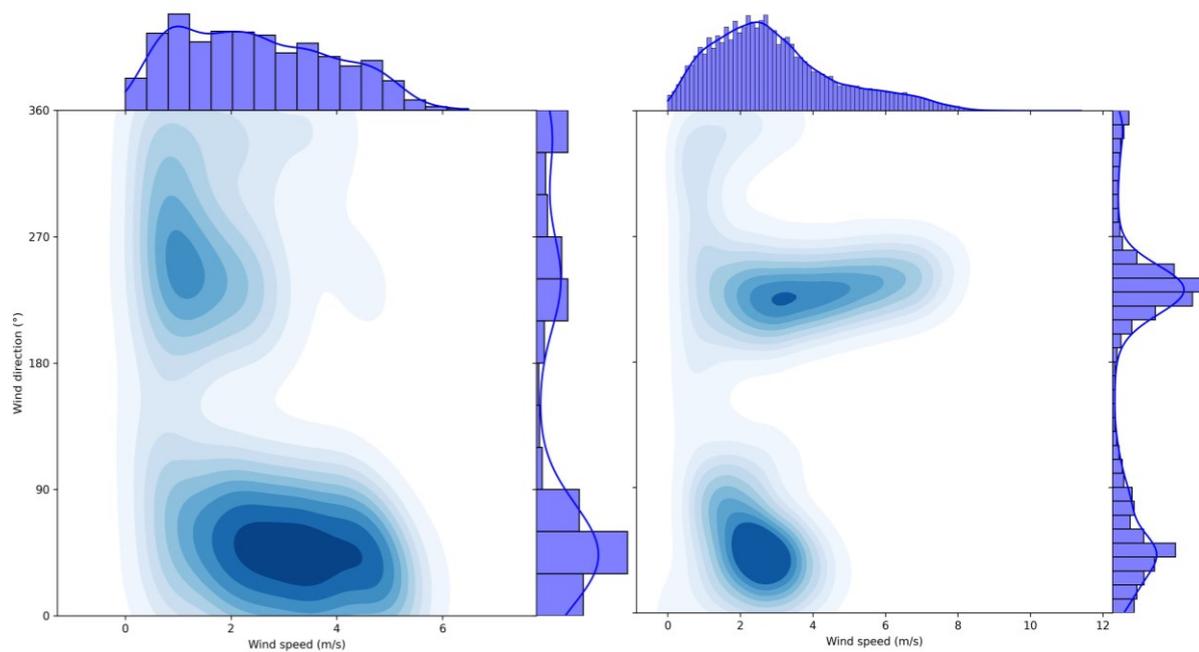


Figure 49: Wind patterns with (left) and without (right) fog collection inland desert. Source: author's elaboration

The findings align with broader atmospheric circulation patterns observed in the region. For instance, the presence of higher wind speeds from northeasterly to northerly directions during fog collection may be influenced by the interaction of local topography with regional atmospheric dynamics, such as the Southeast Pacific High (SEPH) (García et al., 2021). The SEPH, characterized by its southerly flow, contributes to the development of land-sea breezes and other mesoscale phenomena, ultimately impacting the distribution of wind speeds across different wind direction patterns.

Additionally, radiative heating and cooling at the slopes of the Andes and the Altiplano to the west initiate a circulation so called Rutllant cell (Rutllant et al., 2003). This circulation pattern leads to strong westerly and easterly winds during daytime and nighttime, respectively (Schween et al., 2020). The Rutllant cell further contributes to the complexity of wind behavior observed at station OYA\_1211, particularly during different times of the day and night.

In addition, as details García et al. (2021) the wind patterns observed at inland desert measurements exhibit a regular pattern characterized by eastward winds with lower speeds during the second half of the night until hours before noon, followed by westward winds at higher speeds from before noon until the first half of the night. This pattern is influenced by the presence of the Andes and the central depression, which dominate the inland wind system. During the night, wind speeds are low, and direction is mostly from the land, roughly following the slope of the cliff. However, some hours after sunrise, the wind direction switches to southwest, and speed increases, indicating the influence of a land-sea breeze combined with a coastal low-level jet (Garreaud et al., 2008; Soares et al., 2022).

Conversely, without fog collection, wind speeds exhibit a distinct distribution. Wind direction patterns with the highest average speeds are SW (3.89 m/s), WSW (4.37 m/s), and NE (2.59 m/s). Conversely, wind pattern intervals with the lowest average speeds are SSE (1.10 m/s), S (1.42 m/s), and SE (1.54 m/s), suggesting that winds from southeast and south directions tend to have lower speeds when fog is absent.

Such variability in fog distribution and fog water content are expected to happen with climate change at different elevations at the Coastal Cordillera (Del Río et al., 2018). The influence of the ENSO changes, since the beginning of the 21<sup>st</sup> century, presents Interdecadal Pacific Oscillation (IPO) predominantly negative values, suggesting a shift toward the negative (cold) phase. If this evolution is confirmed during the forthcoming years, even drier conditions, will be likely in the arid coast of northern Chile during coming decades, potentially affecting fog episodes (Del Río, 2019; Schulz et al., 2012).

Overall, the comprehensive understanding of wind patterns, incorporating both local topographical effects and broader atmospheric circulation dynamics, provides valuable insights into the intricate relationship between fog collection, regional atmospheric circulation, and local climatic and topographical conditions in the inland desert. Further research integrating these findings with local and broader biosphere-atmosphere dynamics will improve the understanding of the complex fog climate interactions patterns in the region.

The utilization of local spatial statistics to model spatial data, exemplified in this research by the GWR as originally proposed by Fotheringham et al. (2002), has gained widespread recognition among spatial analysts. This surge in popularity is driven by the inability of global statistical techniques to effectively address contemporary scientific needs. The GWR offers a versatile method for examining spatial relationships between dependent and explanatory variables within a unified framework. By seamlessly integrating data exploration, analysis, evaluation, visualization, and mapping capabilities within GIS providing comprehensive output, enabling spatial autocorrelation analysis (Pratt & Chang, 2012), thus, the GWR facilitates a comprehensive analysis of local variations across spatial domains (Brunsdon et al., 1998; Fotheringham et al., 2002).

Traditional regression analysis employs a method that minimizes the squared differences from the fitted line and yields a value for the entire set of observations, assuming that the relationship between variables is consistent throughout the study site (stationary), without considering the possibility of local variations due to the heterogeneity of the space, thus adopting a global perspective. However, this assumption has been recognized as a fundamental weakness by statistical geoscientists, therefore this approach is inadequate when the model relationships vary across the research area (Gutiérrez et al., 2013).

The choice between global and local regression analysis depends on the clarity of the local characteristics of variables. When these characteristics are not readily evident, the global analysis employing fixed coefficients is used, masking the geographical variations of the relations between the variables (Lloyd & Shuttleworth, 2005). Conversely, if the coefficients exhibit spatial heterogeneity and vary depending on the location, as it was proven in Chapter 6, the local analysis is utilized (Fotheringham et al., 2001), assuming that this relationship between geographical data do not have a consistent behavior stationary (Haining, 1977).

The GWR enables the estimation of model parameters and statistics at each sample location, thereby allowing the description of patterns and relationships between variables in a manner that was unobservable with OLS (Fotheringham et al., 2002; Fotheringham et al., 1998). Furthermore, parameter estimation in GWR heavily relies on the weighting function and bandwidth of the kernel used, and selecting appropriate values for these factors is crucial for obtaining desired local variation (Wang et al., 2005). However, as the bandwidth increases, parameter estimates in GWR tend to converge towards those of a global model (Wang & Tenhunen, 2005).

