Discriminating Fog and Clouds in the Namib using Reanalysis Data

Karlsruher Institut für Technologie

Viola Hipler

Karlsruhe Institute of Technology (KIT), Institute for Geography and Geoecology, Reinhard-Baumeister-Platz 1, 76131 Karlsruhe

February 2024

Abstract

Fog is a frequent phenomenon in the Namib Desert and enables the life of a highly specialised flora and fauna. The fog day frequency varies from the coast to inland areas because the offshore stratus is advected at different elevations, becoming fog when touching the ground. Two situations can be identified from the perspective of a coastal station, fog occurrence and low cloud occurrence, but the drivers leading to the one or the other are unknown.

The presented study is a first attempt to fill this knowledge gap and develop a conceptual understanding of the different processes. To identify meteorological drivers, median composites of reanalysis (ERA5) variables were prepared for fog days and for low cloud days, using satellite-derived low cloud cover and station measurements as ground truth. The composites revealed that the main drivers of fog are the synoptic pressure, the temperature inversion at the Namibian coast and the local wind system. A conceptual framework was developed for how these drivers lead to fog or low cloud occurrence. In a second step, the potential of three different machine learning algorithms was assessed, given the task to differentiate fog and low cloud events based on meteorological data. The best classification skill was shown by the logistic regression (67 % accuracy, 69 % F1-score), followed by random forest (64.21 % accuracy, 63 % F1-score). A convolutional neural network overfitted instantly which was attributed to the small sample size.

The presented conceptual framework of the different processes leading to fog and low cloud occurrence is the first explicit disentangling of mechanistic drivers and seasonal variation of fog and low clouds in the Namib. The results show that fog and low clouds can be differentiated based on the drivers by machine learning, which opens the perspective of modeling the fog regime in the past and future Namib Desert.

Eidesstattliche Erklärung

Viola Hipler

Ich versichere wahrheitsgemäß, die Masterarbeit
"Discriminating Fog and Clouds in the Namib using Reanalysis Data"
(Institut für Meteorologie und Klimaforschung - Atmosphärische Spurengase und Fernerkundung)
selbstständig verfasst, alle benutzten Hilfsmittel vollständig und genau angegeben und alles kenntlich gemacht zu haben, was aus Arbeiten anderer unverändert oder mit Abänderungen entnommen wurde sowie die Satzung des KIT zur Sicherung guter wissenschaftlicher Praxis in der jeweils gültigen Fassung beachtet zu haben.

Karlsruhe, den 21. Februar 2024

Acknowledgements

I want to thank my supervisor Hendrik Andersen, who initiated my interest in fog and low clouds in the Namib, for his constant and often proactive support. Thanks also go to Jan Cermak, Sebastian Schmidtlein and the group members of SKL for creating a comfortable and inspiring work environment. Last but not least I am most grateful for the support of my friends and family members, whose enthusiasm and true interest in this work encouraged me during the whole creative process. Thereby, special thanks go to Edvinas Rommel for his substantial support.

Contents

List of Figures

List of Tables

List of Abbreviations

AMJ	April, May, June, fog season (Andersen et al., 2019)
ANN	artificial neural network
CA	continental anticyclone, same as CH
CBH	cloud base height
CH	continental high
CM	Coastal Met (coastal weather station in cetral Namib Desert)
CNN	convolutional neural network
DJF	December, January, February, austral summer
FDF	fog day frequency
FLC	fog and low clouds
FLS	fog and low stratus, synonym to FLC
ITC	intertropical convergence
JJA	June, July, August, austral winter
logreg	logistic regression
MLE	maximum likelihood estimation
MSLP	mean sea level pressure
NLTS	Namib Lower Tropospheric Stability
RF	random forest
RFE	recursive feature elimination
RH	relative humidity
SAA	south Atlantic anticyclone, same as SAH
SAH	south Atlantic high
SON	September, October, November, low cloud season (Andersen et al., 2019)
SST	sea surface temperature
q	specific humidity
t	temperature
u	wind component directing east
v	wind component directing north
w	wind component directing up
z	geopotential

Chapter 1

Introduction

1.1 The relevance of fog in the Namib

From the hyperarid Namib Desert, we know several, partly endemic, species that exploit fog water by evolutionary adaptations of their bodies or behaviour (Burke, 2007). As rain events are extremely rare, the only frequent and reliable water source is the nocturnal fog. Compared to other low clouds, only the ground-touching fog can be reached by the usually tiny or low-growing organisms that try to avoid wind and radiation. A prominent example is the fog-basking beetle *Onymacris unguicularis* that points its backside to the sky and collects the fog water condensated at its legs (Mitchell et al., 2020). While some of the Namib plants rely mostly on groundwater and rare rain events (e.g. *Welwitschia mirabilis* Henschel et al., 2019) and their long lifespan is a characteristic of the Namib flora (Schachtschneider and February, 2010), others are specialised on the sparse but regular water input by fog (e.g. Roth-Nebelsick et al., 2012). The lichen fields, as one characteristic element of the coastal vegetation (Juergens et al., 2013), are kept alive by the short but reliable events of night/morning fog that wet the lichens and thus allow metabolism and photosynthesis (Lange et al., 2007, Warren-Rhodes et al., 2013). Other plants take up fog water through the soil (Ebner et al., 2011), as a thin layer of a few centimeters is moistened by fog events (Li et al., 2018). Not only the water input but also the nutrients transported within the droplets maintain the life of many species, e.g. of the grass *Arthraerua leubnitziae* (Gottlieb et al., 2019).

The special adaptations of Namib organisms to collect fog water are not only beautiful in their diversity but have also inspired technical solutions to harvest fog water for human purposes (Zhong et al., 2018, J. K. Park and Kim, 2019, Lei and Guo, 2020, Yu et al., 2021, Yue et al., 2022). The fog season coincides with peak groundwater salinity which makes fog, as an ion-poor water source, an ideal dilutant to generate drinking water (Shanyengana et al., 2002). Fog water could also enable local agriculture and thus add to Namibia's food supply (Mupambwa et al., 2019).

Fog has maintained life in cold deserts like the Namib and the Atacama for millions of years. With the ongoing rise of atmospheric CO2 and already emerging changes of climate in some regions of the world (Norris et al., 2016), changes will likely happen also in the Namib, potentially threatening a unique ecosystem. While the impact of altered evaporation and rainfall on Namibian species has been assessed (Thuiller et al., 2006), predictions of the future fog regime in the Namib are rare. Fog days might increase at the coast and decrease inland (Haensler et al., 2011), but the main common pitfall of fog prediction studies is the scarcity of available validation data. As fog is defined as decreased ground visibility to lower than 1 km, true validation data can only be obtained on the ground. With 11 weather stations in the central Namib (fig. **1.3**), run by the Southern African Science Service Centre for Climate Change and Adaptive Land Management (SASSCAL) initiative, the data situation is comparably comfortable but far away from complete spatial coverage. As a source of spatially continuous data, many studies use satellite-derived low cloud products as validation. These products cannot, however, distinguish between low clouds and actually ground-touching fog. As the prediction cannot be better than the training or validation, the ecologically most important variable of ground-touching fog can to date not be predicted accurately. A fog occurrence dataset covering the whole Namib is vitally needed to promote research about the future of the Namib ecosystem. This study makes a first step to derive such a dataset of actual fog occurrence.

Satellite data offers the advantage of complete spatial coverage compared to station measurements. Therefore, we see the crucial problem in deriving the cloud base height from the satellite data to determine if a cloud touches the ground and thus provides water to the biosphere. While there are several approaches to determine the cloud base height directly from the imagery on larger scales, e.g. via cloud optical depth or microphysics (Cermak and Bendix, 2011), no approach has yet emerged that would be able to yield the required accuracy for the use case of Namib fog detection. We therefore propose to use meteorological variables from reanalysis to classify the low clouds detected by the satellite as a low, but not ground-touching, cloud or as fog.

There is a close interplay between the climate, including fog occurrence, and species richness and abundance (Hachfeld, Jürgens, et al., 2000). The spatial distribution of fog events determines the abundance of species like the flowering stones *Lithops ruschiorum* (Loots et al., 2019) and the grass *Arthraerua leubnitziae* (Gottlieb et al., 2019) in the coastal belt of the Namib as well as the amount of harvestable water for human use. The timing of fog events can have a remarkable influence on the amount of water available to plants (L. Wang et al., 2019). Despite the fact that Namib flora and fauna depend on fog occurrence (because of their obvious evolutional adaptations), large-scale evidence of this dependence could not be explicitly shown by satellite imagery nor could the relevant spatial and temporal patterns of fog occurrence be monitored. This is because the distinguishing of fog and low clouds is to date inaccurate and has to be approximated

and, in the case of the vegetation dependency, vegetation indices are inadequate to detect differences in such sparse vegetation as in the Namib (Qiao et al., 2020). Fog is a beautiful and interesting phenomenon, as is the life it supports, and the rapid climatic changes only emphasize the need to study it.

1.2 Geography of the Namib

1.2.1 Namib Desert geomorphology

The Namib Desert covers the southwestern coast of the south African subcontinent along roughly 2000 km, including the whole of Namibia's coast and parts of the Angolan coast. It comprises dunes, gravel plains, inselbergs and other landscape features and is bordered by the Great Escarpment, the beginning of the inland plateau. A major part is captured by the Namib-Naukluft National Park and the Namib sand sea is declared a UNESCO heritage site.

The terrain of the Namib is a gently rising flat plain that changes into the rough mountains (up to 2000 m high) of the Great Escarpment about 140 km from the coast (see fig. 1.3). North of the Kuiseb River the bedrock makes up the surface, with occasional gravel fields, deep river gorges and some inselbergs. In the south, the bedrock platform is buried by sandmasses (Logan, 1960). The so-called Namib Sand Sea is an active dune system with shifting dunes driven by strong winds and it overlays a second, ancient dune system.

1.2.2 Climate of the Namib and southern Africa

The Namib is a hyperarid desert with the central part receiving 5-18 mm/year water input from rain and the northern and southern parts receiving below 50 mm and 100 mm, respectively (Eckardt et al., 2013). There may well be consecutive years without any rain at all. However, there is water input from the frequently occurring fog, especially in the so-called fog belt about 20 - 60 km from the coast (up to 87 fog days per year at Swartbank). From the fog belt, the fog day frequency decreases towards the coast and inland towards the Great Escarpment (Seely and Henschel, 1998).

The synoptics of Southern Africa are dominated by the two quasi-permanent high-pressure cells of the subtropical high-pressure belt, the South Atlantic High (SAH) over the South Atlantic and the Indic High over the Indian Ocean (Weischet and Endlicher, 2000, fig. **1.1**). Both are of dynamic nature and part of the global Hadley circulation and can, mostly in austral winter, form a joint broad pressure cell covering the whole southern area of the subcontinent. They originate from tropical air that rises within the intertropical convergence zone (ITC) and comes down, much drier, in the mid-latitudes. The resulting high-pressure cells are stable throughout the year but vary in their latitudinal location

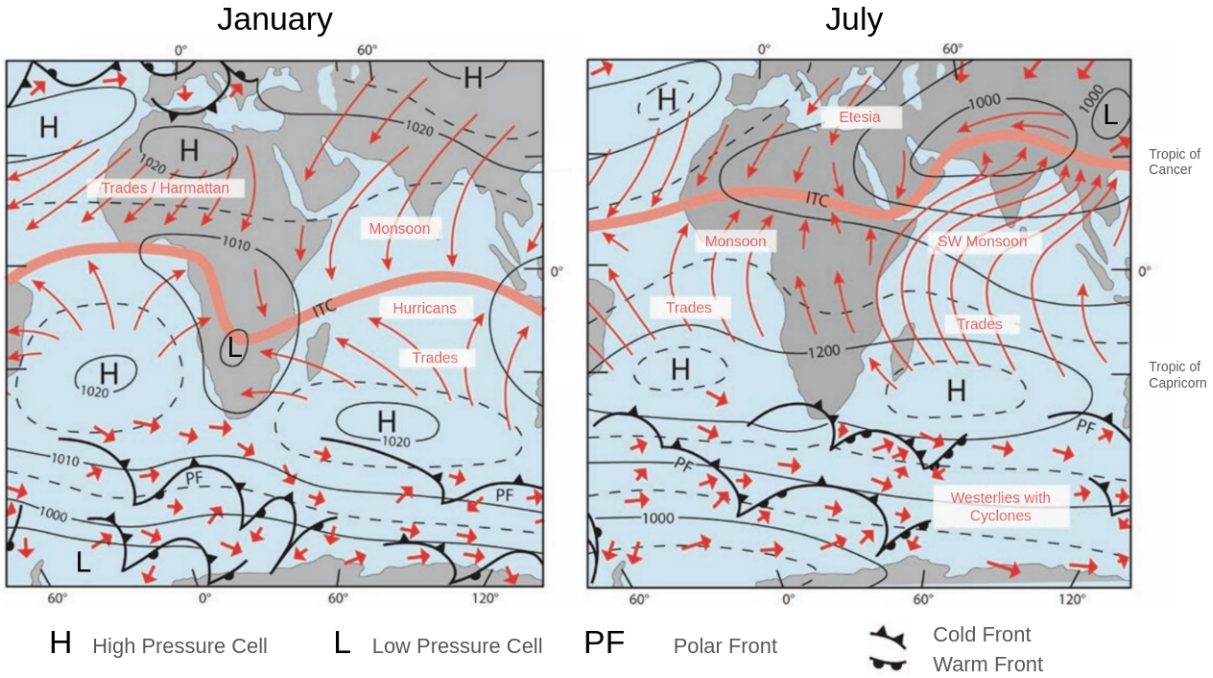


Figure 1.1: Schematic synoptics of Africa in austral summer (left) and winter (right), translated from Baumhauer, 2023, original figure from Weischet and Endlicher, 2000.

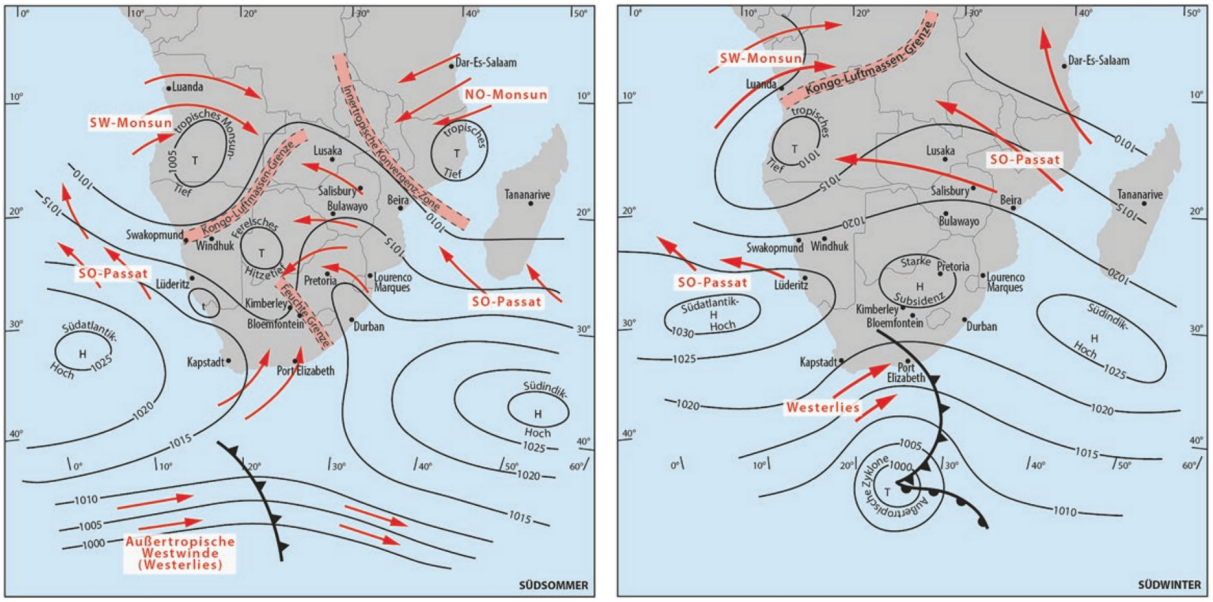


Figure 1.2: Schematic synoptics of Southern Africa in austral summer (left) and winter (right), reprinted from Baumhauer, 2023, original figure from Weischet and Endlicher, 2000. H = high-pressure cell, T = low-pressure cell.