The GWR offers several advantages over traditional regression models. Firstly, it attenuates abrupt changes in local statistics across neighboring data, thereby enhancing the stability of estimations. Secondly, it facilitates the visualization of spatial variability within geographic entities, aiding in the

identification of localized trends and patterns. Finally, GWR enables the analysis of regionally aggregated data, providing insights into broader spatial trends and relationships (Goovaerts, 1998).

Three fundamental principles support GWR, as highlighted by Hanham and Spiker (2005). Firstly, spatial data often lack stationarity, necessitating localized modelling approaches like GWR. Secondly, the spatial structure of data significantly influences the estimation of variable relationships. Thirdly, relationships between variables can exhibit localization and variability across space. GWR addresses these principles by calibrating multiple regressions at each sampled point, capturing localized spatial variation effectively (Zhang et al., 2004). The regression fitting in GWR employs a spatial proximity approach, giving greater weight to observations nearer to the estimation location (Foody, 2003; Fotheringham et al., 2002).

From a bioecological standpoint, local spatial models, such as GWR, play a crucial role in capturing the inherent spatial structuring of natural systems (Legendre, 1993; Legendre, P., & Legendre, L., 1998). By considering spatial heterogeneity and local variation, GWR enhances the understanding of ecological processes and spatial dynamics. This approach provides evidence of the importance of considering local analysis in atmosphere-biosphere interaction phenomena, especially considering that one of the most extensive *Tillandsia* fields in northern Chile is located in this region, whose nutrient uptake comes only from the fog, which is favored by specific characteristics of the terrain (Koch et al., 2019).

In modelling the variability of Near-surface fog, the predictive prowess of this local model is shaped not solely by the robustness of associations between fog collection and topographical features of the study site, but also by the model's proficiency in discerning whether these associations manifest across multiple spatial scales. The GWR offers a structured approach to investigate scale-specific effects by progressively expanding the scope of the local analysis data sample and evaluating their repercussions on the model's predictive performance (Miller, 2012).

When validating regression models, a crucial step involves assessing spatial autocorrelation patterns within residuals. This practice aims to ascertain whether the residuals exhibit any spatial clustering, would indicating potential complete or incomplete capture of the underlying spatial structure within the data (Pratt & Chang, 2012). With a Moran's Index of -0.20 in the residual GWR diagnostic (representing the difference between observed and predicted Near-surface fog values), the model suggests negative spatial autocorrelation in its residuals (see Figure 50). This indicates that high residuals tend to be surrounded by low values, and vice versa, concerning the Near-surface fog water variable. This could indicate spatial patterns in fog distribution not entirely captured by the explanatory variables included in the model, suggesting that other spatial variables not included in the

model could explain part of the variability in Near-surface fog water. It could also indicate that the GWR model may systematically underestimate or overestimate fog presence in certain geographic areas.

In the northeastern Oyarbide measurement areas, specifically at GFC\_1206 and GFC\_1208, the high values of the positive standardized residuals suggest that the model is tending to overestimate Near-surface fog water levels in those areas. On the other hand, in the mostly rest of the measurement areas the standardized residual values are slightly negative, which indicates a slight tendency to underestimate the Near-surface fog water levels in that area. This suggests that the model could be better tuned to capture the characteristics of these specific topographic conditions.

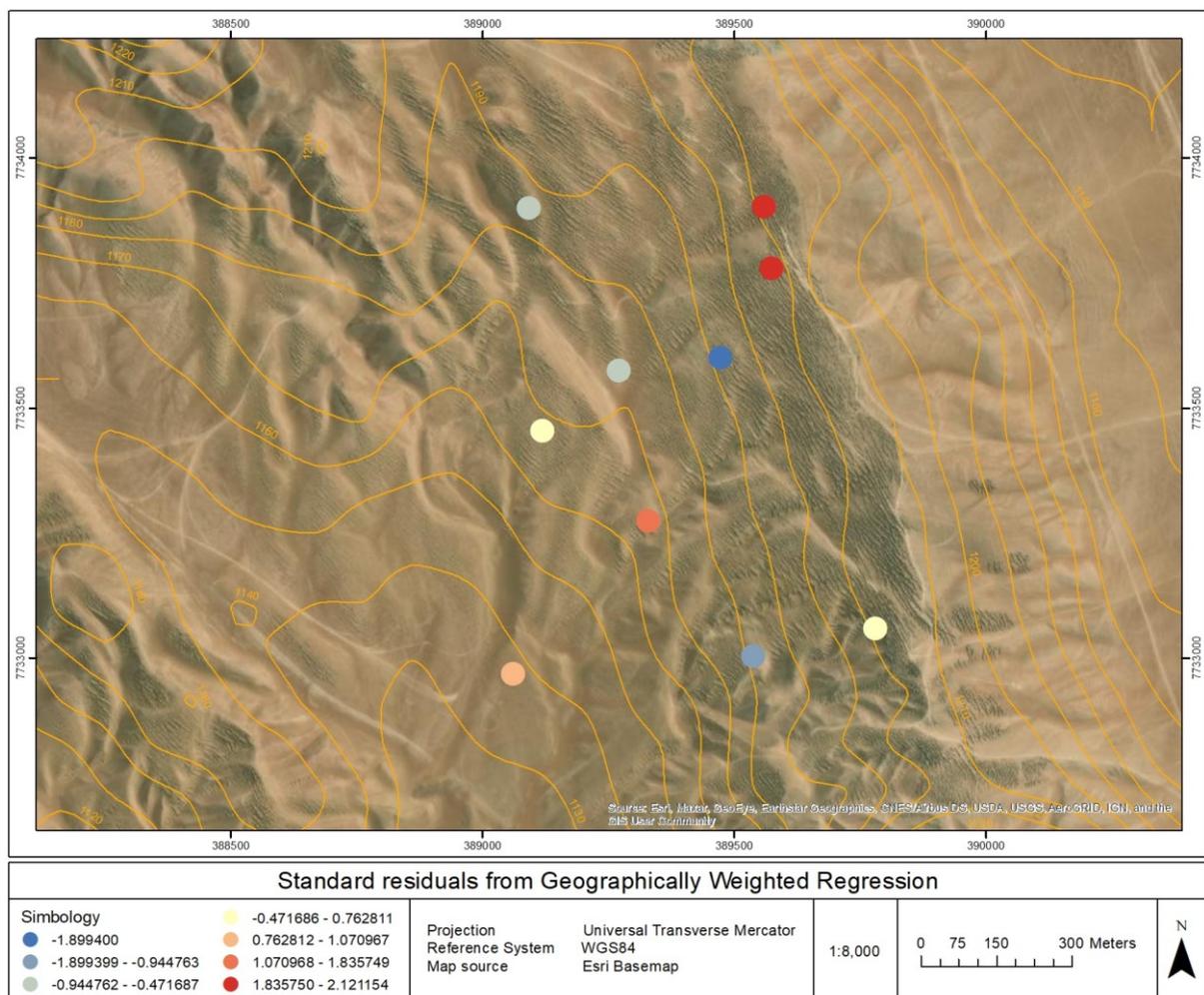


Figure 50: GWR standard residual values distribution. Source: author's elaboration

Spatial analysis enables researchers to determine whether variables are random or tend to cluster (Getis & Ord, 1992), and it helps in understanding the underlying reasons for such spatial distributions

(Fotheringham et al., 2015). On regional bio-cycles, topography plays a crucial role in controlling microclimate and exerts a significant influence on its dynamics (Chen et al., 2007; Hook & Burke, 2000). In the context of fog spatial variability, local geographical features may play a more prominent role than altitude (Błaś et al., 2002).

While GWR offers several advantages in identifying spatial heterogeneity and providing a more detailed and precise analysis, getting heightened levels of granularity and precision (Foody, 2003; Fotheringham et al., 2002; Lloyd & Shuttleworth, 2005), it remains pivotal to take into account the potential reliability conflicts (Páez et al., 2011). Therefore, it is imperative to note that all regression analysis and modelling, by their inherent nature, are essentially non-spatial. Thus, caution must be exercised when deploying these methodologies to analyze spatial data, ensuring meticulous attention is given to the pertinent assumptions and constraints.

Several limitations and challenges have been identified by researchers in the local approach of GWR. Zhang and Shi (2004) suggest that the estimation of local model parameters in GWR relies on spatial weighted proximity, which may lead to a reduced effective number of observations for certain spatial points. Additionally, outliers can significantly impact the estimation of localized model parameters. Moreover, while GWR effectively detects spatial non-stationarity in relationships, interpreting such non-stationarity demands additional contextual and underlying information (Yu & Wu, 2004). Moreover, the interpretation of spatial variation coefficients produced by GWR is not straightforward and often regionally specific, potentially limiting its relevance for specific areas of interest (Platt, 2004).

Furthermore, GWR has been noted to have limitations concerning interpretation and extrapolation. While regression models and diagnostic statistics in GWR can be interpreted similarly to traditional global OLS, local  $R^2$  should not be interpreted with the same confidence as global  $R^2$  (Foody, 2003). The GWR is also not suitable for extrapolating relationships beyond the region where the model was developed. Furthermore, selecting the appropriate kernel and bandwidth for a specific application poses challenges (Foody, 2003). Lastly, as cautioned by Jetz et al. (2005), it's important to acknowledge that GWR models do not fully address spatial autocorrelation and must be interpreted with caution. Additionally, general inferences like those offered by global OLS models cannot be obtained through local GWR models.

In summary, the limitations of GWR encompass issues related to parameter estimation, interpretation, extrapolation, and addressing of the spatial data, which collectively warrant careful consideration and interpretation when utilizing this approach in spatial analysis.

## 7. Conclusions and outlooks in understanding fog climate patterns

This research provides a multifaceted perspective on fog climate dynamics in the Atacama Desert, emphasizing the importance of data quality assurance, regional analysis, and localized investigations.

The analysis of anomalies through the fog climate measurements network underscores the criticality of data integrity in climatic studies. The varying durations and frequencies of missing data and outliers' instances across stations needs meticulous attention when utilizing these datasets for scientific and analytical purposes and climatological modelling. The data quality flow enhances the quality and accuracy of the climatological data housed within the central repository. However, it's important to highlight that the methods outlined in this quality control and data management cannot restore erroneous or missing data once it's lost. What was bad or lost, is irretrievable. Nevertheless, these methods facilitate the substitution of incorrect values and the filling of missing climate data with "reasonable" values that align with the statistical and climatological attributes of the location of every climate station.

The sporadic occurrence of these missing data or outlier intervals, spanning from minutes to months across different climatological stations, suggests several possible causes, such as equipment malfunctions, technical glitches, maintenance periods, or even environmental factors impacting satellite/remote data transmission or recording.

This comprehensive analysis emphasizes the heterogeneous nature of data integrity across whole fog climate measurement network. Further investigations into these discrepancies are crucial to ensure the accuracy and completeness of climatological data, facilitating comprehensive and accurate climatic analyses. Subsequent research endeavors can employ these methodologies to improve the management, application, and analysis of future data sets.

At regional scale, the research highlights notable seasonal variability and geographical influences on fog water collection from coast to inland desert. March and April exhibit the highest increases in fog accumulation, particularly in the interior desert areas. Conversely, May and June experience lower accumulation levels and slower growth.

The highest accumulation of fog water is observed in May along the coast, while higher altitudes record peak accumulation in October. Coastal areas are characterized by relatively stable atmospheric conditions, leading to a more concentrated collection of fog water. Inland desert regions, however, experience fluctuating fog patterns influenced by seasonal and local climatic factors. The

concentration of fog water tends to be lower but displays sporadic instances of significantly higher amounts.

The diurnal variability of fog occurrence, particularly prevalent during the early morning hours preceding noon, is intricately linked to the complex desert-ocean interactions influencing temperature and humidity gradients. Subsequently, prevailing ambient conditions attenuate fog formation, thereby resulting in diminished fog water collection. Consequently, cooler early mornings engender increased fog water collection, thereby exerting discernible trends in fog water collection records.

Hourly fog variability indicates distinct seasonal patterns, with peak occurrence during the Austral Winter. Fog formation is more prevalent during the early morning hours, gradually dissipating as the day progresses. Notably, fog frequency exhibits an inverse correlation with diurnal temperature fluctuations.

Wind patterns play a crucial role in fog events, with prevailing directions affecting fog occurrence and consequently, collection records. Northern winds significantly increase fog occurrence, while fog-free conditions correlate with less pronounced wind patterns.

The records of dew points reveal a multifaceted interplay of atmospheric dynamics, prominently shaped by oceanic factors and the variability of the TIL. As the TIL ascends, there is a discernible increase in dew point measurements attributable to heightened oceanic influences. This ascent of the TIL corresponds with augmented atmospheric moisture content in the data, creating conditions conducive to fog formation and subsequent water condensation processes inland desert. Conversely, during periods characterized by lower TIL levels, dew point recordings typically exhibit a decrease. This dynamic underscores the central role of oceanic drivers in modulating atmospheric moisture content, thereby influencing the formation of fog and the ensuing processes of fog water collection inland desert.

At the local scale, both OLS and GWR models demonstrated adequate capabilities in exploring the spatial variability of Near-surface fog water. While the OLS model provided a global perspective by assuming uniform relationships across the study site, the GWR model effectively captured localized variations, offering deeper insights into the influences of Oyarbide topography as explanatory variables.

Enhancing the comprehension of intricate and dynamic spatial processes is significantly advanced through the application of an integrated methodology. This local spatial modeling approach, such as GWR, combined with a repertoire of analytical and visualization instruments like GIS and remote sensing, not only facilitates the creation of refined models but also allows a better visualization of the

spatial heterogeneities and locally disaggregation statistical data with a heightened level of granularity and precision, contributing substantially to advancing the understanding of complex spatial dynamics in biosphere-atmosphere interactions. Moreover, disaggregating the global  $R^2$  into localized coefficients and analyzing their geographic distribution across Oyarbide provides valuable insights into regions with varying explanatory power, influenced by independent topographical variables in modeling Near-surface fog water variability.

The empirical research in Oyarbide has consistently demonstrated that the foundational assumption of global statistics presupposes homogeneity in the relationships between predictors and the criterion variable across spatial domains, suggesting a uniform response to a factor throughout the study site. However, real-world dynamics, particularly evident in the inland desert through local regression analysis, challenge this assumption. The relationships between atmospheric-topographical variables display significant heterogeneity, manifesting as geographical variations where the terrain roughness generates notably divergent responses in the Near-surface fog variability across the study site.