with the annual movement of the ITC. Dry tropical air also comes down between the ITC and the mid-latitudes, warming on its descent and sitting on the moister and cooler trade winds, thus installing stability (Baumhauer, 2023). Especially in front of the Namibian coast, the warm and dry air coming down within the SAH overlays the cold and moist marine boundary layer with a marked temperature inversion. The inversion is situated at a shallow height of about 500 m and thus prevents convective ascent that could lead to rain.

Equally important, but of thermal nature, is the winterly thermal high over the south-east African plateaus that is in summer replaced by an extension of the Indic High (Continental High, CH) (fig. **1.2**,Weischet and Endlicher, 2000). The eastward-traveling anticyclones from the high latitudes can merge with the CH (Logan, 1960). The cold airmass is then further distributed by the resulting high-pressure cell. In austral summer, the CH turns into a thermal low at the ground because of the heating of the land surface and pulls the ITC southward.

Over the ocean, the wind is driven by the SAH and blows mostly from SSE parallel to the Namibian coast (Logan, 1960). The coastal area is dominated by a local wind system driven by temperature differences between the ocean, flat plain, and mountain surface (Weischet and Endlicher, 2000). During daytime, the land surface is quickly heated by the sun while the ocean surface heats up slower and is also cool from upwelling arctic waters of the Benguela current. This gradient generates a moderate SW sea breeze of 5 - 10 m/s (Seely and Henschel, 1998) which brings moist cool air from the ocean to the land. It begins at the coast during the late morning and occurs later and weaker inland.

Especially in austral summer (December, January, February, DJF), a second temperature gradient develops between the hot mountains of the Great Escarpment and the cooler flat plain and causes a strong NW wind of 10 - 15 m/s to start blowing in the late afternoon (Seely and Henschel, 1998). The wind keeps blowing until around midnight and overlasts the sea breeze which fades off at dusk. With the afternoon NW wind, fog or low stratus clouds from the ocean creep onto the land (Andersen et al., 2019) and hinder nocturnal radiative cooling. The bare inland and mountain regions cool down faster, thus reversing the temperature gradient of the day, especially in austral winter (June, July, August, JJA). The resulting SE wind blows with 5 - 10 m/s from the Escarpment down to the sea, increasing until dawn. Also in winter, strong, warm, and dry bergwinds can occur, blowing from the Escarpment (Seely and Henschel, 1998) and increasing temperature on the ground by up to 10 K (Weischet and Endlicher, 2000). This situation occurs when the westerlies swing out poleward, giving place to the SAH ridging zonally over the whole tip of Africa and shoveling moist air from the Indian Ocean over the subcontinent to a low over Angola (Weischet and Endlicher, 2000).

The local wind system is overlain by the trade winds of the global Hadley circulation. They blow from the subtropical high-pressure belt to the ITC and are deflected eastward

by the Coriolis force over the continent. When going over the Great Escarpment to the ocean, these winds subside because of divergence and reinforce the shallow inversion towards the moist and cold marine boundary layer.

1.3 Current knowledge about fog formation and occurrence in the Namib

Recent studies make use of satellite data to get a grip on meteorological phenomena in the Namib. Because the cloud base cannot be determined accurately from space and because the exact same cloud can occur as a cloud at the coast and as fog further inland, the category of FLC or FLS is used, bundling fog and low clouds. The coastal fog is relatively rare compared to the low cloud reaching further inland (Seely and Henschel, 1998), so findings concerning FLC/FLS can - with care - be attributed to low clouds. Most knowledge about Namib fog comes from the station network of the SASSCAL initiative (fig. **1.3**).

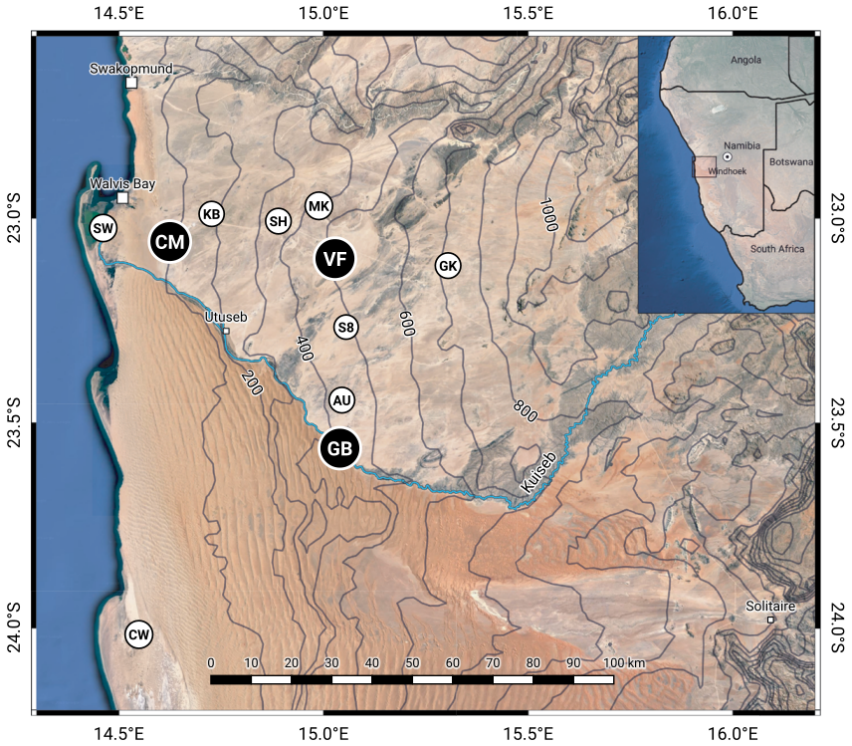


Figure 1.3: Station network in the central Namib, reprinted from Spirig et al., 2019. Black stations were intensively measured during September 2017 by Spirig et al., 2019. CM = Coastal Met, VF = Vogelfederberg, GB = Gobabeb.

1.3.1 Fog types

The cold upwelling cells of the Benguela current at the western African coast are a quasi-permanent source of a low stratus cloud deck over the Atlantic. At the same time, the cold ocean surface and the shallow inversion above it prevent the formation of convective clouds, making rain events extremely rare in the Namib Desert (Logan, 1960).

The ocean cloud deck is driven inland by winds and is recorded as fog where it touches the land and hence reduces ground visibility. Mainly in the months of April, May, June (AMJ) the stratus formed over the ocean is low with a base height of below 200 m.s.l. and intercepts the land already at the coast (Andersen et al., 2019). It is transported inland by the SW sea breeze which starts blowing in the afternoon (Seely and Henschel, 1998, Eckardt et al., 2013). The stratus occurs as fog at the coast on more than 100 days per year, reaching up to 15 km inland (Seely and Henschel, 1998). In an intensive measuring campaign at Coastal Met station (CM) during September 2017, the nighttime wind direction during fog events was found to be predominantly N to NNE (Spirig et al., 2019). This fog type is called 'coastal fog', 'advected coastal stratus' or 'advection fog'. In the following, it will be referred to as 'fog'.