By advancing the scientific understanding of fog climate dynamics in this hyper arid environment, these findings enhance broader scientific understanding of the atmosphere-biosphere interactions and contribute to the fog ecosystems modeling in the climate change context.

---

## References

- Abdul-Rahim, A. S., Kim, Y., & Yue, L. (2022). Investigating Spatial Heterogeneity of the Environmental Kuznets Curve for Haze Pollution in China. *Atmosphere*, *13*(5), Article 806. <https://doi.org/10.3390/atmos13050806>
- Alexandersson, H. (1986). A homogeneity test applied to precipitation data. *Journal of Climatology*, *6*(6), 661–675. <https://doi.org/10.1002/joc.3370060607>
- Anselin, L. (1995). Local Indicators of Spatial Association—LISA. *Geographical Analysis*, *27*(2), 93–115. <https://doi.org/10.1111/j.1538-4632.1995.tb00338.x>
- Bendix, J. (2002). A satellite-based climatology of fog and low-level stratus in Germany and adjacent areas. *Atmospheric Research*, *64*(1-4), 3–18. [https://doi.org/10.1016/S0169-8095\(02\)00075-3](https://doi.org/10.1016/S0169-8095(02)00075-3)
- Berterretche, M., Hudak, A. T., Cohen, W. B., Maieresperger, T. K., Gower, S. T., & Dungan, J. (2005). Comparison of regression and geostatistical methods for mapping Leaf Area Index (LAI) with Landsat ETM+ data over a boreal forest. *Remote Sensing of Environment*, *96*(1), 49–61. <https://doi.org/10.1016/j.rse.2005.01.014>
- Błaś, M., Sobik, M., Quiel, F., & Netzel, P. (2002). Temporal and spatial variations of fog in the Western Sudety Mts., Poland. *Atmospheric Research*, *64*(1-4), 19–28. [https://doi.org/10.1016/S0169-8095\(02\)00076-5](https://doi.org/10.1016/S0169-8095(02)00076-5)
- Borges, P. de A., Franke, J., Da Anunciação, Y. M. T., Weiss, H., & Bernhofer, C. (2016). Comparison of spatial interpolation methods for the estimation of precipitation distribution in Distrito Federal, Brazil. *Theoretical and Applied Climatology*, *123*(1-2), 335–348. <https://doi.org/10.1007/s00704-014-1359-9>
- Borthagaray, A. I., Fuentes, M. A., & Marquet, P. A. (2010). Vegetation pattern formation in a fog-dependent ecosystem. *Journal of Theoretical Biology*, *265*(1), 18–26. <https://doi.org/10.1016/j.jtbi.2010.04.020>
- Brunsdon, C., Fotheringham, A. S., & Charlton, M. (1998). Geographically Weighted Regression. *Journal of the Royal Statistical Society: Series D (the Statistician)*, *47*(3), 431–443. <https://doi.org/10.1111/1467-9884.00145>
- Brunsdon, C., Fotheringham, A. S., & Charlton, M. E. (1996). Geographically Weighted Regression: A Method for Exploring Spatial Nonstationarity. *Geographical Analysis*, *28*(4), 281–298. <https://doi.org/10.1111/j.1538-4632.1996.tb00936.x>

- Buishand, T. A. (1982). Some methods for testing the homogeneity of rainfall records. *Journal of Hydrology*, 58(1-2), 11–27. [https://doi.org/10.1016/0022-1694\(82\)90066-X](https://doi.org/10.1016/0022-1694(82)90066-X)
- Cáceres, L., Gómez-Silva, B., Garró, X., Rodríguez, V., Monardes, V., & McKay, C. P. (2007). Relative humidity patterns and fog water precipitation in the Atacama Desert and biological implications. *Journal of Geophysical Research: Biogeosciences*, 112(G4), Article G04S14. <https://doi.org/10.1029/2006JG000344>
- Cao, L.-J., & Yan, Z.-W. (2012). Progress in Research on Homogenization of Climate Data. *Advances in Climate Change Research*, 3(2), 59–67. <https://doi.org/10.3724/SP.J.1248.2012.00059>
- Cereceda, P., & Schemenauer, R. S. (1991). The Occurrence of Fog in Chile. *Journal of Applied Meteorology*, 30(8), 1097–1105. [https://doi.org/10.1175/1520-0450\(1991\)030<1097:TOOFIC>2.0.CO;2](https://doi.org/10.1175/1520-0450(1991)030<1097:TOOFIC>2.0.CO;2)
- Cereceda, P., Larraín, H., Lázaro, P., Osses, P., Schemenauer, R. S., & Fuentes, L. (1999). Campos de Tillandsiales y niebla en el desierto de Tarapaca. *Revista De Geografía Norte Grande*, 26, 3–13. Retrieved from <https://ojs.uc.cl/index.php/RGNG/article/view/42741>.
- Cereceda, P., Larrain, H., Osses, P., Farías, M., & Egaña, I. (2008a). The climate of the coast and fog zone in the Tarapacá Region, Atacama Desert, Chile. *Atmospheric Research*, 87(3-4), 301–311. <https://doi.org/10.1016/j.atmosres.2007.11.011>
- Cereceda, P., Larrain, H., Osses, P., Farías, M., & Egaña, I. (2008b). The spatial and temporal variability of fog and its relation to fog oases in the Atacama Desert, Chile. *Atmospheric Research*, 87(3-4), 312–323. <https://doi.org/10.1016/j.atmosres.2007.11.012>
- Cereceda, P., Osses, P., Larraín, H., & Pérez, M. (2007, July 22). *Biogeographical and phytosociological characteristics of the Tillandsia fields at the fog oasis of Cerro Guanaco, Tarapaca, Chile*. Fourth International Conference on Fog, Fog Collection and Dew, La Serena, Chile.
- Cereceda, P., Osses, P., Larrain, H., Farías, M., Lagos, M., Pinto, R., & Schemenauer, R. S. (2002). Advective, orographic and radiation fog in the Tarapacá region, Chile. *Atmospheric Research*, 64(1-4), 261–271. [https://doi.org/10.1016/S0169-8095\(02\)00097-2](https://doi.org/10.1016/S0169-8095(02)00097-2)
- Cerlini, P. B., Silvestri, L., & Saraceni, M. (2020). Quality control and gap-filling methods applied to hourly temperature observations over central Italy. *Meteorological Applications*, 27(3), Article e1913. <https://doi.org/10.1002/met.1913>
- Cermak, J., Eastman, R. M., Bendix, J., & Warren, S. G. (2009). European climatology of fog and low stratus based on geostationary satellite observations. *Quarterly Journal of the Royal Meteorological Society*, 135(645), 2125–2130. <https://doi.org/10.1002/qj.503>

- 
- Chen, X. F., Chen, J. M., An, S. Q., & Ju, W. M. (2007). Effects of topography on simulated net primary productivity at landscape scale. *Journal of Environmental Management*, *85*(3), 585–596. <https://doi.org/10.1016/j.jenvman.2006.04.026>
- David, F. N., & Tukey, J. W. (1977). Exploratory Data Analysis. *Biometrics*, *33*(4), 5–23. <https://doi.org/10.2307/2529486>
- Del Río, C. (2019). *Spatiotemporal characteristics of coastal fog in the Atacama Desert. A remote sensing based analysis of the past, present and future distribution and variability of low clouds under climate change in a hyper-arid region of northern Chile* [Doctoral Thesis]. Heidelberg University, Germany. <https://doi.org/10.11588/heidok.00027922>
- Del Río, C., Lobos-Roco, F., Latorre, C., Koch, M. A., García, J.-L., Osses, P., Lambert, F., Alfaro, F., & Siegmund, A. (2021). Spatial distribution and interannual variability of coastal fog and low clouds cover in the hyperarid Atacama Desert and implications for past and present *Tillandsia landbeckii* ecosystems. *Plant Systematics and Evolution*, *307*(5), Article 58. <https://doi.org/10.1007/s00606-021-01782-z>
- Del Río, C., Rivera, D., Siegmund, A., Wolf, N., Cereceda, P., Larraín, H., Lobos, F., Garcia, J. L., Osses, P., Zanetta, N., & Lambert, F. (2018). ENSO Influence on Coastal Fog-Water Yield in the Atacama Desert, Chile. *Aerosol and Air Quality Research*, *18*(1), 127–144. <https://doi.org/10.4209/aaqr.2017.01.0022>
- Doraiswamy, P. C., Pasteris, P. A., Jones, K. C., Motha, R. P., & Nejedlik, P. (2000). Techniques for methods of collection, database management and distribution of agrometeorological data. *Agricultural and Forest Meteorology*, *103*(1-2), 83–97. [https://doi.org/10.1016/S0168-1923\(00\)00120-9](https://doi.org/10.1016/S0168-1923(00)00120-9)
- Easterling, D. R., & Peterson, T. C. (1995). A new method for detecting undocumented discontinuities in climatological time series. *International Journal of Climatology*, *15*(4), 369–377. <https://doi.org/10.1002/joc.3370150403>
- Farías, M., Cerceda, P., Osses, P., & Núñez, R. (2005). Comportamiento espacio-temporal de la nube estratocúmulo, productora de niebla en la costa del desierto de Atacama (21° lat. S., 70° long. W.), durante un mes de invierno y otro de verano. *Investigaciones Geográficas*(56), 43–61. <https://doi.org/10.14350/rig.30096>
- Feng, S., Hu, Q., & Qian, W. (2004). Quality control of daily meteorological data in China, 1951–2000: a new dataset. *International Journal of Climatology*, *24*(7), 853–870. <https://doi.org/10.1002/joc.1047>
- Fernandes, R., & Leblanc, S. G. (2005). Parametric (modified least squares) and non-parametric (Theil–Sen) linear regressions for predicting biophysical parameters in the presence of measurement

- errors. *Remote Sensing of Environment*, 95(3), 303–316. <https://doi.org/10.1016/j.rse.2005.01.005>
- Fessehaye, M., Abdul-Wahab, S. A., Savage, M. J., Kohler, T., Gherezghiher, T., & Hurni, H. (2014). Fog-water collection for community use. *Renewable and Sustainable Energy Reviews*, 29, 52–62. <https://doi.org/10.1016/j.rser.2013.08.063>
- Foody, G. M. (2003). Geographical weighting as a further refinement to regression modelling: An example focused on the NDVI–rainfall relationship. *Remote Sensing of Environment*, 88(3), 283–293. <https://doi.org/10.1016/j.rse.2003.08.004>
- Fotheringham, A. S., Brunsdon, C., & Charlton, M. (2002). *Geographically Weighted Regression: The Analysis of Spatially Varying Relationships* (1st ed.). Wiley.
- Fotheringham, A. S., Charlton, M. E., & Brunsdon, C. (1998). Geographically Weighted Regression: A Natural Evolution of the Expansion Method for Spatial Data Analysis. *Environment and Planning a: Economy and Space*, 30(11), 1905–1927. <https://doi.org/10.1068/a301905>
- Fotheringham, A. S., Charlton, M. E., & Brunsdon, C. (2001). Spatial Variations in School Performance: A Local Analysis Using Geographically Weighted Regression. *Geographical and Environmental Modelling*, 5(1), 43–66. <https://doi.org/10.1080/13615930120032617>
- Fotheringham, A. S., Crespo, R., & Yao, J. (2015). Geographical and Temporal Weighted Regression (GTWR). *Geographical Analysis*, 47(4), 431–452. <https://doi.org/10.1111/gean.12071>
- Fotheringham, A. S., Yue, H., & Li, Z. (2019). Examining the influences of air quality in China's cities using multi-scale geographically weighted regression. *Transactions in GIS*, 23(6), 1444–1464. <https://doi.org/10.1111/tgis.12580>
- Fox, J. C., Ades, P. K., & Bi, H. (2001). Stochastic structure and individual-tree growth models. *Forest Ecology and Management*, 154(1-2), 261–276. [https://doi.org/10.1016/S0378-1127\(00\)00632-0](https://doi.org/10.1016/S0378-1127(00)00632-0)
- Frescino, T. S., Edwards, T. C., & Moisen, G. G. (2001). Modeling Spatially Explicit Forest Structural Attributes Using Generalized Additive Models. *Journal of Vegetation Science*, 12(1), 15–26. <https://doi.org/10.2307/3236670>
- García, J.-L., Lobos-Roco, F., Schween, J. H., Del Río, C., Osses, P., Vives, R., Pezoa, M., Siegmund, A., Latorre, C., Alfaro, F., Koch, M. A., & Loehnert, U. (2021). Climate and coastal low-cloud dynamic in the hyperarid Atacama fog Desert and the geographic distribution of *Tillandsia landbeckii* (Bromeliaceae) dune ecosystems. *Plant Systematics and Evolution*, 307(57), 1–22. <https://doi.org/10.1007/s00606-021-01775-y>

- 
- Garreaud, R., Barichivich, J., Christie, D. A., & Maldonado, A. (2008). Interannual variability of the coastal fog at Fray Jorge relict forests in semiarid Chile. *Journal of Geophysical Research: Biogeosciences*, *113*(G4), Article G04011. <https://doi.org/10.1029/2008JG000709>
- Getis, A., & Ord, J. K. (1992). The Analysis of Spatial Association by Use of Distance Statistics. *Geographical Analysis*, *24*(3), 189–206. <https://doi.org/10.1111/j.1538-4632.1992.tb00261.x>
- Goovaerts, P. (1998). *Mathematical Geology*, *30*(1), 21–42. <https://doi.org/10.1023/A:1021757104135>
- Goovaerts, P. (2000). Geostatistical approaches for incorporating elevation into the spatial interpolation of rainfall. *Journal of Hydrology*, *228*(1-2), 113–129. [https://doi.org/10.1016/S0022-1694\(00\)00144-X](https://doi.org/10.1016/S0022-1694(00)00144-X)
- Grunow, J. (1952). Nebelniederschlag. Bedeutung und Erfassung einer Zusatzkomponente des Niederschlags. *Berichte Des Deutschen Wetterdienst in Der US-Zone*, *42*, 30–34.
- Gutiérrez, J., Cardozo, O. D., & García, J. C. (2013). Estimación directa de la demanda de transporte a nivel de estación mediante el uso de la regresión geográficamente ponderada, *Congreso Chileno De Ingeniería De Transporte (16)*, pp. 1–15. Retrieved from <https://revistaeggp.uchile.cl/index.php/CIT/article/view/28422>.
- Haining, R. (1977). Model Specification in Stationary Random Fields. *Geographical Analysis*, *9*(2), 107–129. <https://doi.org/10.1111/j.1538-4632.1977.tb00566.x>
- Hanham, R., & Spiker, J. S. (2005). Urban Sprawl Detection Using Satellite Imagery and Geographically Weighted Regression. In *Geo-Spatial Technologies in Urban Environments* (pp. 137–151). Springer-Verlag. [https://doi.org/10.1007/3-540-26676-3\\_12](https://doi.org/10.1007/3-540-26676-3_12)
- Hook, P. B., & Burke, I. C. (2000). Biogeochemistry in a Shortgrass Landscape: Control by Topography, Soil Texture, and Microclimate. *Ecology*, *81*(10), 2686–2703. [https://doi.org/10.1890/0012-9658\(2000\)081\[2686:BIASLC\]2.0.CO;2](https://doi.org/10.1890/0012-9658(2000)081[2686:BIASLC]2.0.CO;2)
- Hurvich, C. M., Simonoff, J. S., & Tsai, C.-L. (1998). Smoothing Parameter Selection in Nonparametric Regression Using an Improved Akaike Information Criterion. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, *60*(2), 271–293. <https://doi.org/10.1111/1467-9868.00125>
- Jaksic, F. M., Torres-Mura, J. C., Cornelius, C., & Marquet, P. A. (1999). Small mammals of the Atacama Desert (Chile). *Journal of Arid Environments*, *42*(2), 129–135. <https://doi.org/10.1006/jare.1999.0512>
- Jarque, C. M., & Bera, A. K. (1987). A Test for Normality of Observations and Regression Residuals. *International Statistical Review / Revue Internationale De Statistique*, *55*(2), 163–172. <https://doi.org/10.2307/1403192>