Most of the year, from July to March, the stratus is higher, at 300 - 500 m.s.l (cloud base height) and has to travel about 20 - 60 km to intercept the rising land, where it is recorded as fog. Because of the travel time, the interception starts later in the evening at about 18 - 22 UTC (4 - 8 PM local time) and the stratus is dissolved in the early morning (Andersen et al., 2019). On rare occasions, the stratus even reaches the Great Escarpment (less than 10 days per year (Eckardt et al., 2013)). Annual fog day frequencies (FDF) from different stations in the central Namib indicate that high fog is more frequent than coastal fog, as the FDF increases until about 20 - 60 km distance from the coast and then declines (Lancaster et al., 1984). This type of inland fog is associated with NNE winds at the surface which was explained as follows by Seely and Henschel, 1998: The cloud deck is driven inland by NW winds late in the night and mixes with downward easterly mountain-plain winds from the Escarpment. It cannot escape upwards because of the inversion layer. The resulting mixed foggy air at the ground moves NNE. The nocturnal NE to NNE wind direction was confirmed in September 2017 at Gobabeb station (Seely and Henschel, 1998, GB in fig. **1.3**). The phenomenon is called 'high fog' or 'cloud interception fog' and later in this work 'low clouds' (adopting the perspective of a coastal station). The peak months of low clouds, March and September, match the coexistence of the two responsible winds, NW plain-mountain wind and SE mountain-plain wind, as one dominates in summer and the other in winter, respectively. Satellite imagery revealed a peak season in September, October, November (SON) (Andersen et al., 2019).

Besides high fog and coastal fog, there are the two occasionally occurring fog types radiation fog and frontal fog. Radiation fog forms in place over cooling surfaces and is

mostly found in valleys. Though it is fed by moist coastal air mixing with easterly cold air from the mountains, it is not termed 'advected' because the coastal air is clear and condensation takes place in situ (Spirig, 2022). The last type, frontal fog with drizzle, is associated with the passage of cold fronts in winter (Eckardt et al., 2013, Seely and Henschel, 1998).

The debate about the origin and formation of Namib fog has been reawakened by a study classifying more than half of the fog events as radiation fog (Kaseke et al., 2017). The number comes from categorising mixed fog as radiation fog, with mixed fog being Ocean water vapour mixed with local moist air and condensated because of radiative cooling. Going with the fog type definition of the study, at Gobabeb station which lies at the edge of the Namib fog zone near Kuiseb river (GB in fig. **1.3**), only 40 % of the fog events were ocean water-derived according to isotope analysis. The authors suggest that the advective fog zone might be retreating towards the coast and that rivers and groundwater, brought up by vegetation, contribute to the majority of fog events at this distance from the coast (Kaseke et al., 2017). However, the crucial fog probes were collected at 8 AM local time when the fog might already be dissolving and altering its isotope ratio, as was pointed out by Spirig, 2022. Further isotope analyses indicate that Namib fog can be of mixed type especially from 30 - 50 km inland on, where advected moisture mixes with local moisture (Kaseke et al., 2018). The isotope method yielded plausible fog types compared to meteorological conditions during the fog events (Kaseke and Wang, 2022). Recent findings from the gravel plains, as well as from the dunes near Gobabeb, also indicate that water vapour from deeper soil layers adds to fog formation (Adhikari and Wang, 2020). Most studies, however, agree that these two fog types are rare and that most fog events are of advective nature (Seely and Henschel, 1998, Eckardt et al., 2013). The evidence comes from meteorological understanding (e.g. Seely and Henschel, 1998) and satellite imagery (e.g. Andersen et al., 2019). Also, marine bacteria were found in fog droplets as far as 50 km from the coast (Evans et al., 2019). In line with most of the literature and because satellite imagery shows the advection of the cloud deck in most fog cases, this study focuses on fog and low clouds. While most studies agree on the nature of the Namib fog being advective, the drivers causing the cloud deck to come in as fog or as low cloud remain unclear.

1.3.2 Potential drivers of coastal fog and low clouds

The seasonalities of low clouds and fog suggest a relation to drivers that expose similar seasonal patterns. For fog, the drivers should match the peak in AMJ, for low clouds in SON. As Seely and Henschel, 1998 points out for low clouds, the peak occurrence can also be matched with the co-existence of two necessary drivers, like the NW plain-mountain wind and SE mountain-plain wind. Because of the distinct seasonality of both fog types,

not only the local wind system (which underlies seasonality as well) but also the synoptic-scale weather situation could be relevant. For the joint category of FLC the linkage to synoptic patterns has already been shown with satellite data (Andersen et al., 2020).

Based on the qualitative explanation of fog occurrence, drivers on different scales are necessary to allow for fog and low cloud occurrence. The large-scale prerequisites are: the high-pressure cells of the subtropical high-pressure belt, sending out trade winds; Cold sea surface temperatures at the coast to establish a cold marine boundary layer with a shallow inversion above it; And on a more local scale winds to blow in the stratus deck. However, the specific necessary interactions between these drivers, leading to fog or to low clouds, are poorly understood and beg further study. To date, it is not even known if fog and low clouds are two distinct phenomena triggered by different meteorological conditions.

The extent of the Benguela upwelling cell was a significant indicator of fog events in Lüderitz, a coastal city in the South of the Namib (Olivier and Stockton, 1989). Based on former studies, expert knowledge and their investigations, these authors explain the formation of fog for Lüderitz as follows: Warm oceanic air is transported by onshore winds over the cold upwelling strip at the coast, where the stratus forms. For condensation, the initial air temperature has to be higher than the sea surface temperature. The onshore wind may come from several sources: On 89 % of all fog days in 1984 it came from a coastal low south of Lüderitz and if the upwelling strip is too broad, the low only reaches over cold water and no clouds are formed. Coastal lows are also linked to temperature inversion layers that are necessary to prevent the formed stratus from rising. The onshore wind may also be associated with the (SAH), with an approaching cold front or it may be a sea breeze in the afternoon resulting from the local land-ocean wind system. The cold upwelling itself is driven by offshore winds that push the surface water away from the coast, making the cold water from below come up. The easterly surface wind may occur as the rotated southerly geostrophic wind resulting from the large-scale pressure gradient between the continental low and the oceanic high.

1.3.3 Fog retrieval from space

From the satellite perspective, the topmost cloud can, depending on its opacity, hide everything underneath, which is why cloudy areas have to be cut out from the image for many research purposes dealing with ground features. Remote sensing research in some areas of the world is thus limited, and refining cloud masking algorithms is a flowering field of research (Skakun et al., 2022). Fog and low clouds (FLC) can hide under higher clouds and are difficult to identify from space by passive sensors. About 30 % of FLC situations over Europe were not detected from geostationary passive satellite images compared to a space-based active LIDAR sensor (Cermak, 2018). But also in a study distinguishing fog

and low clouds in the Namib with active space-based LIDAR (Qiao et al., 2022), imagery with middle or high clouds was excluded as one of the first steps prior to any analysis.

Assuming that the low cloud is not hidden, the passive satellites still peek on it from above and can often only see the cloud top. The cloud top height can be determined from space and gives an estimation of the vertical position of the cloud (Cermak, 2006). However, low stratus clouds and fog are often treated as one category named broadly 'fog' or more accurately 'fog and low clouds' (FLC) or 'fog and low stratus' (FLS) (Cermak and Bendix, 2011).

There are several approaches for distinguishing between low clouds and ground fog (Lakra and Avishek, 2022). The oldest and most obvious approach is to estimate the cloud base height by subtracting the thickness of the cloud from the cloud top height. The thickness can be estimated roughly by parameterisations of e.g. the brightness–temperature difference (Ellrod, 1995) or liquid water path (Hutchison, 2002). Assuming an adiabatic cloud profile, the thickness is related to the root of the liquid water path, but a feasibility study indicates that this assumption may not hold true for low stratus (Bendix et al., 2005). Pseudosounding gives insights into vertical structures in cloud optical depth, which have to be linked to an assumed cloud profile by radiative transfer (Schüller et al., 2005). For low stratus clouds, the liquid water distribution within the cloud can be optimised with a microphysical model to match the liquid water content gained from satellite imagery (Cermak and Bendix, 2011). With that approach, the cloud base could be estimated with a hit rate of 81 % and a threat score of 62 %.

Compared to these passive remote sensing approaches, active sensors offer detailed information about the vertical atmospheric structure (Cermak, 2018). LIDAR data from the Cloud-Aerosol Lidar and Infrared Pathfinder Satellite Observations (CALIPSO) was recently used to distinguish clear, fog and low cloud situations at nighttime in the Namib (Qiao et al., 2022). The authors chose a lower detection limit of 30 m which is similar to ceilometer limits. Fog events showed increased backscatter within 30 - 120 m above ground with a correlation to fog deposition in 30 - 60 m, and low clouds were linked to higher backscatter in 100 - 200 m. The polar orbiter crosses the Namib every 16 days and can therefore not yield a temporally continuous time series of fog and low cloud occurrence. It is therefore not suitable to gain information on the mechanisms producing fog and low clouds in the Namib. There is, however, a great potential to combine lidar-derived fog / low cloud information with passive sensors mounted on satellites with higher temporal coverage. Nevertheless this study exposes the current inability to produce temporally full-coverage datasets of Namib fog and low cloud occurrence from satellite imagery.

Besides the detection of fog and low clouds from satellite data alone, additional meteorological data can support the classification (Lakra and Avishek, 2022). A global-coverage, since 1940 temporally continuous dataset is the ERA5, the 5th generation reanalysis product of the European Centre for Medium-Range Weather Forecasts (ECMWF). This data

has been used successfully to explore fog formation in Casablanca (Bari et al., 2015) and to determine atmospheric stability and aerosol interactions with the ocean cloud deck in front of the Northern Namibian coast (Andersen and Cermak, 2015).

Based on the identified research gaps the question emerges how to fill them. A better understanding of the mechanisms leading to either fog or low clouds would most likely inspire the debate on how to map fog occurrence and deduce it from satellite imagery. The relevant fog and low cloud drivers could be used to refine the umbrella category 'FLC' on satellite products and specify it as fog or low cloud. To link the drivers and initial satellite imagery to a specified classification outcome, statistical models are needed.

1.4 The potential of machine learning

Machine learning offers the advantage of taking potentially meaningful variables as input and, given the true outcome values for these variables, automatically deducing links between input and output. In simple models like logistic regression (logreg), these links are transparent and easily understandable for the human user, yet the model skill may be unsatisfying for complex problems. With increasing complexity of the algorithm, the skill may increase but the links may not be transparent to the human user anymore. A well-established algorithm of medium complexity is Random Forest, which has been applied successfully to classification tasks in many research fields (e.g. Talukdar et al., 2020).

Deep learning, as a subfield of machine learning, has shown the ability to outperform former approaches like decision trees-based learning (like random forest (RF)), especially in perceptional tasks like image recognition that are not easy to describe mathematically. Their learned strategies to solve a classification task, is, though, not easily backtracked so that they are sometimes called a 'black box'. As processes in the atmosphere are often very complex and rely on patterns on different temporal and spatial scales, deep learning has seen a lot of successful applications in the field of atmospheric science (X. Wang et al., 2023, Ait Ouadil et al., 2023).

In the case of fog, deep learning has mostly served for near-realtime fog prediction (nowcasting) in air traffic (Lakra and Avishek, 2022). For example, Convolutional Neural Networks (CNNs) outperformed RF in predicting fog in the next hour from images of the foregoing hour at Daesan Port, South Korea (J. Park et al., 2022). Fog prediction from time series of meteorological variables was implemented with an artificial neural network (ANN) for Canberra Airport, California (Fabbian et al., 2007). A global retrieval of marine liquid water clouds has been done with an ANN using satellite data and reanalysis (Andersen et al., 2017). The ANN was trained on different regions and outperformed bivariate statistics by far and multivariate statistics only slightly and in specific contexts.

With statistical models, the known conceptual relationships can be quantified and assessed in their importance for the process. This is most easily done in the logreg and

is also feasible within a RF model. Even in deep learning the learned patterns can be visualised. Thus, there can be a benefit from modeling for the conceptual understanding of a problem.

Based on the identification of relevant drivers, fog and low clouds in the Namib could be modeled by statistical models. The applicability of different model types and the suitable level of complexity for this case is, however, currently unknown.

1.5 Summary of knowledge gaps, hypotheses, and outline of thesis

From the given literature review on fog and low clouds in the Namib, it becomes clear that we still lack detailed knowledge on the drivers of the one or the other situation. Both the large-scale synoptics (Andersen et al., 2020) and the local wind system (Seely and Henschel, 1998) have been linked to fog occurrence. Our understanding remains fragmented though because either only low clouds are focused (Seely and Henschel, 1998) or fog and low clouds are treated as one group compared to clear days (Andersen et al., 2020, Veloso et al., 2024). Some research has also been done on single drivers like the sea surface temperature (Olivier and Stockton, 1989), but a general concept of what meteorological situation leads to fog versus low clouds is lacking.

To date, there is no satisfying routine for the Namib to classify fog occurrence from satellite and distinguish it from low clouds. The existing approaches are either inaccurate or require different datasets for night and day or suffer from other inconveniences. Machine learning methods have been proven extremely powerful in solving classification tasks. With a deepened understanding of the meteorological mechanisms leading to fog or low clouds, the relevant drivers can be used to train a statistical model to predict fog or low cloud occurrence. Deep learning as a subfield of machine learning has been shown to outperform simpler models in some perceptual tasks and can therefore be expected to excel in fog and low cloud prediction.

Hence I propose the following hypotheses:

1. Fog and low clouds are characterised by different patterns of meteorological drivers.

2. These meteorological drivers can be used to distinguish fog and low clouds
   2.1 with machine learning approaches
   2.2 and even better with Deep Learning.

To treat the hypotheses, Coastal Met station (CM) at the coast of the central Namib is chosen as a suitable ground truth source. It offers fog records on the ground as well as eventual ceilometer data from former campaigns (Spirig et al., 2019). For H1, median composites of several ERA5 variables will be examined for relevant meteorological

elements. These elements will then be used to train three different machine learning algorithms on classifying fog and low clouds to address H2.

Chapter 2

Data

In this section the origin and structure of the raw data is explained, followed by a description of how the datasets for hypothesis 1 are derived.

2.1 Data sources

Three raw data sources have been used: the satellite-derived fog and low cloud product over the central Namib (FLC product), the fog measurements and ceilometer data from Coastal Met weather station at the central Namib coast, and several meteorological variables of the ERA5 reanalysis product.

2.1.1 FLC product

The fog and low cloud (FLC) product is a raster of low cloud cover over the Namib region. is derived from the infrared bands of the geostationary satellite SEVIRI (Spinning Enhanced Visible and Infrared Imager). FLC are detected with a probability of 94 % and a false-alarm rate of 12 %, which was assessed using station data (Andersen and Cermak, 2018). SEVIRI has a nadir resolution of 3 km. The FLC product covers the whole Namib. I used the FLC product from 01.01.2004 - 31.12.2020 with a temporal resolution of 15 minutes. The values are empty in 2005-12, 2006-11, 2007-03 and 2007-09 due to satellite outage.