- 
- Jetz, W., Rahbek, C., & Lichstein, J. W. (2005). Local and global approaches to spatial data analysis in ecology. *Global Ecology and Biogeography*, *14*(1), 97–98. <https://doi.org/10.1111/j.1466-822X.2004.00129.x>
- Kendall, M. G. (1975). *Rank correlation methods* (4th ed., 2nd impr). C. Griffin.
- Kim, T.-W., Ahn, H., Ahn, J.-H., Byeon, S.-H., Wi, S.-W., & Kim, M. (2007). Estimating Missing Values of Daily Rainfall Using Classification Techniques. *Proceedings of the 32nd IAHR World Congress (Venice, 2007)*, 1–8.
- Koch, M. A., Kleinpeter, D., Auer, E., Siegmund, A., Del Rio, C., Osses, P., García, J.-L., Marzol, M. V., Zizka, G., & Kiefer, C. (2019). Living at the dry limits: ecological genetics of *Tillandsia landbeckii* lomas in the Chilean Atacama Desert. *Plant Systematics and Evolution*, *305*(10), 1041–1053. <https://doi.org/10.1007/s00606-019-01623-0>
- Koenker, R. (1981). A note on studentizing a test for heteroscedasticity. *Journal of Econometrics*, *17*(1), 107–112. [https://doi.org/10.1016/0304-4076\(81\)90062-2](https://doi.org/10.1016/0304-4076(81)90062-2)
- Kruskal, W. H., & Wallis, W. A. (1952). Use of Ranks in One-Criterion Variance Analysis. *Journal of the American Statistical Association*, *47*(260), 583–621. <https://doi.org/10.1080/01621459.1952.10483441>
- Lange, C. (2003). Fog frequency and chemical composition of fog water—a relevant contribution to atmospheric deposition in the eastern Erzgebirge, Germany. *Atmospheric Environment*, *37*(26), 3731–3739. [https://doi.org/10.1016/S1352-2310\(03\)00350-9](https://doi.org/10.1016/S1352-2310(03)00350-9)
- Larrain, H., Velásquez, F., Cereceda, P., Espejo, R., Pinto, R., Osses, P., & Schemenauer, R. S. (2002). Fog measurements at the site “Falda Verde” north of Chañaral compared with other fog stations of Chile. *Atmospheric Research*, *64*(1-4), 273–284. [https://doi.org/10.1016/S0169-8095\(02\)00098-4](https://doi.org/10.1016/S0169-8095(02)00098-4)
- Latorre, C., González, A. L., Quade, J., Fariña, J. M., Pinto, R., & Marquet, P. A. (2011). Establishment and formation of fog-dependent *Tillandsia landbeckii* dunes in the Atacama Desert: Evidence from radiocarbon and stable isotopes. *Journal of Geophysical Research: Biogeosciences*, *116*(G3), Article G03033. <https://doi.org/10.1029/2010JG001521>
- Legendre, P. (1993). Spatial Autocorrelation: Trouble or New Paradigm? *Ecology*, *74*(6), 1659–1673. <https://doi.org/10.2307/1939924>
- Legendre, P., & Legendre, L. (1998). *Numerical ecology* (2. Engl. ed., reprinted.). *Developments in environmental modelling: Vol. 20*. Elsevier Science.
- Lehnert, L. W., Thies, B., Trachte, K., Achilles, S., Osses, P., Baumann, K., Bendix, J., Schmidt, J., Samolov, E., Jung, P., Leinweber, P., Karsten, U., & Büdel, B. (2018). A Case Study on Fog/Low Stratus Occurrence at Las Lomitas, Atacama Desert (Chile) as a Water Source for Biological Soil

- Crusts. *Aerosol and Air Quality Research*, 18(1), 254–269.  
<https://doi.org/10.4209/aaqr.2017.01.0021>
- Liu, W., Han, Y., Li, J., Tian, X., & Liu, Y. (2018). Factors affecting relative humidity and its relationship with the long-term variation of fog-haze events in the Yangtze River Delta. *Atmospheric Environment*, 193, 242–250. <https://doi.org/10.1016/j.atmosenv.2018.09.015>
- Lloyd, C., & Shuttleworth, I. (2005). Analysing Commuting Using Local Regression Techniques: Scale, Sensitivity, and Geographical Patterning. *Environment and Planning a: Economy and Space*, 37(1), 81–103. <https://doi.org/10.1068/a36116>
- Lund, R., & Reeves, J. (2002). Detection of Undocumented Changepoints: A Revision of the Two-Phase Regression Model. *Journal of Climate*, 15(17), 2547–2554. [https://doi.org/10.1175/1520-0442\(2002\)015<2547:DOUCAR>2.0.CO;2](https://doi.org/10.1175/1520-0442(2002)015<2547:DOUCAR>2.0.CO;2)
- Mair, A., & Fares, A. (2010). Assessing Rainfall Data Homogeneity and Estimating Missing Records in Mākaha Valley, O‘ahu, Hawai‘i. *Journal of Hydrologic Engineering*, 15(1), 61–66. [https://doi.org/10.1061/\(ASCE\)HE.1943-5584.0000145](https://doi.org/10.1061/(ASCE)HE.1943-5584.0000145)
- Mann, H. B. (1945). Nonparametric Tests Against Trend. *Econometrica*, 13(3), 245–259. <https://doi.org/10.2307/1907187>
- Mikulane, S., Siegmund, A., Del Río, C., Koch, M. A., Osses, P., & García, J.-L. (2022). Remote sensing based mapping of Tillandsia fields - A semi-automatic detection approach in the hyperarid coastal Atacama Desert, northern Chile. *Journal of Arid Environments*, 205, Article 104821. <https://doi.org/10.1016/j.jaridenv.2022.104821>
- Miller, A. (1976). The climate of Chile: World Survey Climatol., 12, 113–145. <https://doi.org/10.1002/qj.49710343520>
- Miller, J. A. (2012). Species distribution models. *Progress in Physical Geography: Earth and Environment*, 36(5), 681–692. <https://doi.org/10.1177/0309133312442522>
- Moat, J., Orellana-Garcia, A., Tovar, C., Arakaki, M., Arana, C., Cano, A., Faundez, L., Gardner, M., Hechenleitner, P., Hepp, J., Lewis, G., Mamani, J.-M., Miyasiro, M., & Whaley, O. Q. (2021). Seeing through the clouds – Mapping desert fog oasis ecosystems using 20 years of MODIS imagery over Peru and Chile. *International Journal of Applied Earth Observation and Geoinformation*, 103, Article 102468. <https://doi.org/10.1016/j.jag.2021.102468>
- Montecinos, S., Cereceda, P., & Rivera, D. (2018). Fog collection and its relationship with local meteorological variables in a semiarid zone in Chile. *Atmósfera*, 31(2), 143–153. <https://doi.org/10.20937/atm.2018.31.02.03>
- Montgomery, D. C., Peck, E. A., & Vining, G. G. (2021). *Introduction to linear regression analysis* (6th edition). *Wiley series in probability and statistics*. Wiley.