The algorithm applies spectral thresholds to the satellite image in order to exclude high ice clouds. It then compares the remaining land surface and low clouds to a monthly and an annual composite representing cloud-free conditions. The resulting FLC pixels are checked for context-plausibility by an iterative counting of surrounding FLC pixels (Andersen and Cermak, 2018).

2.1.2 Coastal Met station

The Coastal Met weather station is located at 14.62595 E, 23.05631 S near Walvis Bay in the Namib Desert (see 1.3). The surrounding landscape consists of *Arthraerua*-dominated fog belt shrublands with lichen fields further inland and *Stipagrostis sabulicola*-Cladoraphis-mobile dune grasslands further south.

It is one of 11 FogNet stations and belongs to the Southern African Science Service Centre for Climate Change and Adaptive Land Management (SASSCAL) initiative. Within the station network, Coastal Met is one of the nearest stations to the ocean with a distance to the coast of about 15 km and an elevation of 94 m msl (Spirig et al., 2019). The meteorological variables are recorded every minute. The relevant variable for this study is the fog precipitation, which is measured by Juvik fog collectors (detailed information on the instrument see Juvik and Nullet, 1995). Fog water is collected by a cylindrical screen and drips onto a tipping gauge that records 0.1 mm per received drop. The exact moment of when the drop falls is somewhat arbitrary so the measurement is suitable to qualitatively state a fog event rather than quantitatively determine the exact fog duration or deduce fog properties. I use station data from 01.07.2014 to 31.12.2020. Additional ceilometer data is available for Coastal Met station from September 2017 to February 2018. The ceilometer is an active lidar instrument that measures single-wavelength backscatter. It can thereby detect the cloud base height (CBH) of one or several clouds, depending on their transparency. The instrument at CM is a CL31-2 from Vaisala (Wiegner et al., 2019).

2.1.3 ERA5 reanalysis

The ERA5 is a global coverage reanalysis product with hourly meteorological variables from 1940 to present and a spatial resolution of 0.25 degrees (Hersbach et al., 2020). Multiple data sources like station data, radiosondes and satellite imagery are jointly used to model the weather of the past. The product is maintained by the European Centre for Medium-Range Weather Forecasts (ECMWF), an intergovernmental organisation supported by 35 states. The known issues of ERA5 are not of concern in the Namib Desert, apart from potential jumps in boundary layer wind speed at the transition from one to the other 12 h assimilation window (Hersbach et al., 2020). The ERA5 variables are available either on pressure levels with 25 hPa resolution (50 hPa above 750) or on single level, like the mean sea level pressure (MSLP). In this study, the variables temperature (t), relative humidity (RH), specific humidity (q), wind components directing east (u), north (v) and up (w), geopotential (z) on pressure levels up to 500 hPa were downloaded, as well as mean sea level pressure (MSLP) and sea surface temperature (SST). The ERA5 data can be downloaded in NetCDF format and was handled using the xarray python library (Hoyer and Hamman, 2017).

2.2 Derived data

The general workflow to prepare fog and low cloud median composites for hypothesis 1 is visualised in fig. **2.1**. To find discriminative patterns for fog and low clouds at CM, the first step is to create a dataset of fog and low cloud occurrences at the station. The FLC product is used to identify low cloud-covered nights which are then divided into fog and low cloud nights using the ground observations of fog at the station. For these nights, ERA5 variables are downloaded and reduced to median composites on fog and low cloud days.

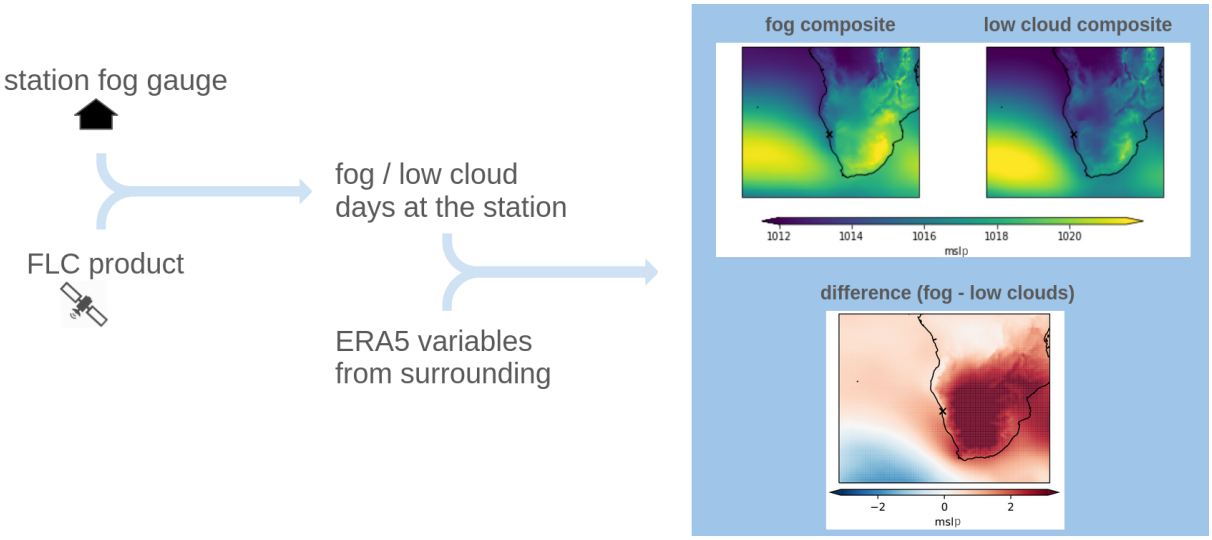
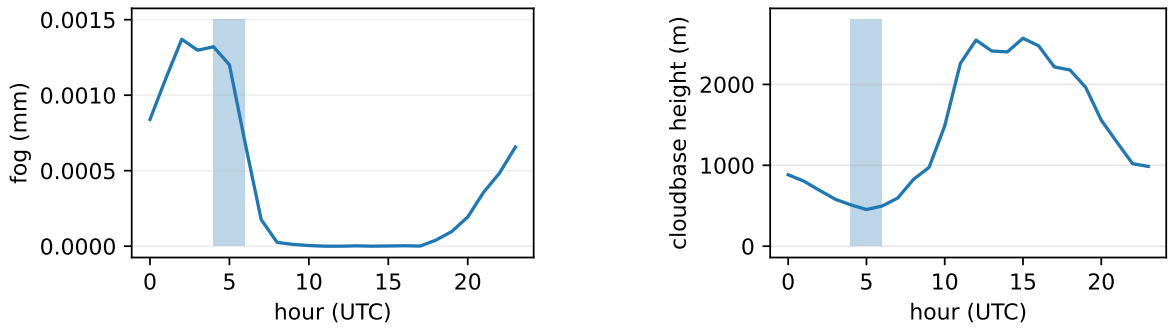


Figure 2.1: General workflow of calculating median composites of ERA5 variables on fog days and low cloud days at Coastal Met, Namibia. FLC = fog and low cloud product from Andersen and Cermak, 2018.

2.2.1 Preparing ground truth fog events

A daily time slot was chosen to select days with FLC occurrence from the FLC product. I took 4 - 6 UTC because 5 UTC was shown as the relevant time point of maximal FLC occurrence at CM station (Andersen and Cermak, 2018) and 4 - 6 is a suitable time interval symmetrically around it, capturing the fog precipitation peak (fig. **2.2a**). Additionally, the mean hourly cloud base height is lowest at 5 UTC (fig. **2.2b**), indicating the time of maximal fog occurrence compared to low cloud occurrence. That makes 9 records per day (4:00, 4:15, 4:30, ..., 5:30, 5:45, 6:00). I chose a robustness criterion of 5 FLC positive records out of the 9 to accept FLC for that time slot. This conservative criterion shall minimise the risk of including clear days. In total, 961 FLC-covered days were selected.

The station data was selected at the corresponding time window 4 - 6 UTC on FLC-positive days. In order to label each day as a fog or low cloud event, the fog days are determined using the station measurements. Of all measured variables at the station, the



(a) fog precipitation (b) cloud base height

Figure 2.2: Hourly mean of fog precipitation in drops (2014 - 2021) and cloud base height from ceilometer (2017 - 2018) at Coastal Met, Namibia.

fog precipitation is the most reliable for detecting fog events, more than net radiation, longwave downward radiation or surface temperature (Spirig et al., 2019). One drop in the fog gauge within the two-hour slot (temporal resolution of 1 minute) was considered enough to make this day a fog day. The station data contained 166 fog days that had been excluded by the conservative criterion applied to the FLC product. These days were added to the fog/low cloud dataset because they indicated fog within the 4 - 6 UTC time frame (fig. **2.3**). Consistency of the fog day measure was checked by calculating the sum

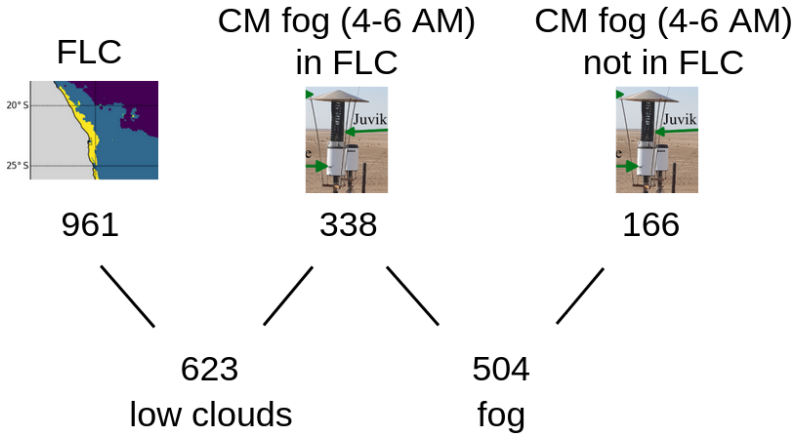


Figure 2.3: Schematic illustration of how a dataset of fog and low cloud occurrences at Coastal Met station was created. Numbers indicate total numbers of days within 2014-07-01 - 2020-12-31.

of drops at the station from 0 to 8 UTC instead of 4 - 5 UTC for each day, which revealed additional 175 missed fog nights in total but yielded the same seasonal pattern (fig. **2.4**). These fog nights were not included in the final dataset because it is not sure that there was fog in the examined 4-6 UTC slot.

The fog / non-fog data set was checked for consistency with the relative humidity measured at the station (fig. **2.5**). Relative humidity during fog conditions is usually 100 % but can be slightly under- or oversaturated due to condensation and associated

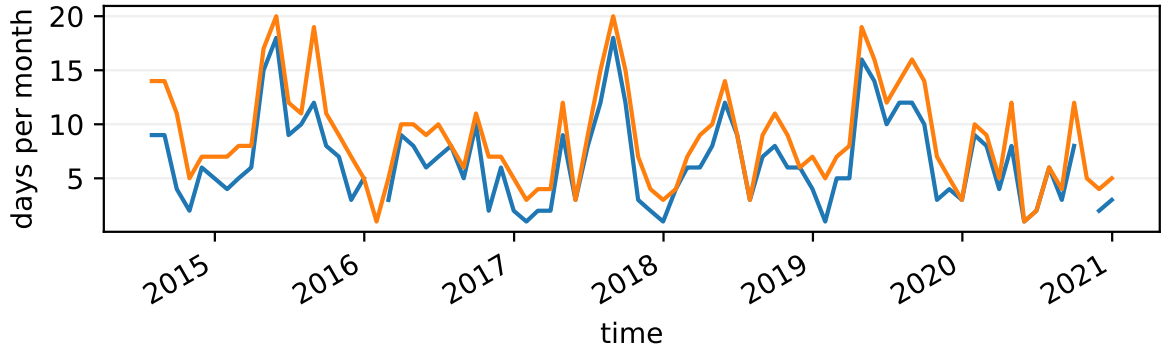


Figure 2.4: Number of days with fog recorded during 4 - 6 UTC (blue) and during 0 - 8 UTC (orange) at Coastal Met, Namibia.

warming of the surrounding air (Gerber, 1981). During the two considered hours, the relative humidity can vary, especially in cases of fog dissipation, and can cause low average humidity. The lowest average relative humidity during a fog event is 79.61 %, which was still considered plausible, so the fog determination proves reasonable and there was no need to exclude values manually.

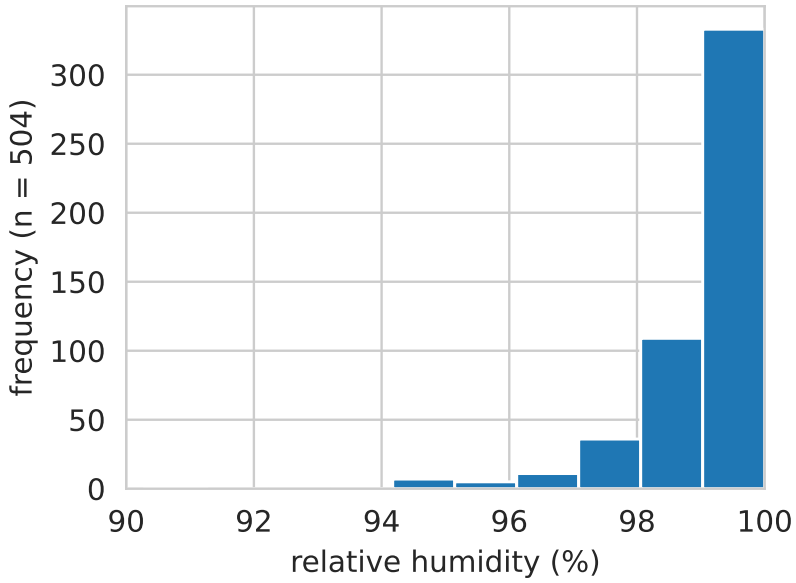


Figure 2.5: Mean relative humidity at 4 - 6 UTC during fog events at Coastal Met, Namibia.

Consistency check by fog climatology

The climatology of Namib fog events is well described (Seely and Henschel, 1998) so a good health check for FLC and station data is to compare them to climatology. The whole available time series of FLC days at Coastal Met is shown in **fig. 2.6**, displaying

the number of days where the station pixel was covered by any low cloud, may it be a cloud or fog, in the early morning from 4 - 6 UTC. A pronounced seasonality can be seen in 2014 - 2019 with a higher FLC frequency in austral summer than in early austral winter. This seasonality is less obvious but still present in the other years. For the time of available station data, the FLC days are, based on fog precipitation measurements, divided into coastal fog (338 days), and low clouds (623 days) (the latter also referred to as 'high fog' in literature), and are additionally enriched by fog days overlooked by the satellite, 166 days. Low cloud days make up the majority part (623 days) and occur mainly in austral summer, thus creating the overall seasonality in FLC days. Fog occurs less (504 days) and with a peak in early winter (**fig. 2.7**).