- 
- Niedzielski, T., & Migala, K. (2018). *Geoinformatics and Atmospheric Science* (1st edition). Birkhäuser.  
<https://doi.org/10.1007/978-3-319-66092-9>
- Olaya, V. (2009). Chapter 6 Basic Land-Surface Parameters. In *Developments in Soil Science. Geomorphometry - Concepts, Software, Applications* (Vol. 33, pp. 141–169). Elsevier.  
[https://doi.org/10.1016/S0166-2481\(08\)00006-8](https://doi.org/10.1016/S0166-2481(08)00006-8)
- Osborne, P. E., Foody, G. M., & Suárez-Seoane, S. (2007). Non-stationarity and local approaches to modelling the distributions of wildlife. *Diversity and Distributions*, 13(3), 313–323.  
<https://doi.org/10.1111/j.1472-4642.2007.00344.x>
- Osses, P., Cereceda, P., Larrain, H., & Astaburuaga, J. P. (2007). *Distribution and geographical factors of the Tillandsia fields in the coastal mountain range of the Tarapacá Region, Chile*. Fourth International Conference on Fog, Fog Collection and Dew, La Serena, Chile.
- Osses, P., Escobar, R., Del Rio, C., Garcia, R., & Vargas, C. (2017). El Clima desértico costero con nublados abundantes del desierto de Atacama y su relación con los recursos naturales energía solar y agua de niebla. Caso de estudio Alto Patache (20,5°S), región de Tarapacá, Chile. *Revista De Geografía Norte Grande*(68), 33–48. <https://doi.org/10.4067/S0718-34022017000300033>
- Páez, A., Farber, S., & Wheeler, D. (2011). A Simulation-Based Study of Geographically Weighted Regression as a Method for Investigating Spatially Varying Relationships. *Environment and Planning a: Economy and Space*, 43(12), 2992–3010. <https://doi.org/10.1068/a44111>
- Paulhus, J. L. H., & Kohler, M. A. (1952). Interpolation of Missing Precipitation Records. *Monthly Weather Review*, 80(8), 129–133. [https://doi.org/10.1175/1520-0493\(1952\)080<0129:IOMPR>2.0.CO;2](https://doi.org/10.1175/1520-0493(1952)080<0129:IOMPR>2.0.CO;2)
- Pettitt, A. N. (1979). A Non-Parametric Approach to the Change-Point Problem. *Applied Statistics*, 28(2), 126–135. <https://doi.org/10.2307/2346729>
- Platt, R. V. (2004). Global and local analysis of fragmentation in a mountain region of Colorado. *Agriculture, Ecosystems & Environment*, 101(2-3), 207–218.  
<https://doi.org/10.1016/j.agee.2003.09.005>
- Pratt, B., & Chang, H. (2012). Effects of land cover, topography, and built structure on seasonal water quality at multiple spatial scales. *Journal of Hazardous Materials*, 209-210, 48–58.  
<https://doi.org/10.1016/j.jhazmat.2011.12.068>
- Roberts, D. W. (1986). Ordination on the basis of fuzzy set theory. *Vegetatio*, 66(3), 123–131.  
<https://doi.org/10.1007/BF00039905>

- 
- Rundel, P. W., Dillon, M. O., Palma, B., Mooney, H. A., Gulmon, S. L., & Ehleringer, J. R. (1991). The Phytogeography and Ecology of the Coastal Atacama and Peruvian Deserts. *Aliso*, *13*(1), 1–49. <https://doi.org/10.5642/aliso.19911301.02>
- Rutllant, J. A., Fuenzalida, H., & Aceituno, P. (2003). Climate dynamics along the arid northern coast of Chile: The 1997–1998 Dinámica del Clima de la Región de Antofagasta (DICLIMA) experiment. *Journal of Geophysical Research: Biogeosciences*, *108*(D17), Article 4538. <https://doi.org/10.1029/2002JD003357>
- Sachdeva, M., & Fotheringham, A. S. (2020). The Geographically Weighted Regression Framework. *Geographic Information Science & Technology Body of Knowledge*, *2020*(Q4). <https://doi.org/10.22224/gistbok/2020.4.7>
- Sarricolea, P., Meseguer Ruiz, O., & Romero-Aravena, H. (2017). Tendencias de la precipitación en el Norte Grande de Chile y su relación con las proyecciones de cambio climático. *Diálogo Andino*(54), 41–50. <https://doi.org/10.4067/S0719-26812017000300041>
- Schemenauer, R. S., & Cereceda, P. (1994a). Fog collection's role in water planning for developing countries. *Natural Resources Forum*, *18*(2), 91–100. <https://doi.org/10.1111/j.1477-8947.1994.tb00879.x>
- Schemenauer, R. S., & Cereceda, P. (1994b). A Proposed Standard Fog Collector for Use in High-Elevation Regions. *Journal of Applied Meteorology*, *33*(11), 1313–1322. [https://doi.org/10.1175/1520-0450\(1994\)033<1313:APSFCE>2.0.CO;2](https://doi.org/10.1175/1520-0450(1994)033<1313:APSFCE>2.0.CO;2)
- Schulz, N., Boisier, J. P., & Aceituno, P. (2012). Climate change along the arid coast of northern Chile. *International Journal of Climatology*, *32*(12), 1803–1814. <https://doi.org/10.1002/joc.2395>
- Schween, J. H., Hoffmeister, D., & Löhnert, U. (2020). Filling the observational gap in the Atacama Desert with a new network of climate stations. *Global and Planetary Change*, *184*, Article 103034. <https://doi.org/10.1016/j.gloplacha.2019.103034>
- Searcy, J. K., & Hardison, C. H. (1960). *Double Mass Curves. Manual of Hydrology: Part 1.: General Surface-Water Techniques*. Methods and practices of the Geological Survey. United States Government Printing Office.
- Smith, D. M., Kniveton, D. R., & Barrett, E. C. (1998). A Statistical Modeling Approach to Passive Microwave Rainfall Retrieval. *Journal of Applied Meteorology*, *37*(2), 135–154. [https://doi.org/10.1175/1520-0450\(1998\)037<0135:ASMATP>2.0.CO;2](https://doi.org/10.1175/1520-0450(1998)037<0135:ASMATP>2.0.CO;2)
- Soares, P. M. M., Cardoso, R. M., Semedo, Á., Chinita, M. J., & Ranjha, R. (2022). Climatology of the Iberia coastal low-level wind jet: weather research forecasting model high-resolution results. *Tellus a: Dynamic Meteorology and Oceanography*, *66*(1), Article 22377. <https://doi.org/10.3402/tellusa.v66.22377>

- 
- Szentimrey, T. (1999). Multiple Analysis of Series for Homogenization (MASH). *Proc. Second Seminar for Homogenization of Surface Climatological Data, WMO WCDMP-41(OMSZ)*, 27–46.
- Tay, R. (2017). Correlation, Variance Inflation and Multicollinearity in Regression Model. *Journal of the Eastern Asia Society for Transportation Studies*, 12, 2006–2015. <https://doi.org/10.11175/easts.12.2006>
- Tobler, W. R. (1970). A Computer Movie Simulating Urban Growth in the Detroit Region. *Economic Geography*, 46, 234–240. <https://doi.org/10.2307/143141>
- Trewin, B. (2013). A daily homogenized temperature data set for Australia. *International Journal of Climatology*, 33(6), 1510–1529. <https://doi.org/10.1002/joc.3530>
- Vautard, R., Yiou, P., & van Oldenborgh, G. J. (2009). Decline of fog, mist and haze in Europe over the past 30 years. *Nature Geoscience*, 2(2), 115–119. <https://doi.org/10.1038/ngeo414>
- Wang, Q., Ni, J., & Tenhunen, J. (2005). Application of a geographically-weighted regression analysis to estimate net primary production of Chinese forest ecosystems. *Global Ecology and Biogeography*, 14(4), 379–393. <https://doi.org/10.1111/j.1466-822X.2005.00153.x>
- Weischet, W. (1975). Las condiciones climáticas del Desierto de Atacama como desierto extremo de la Tierra. *Revista De Geografía Norte Grande*, 1(3-4), pp. 363–373. Retrieved from <https://revistanortegrande.uc.cl/index.php/RGNG/article/view/39689>.
- Westbeld, A., Klemm, O., Griebbaum, F., Sträter, E., Larrain, H., Osses, P., & Cereceda, P. (2009). Fog deposition to a *Tillandsia* carpet in the Atacama Desert. *Annales Geophysicae*, 27(9), 3571–3576. <https://doi.org/10.5194/angeo-27-3571-2009>
- Windle, M. J. S., Rose, G. A., Devillers, R., & Fortin, M.-J. (2010). Exploring spatial non-stationarity of fisheries survey data using geographically weighted regression (GWR): an example from the Northwest Atlantic. *ICES Journal of Marine Science*, 67(1), 145–154. <https://doi.org/10.1093/icesjms/fsp224>
- World Meteorological Organization (WMO). (2008). *Guide to meteorological instruments and methods of observation* (7th ed.). WMO: No. 8. Geneva, Switzerland: World Meteorological Organization.
- World Meteorological Organization (WMO). (2011). *Guide to Climatological Practices*. WMO: No. 100. Geneva, Switzerland: World Meteorological Organization.
- Wolf, N., Siegmund, A., Del Río, C., Osses, P., & García, J. L. (2016). Remote sensing-based detection and spatial pattern analysis for geo-ecological niche modeling of *Tillandsia* spp. in the Atacama, Chile. *ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XLI-B2, 251–256. <https://doi.org/10.5194/isprsarchives-XLI-B2-251-2016>

- Yu, D., & Wu, C. (2004). Understanding Population Segregation from Landsat ETM+ Imagery: A Geographically Weighted Regression Approach. *GIScience & Remote Sensing*, 41(3), 187–206. <https://doi.org/10.2747/1548-1603.41.3.187>
- Zahumensky, I. (2004). Guidelines on Quality Control Procedures for Data from Automatic Weather Stations (WMO-No. 955).
- Zhang, L., & Gove, J. H. (2005). Spatial Assessment of Model Errors from Four Regression Techniques. *Forest Science*, 51(4), 334–346. <https://doi.org/10.1093/forests/51.4.334>
- Zhang, L., & Shi, H. (2004). Local Modeling of Tree Growth by Geographically Weighted Regression. *Forest Science*, 50(2), 225–244. <https://doi.org/10.1093/forests/50.2.225>
- Zhang, L., Huiquan, B., Pengfei, C., & Davis, C. J. (2004). Modeling spatial variation in tree diameter–height relationships. *Forest Ecology and Management*, 189(1-3), 317–329. <https://doi.org/10.1016/j.foreco.2003.09.004>
- Zhao, N., Yang, Y., & Zhou, X. (2010). Application of geographically weighted regression in estimating the effect of climate and site conditions on vegetation distribution in Haihe Catchment, China. *Plant Ecology*, 209(2), 349–359. <https://doi.org/10.1007/s11258-010-9769-y>
- Zhou, Q., Wang, C., & Fang, S. (2019). Application of geographically weighted regression (GWR) in the analysis of the cause of haze pollution in China. *Atmospheric Pollution Research*, 10(3), 835–846. <https://doi.org/10.1016/j.apr.2018.12.012>





**Eidesstattliche Versicherung gemäß § 8 der Promotionsordnung für die Gesamtfakultät für Mathematik, Ingenieur- und Naturwissenschaften der Universität Heidelberg / Sworn Affidavit according to § 8 of the doctoral degree regulations of the Combined Faculty of Mathematics, Engineering and Natural Sciences at the Heidelberg University**

1. Bei der eingereichten Dissertation zu dem Thema / The thesis I have submitted entitled

.....

handelt es sich um meine eigenständig erbrachte Leistung / is my own work.

2. Ich habe nur die angegebenen Quellen und Hilfsmittel benutzt und mich keiner unzulässigen Hilfe Dritter bedient. Insbesondere habe ich wörtlich oder sinngemäß aus anderen Werken übernommene Inhalte als solche kenntlich gemacht. / I have only used the sources indicated and have not made unauthorised use of services of a third party. Where the work of others has been quoted or reproduced, the source is always given.

3. Die Arbeit oder Teile davon habe ich wie folgt/bislang nicht<sup>1)</sup> an einer Hochschule des In- oder Auslands als Bestandteil einer Prüfungs- oder Qualifikationsleistung vorgelegt. / I have not yet/have already<sup>1)</sup> presented this thesis or parts thereof to a university as part of an examination or degree.

Titel der Arbeit / Title of the thesis:.....

Hochschule und Jahr / University and year:.....

Art der Prüfungs- oder Qualifikationsleistung / Type of examination or degree:.....

4. Die Richtigkeit der vorstehenden Erklärungen bestätige ich. / I confirm that the declarations made above are correct.

5. Die Bedeutung der eidesstattlichen Versicherung und die strafrechtlichen Folgen einer unrichtigen oder unvollständigen eidesstattlichen Versicherung sind mir bekannt. / I am aware of the importance of a sworn affidavit and the criminal prosecution in case of a false or incomplete affidavit

Ich versichere an Eides statt, dass ich nach bestem Wissen die reine Wahrheit erklärt und nichts verschwiegen habe. / I affirm that the above is the absolute truth to the best of my knowledge and that I have not concealed anything.

.....  
Ort und Datum / Place and date

.....  
Unterschrift / Signature

<sup>1)</sup> Nicht Zutreffendes streichen. Bei Bejahung sind anzugeben: der Titel der andernorts vorgelegten Arbeit, die Hochschule, das Jahr der Vorlage und die Art der Prüfungs- oder Qualifikationsleistung. / Please cross out what is not applicable. If applicable, please provide: the title of the thesis that was presented elsewhere, the name of the university, the year of presentation and the type of examination or degree.