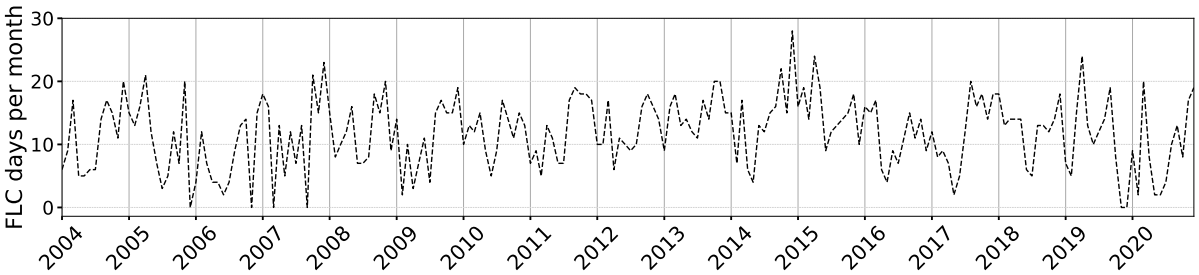


Figure 2.6: Time series of total monthly fog and low cloud (FLC) days at Coastal Met station, Namibia, as shown by the FLC product created by Andersen and Cermak, 2018.

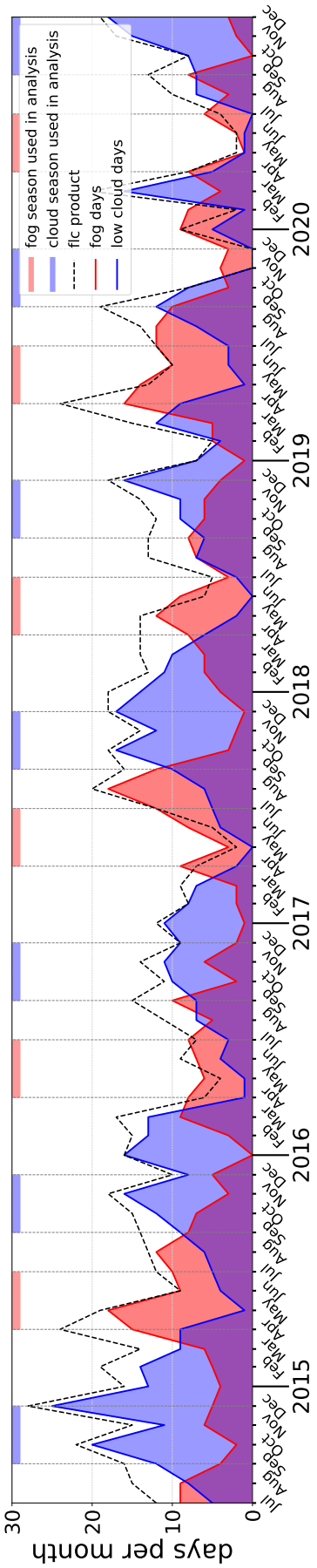


Figure 2.7: Time series of fog and low cloud days at Coastal Met, Namibia, derived from fog and low cloud (FLC) product (Andersen and Cermak, 2018) and fog measurements at the station (data cut to available station records).

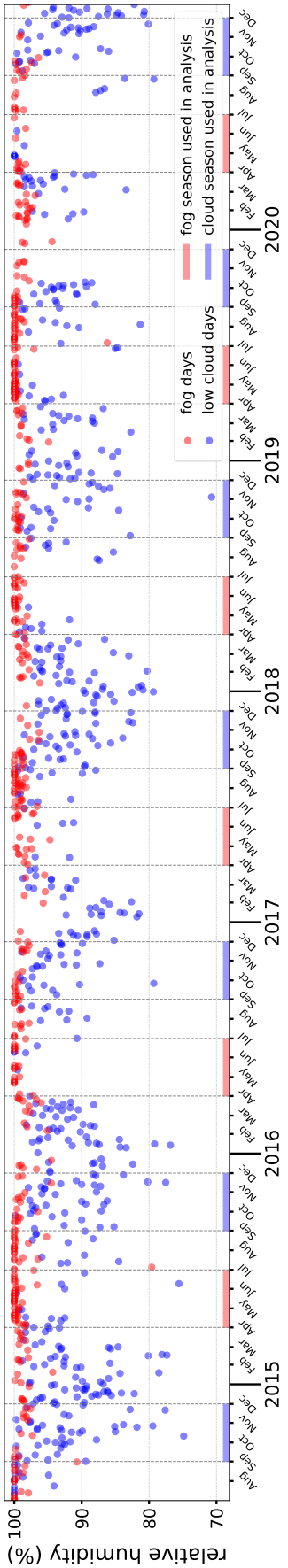


Figure 2.8: Time series of relative humidity measured at Coastal Met, Namibia.

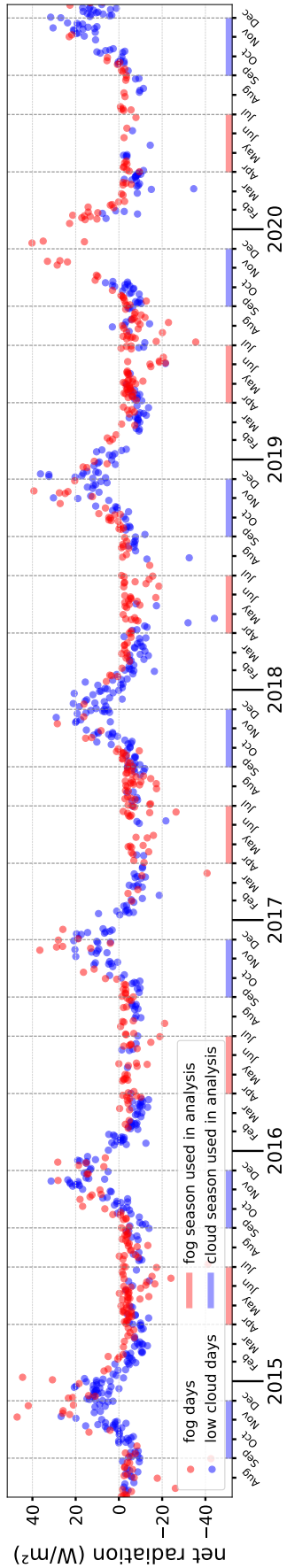


Figure 2.9: Time series of net radiation measured at Coastal Met, Namibia.

The fog day frequency (FDF) which is commonly reported in literature commonly reported, includes any type of local fog event (that is, also the rare events of radiation fog and frontal fog) at the reference location. The FDF is reported to be higher in austral winter (May - Sep) at coastal stations (capturing the coastal fog) and higher in austral summer (Aug - Oct, Mar) at inland stations (capturing low clouds) (Seely and Henschel, 1998, Lancaster et al., 1984). While not all of the low clouds at Coastal Met might travel far enough to arrive as fog at the inland stations, we assume that a majority do and that especially the FDF seasonality is transferable. Our differentiation of fog and low clouds based on station fog gauges thus yields a plausible target variable.

Based on the findings of Andersen et al., 2019, the data is split into seasons for some of the analyses. **Fig. 2.7** confirms that these seasons appropriately describe different phases of the annual cycle of fog and low cloud events at Coastal Met station. The April-May-June season (AMJ) captures the peak coastal fog months while the September-October-November season (SON) is an equally long (and thus comparable) period capturing the low cloud occurrence. The time between these two periods is summarised in two further seasons. The figure also points out that the peak occurrence of fog and low clouds can vary considerably between two years, as 2017 shows a late fog peak in August and September compared to an early 2018 peak in May and June. The usual low cloud peak in austral summer is missing in November / December 2020, which makes 2020, together with irregular fog peaks, an exceptional year.

Other variables measured at Coastal Met weather station were also analysed to confirm the reliability of the fog/cloud differentiation. **Fig. 2.8** shows that the declared fog events go with high relative humidity (RH) near the ground (98.2 % of fog days over 95 % RH) compared to low cloud days. The net radiation peaks around October to February (fig. **2.9**) due to earlier sunrise in austral summer. In austral winter, when 5 UTC (7 AM local time) is before sunrise, the net radiation during fog events tends to be zero compared to negative net radiation during low cloud events. At clear sky conditions, the surface releases longwave radiation to the atmosphere, while during fog the radiation is scattered back downward. Balanced longwave radiation during nighttime FLC occurrence was also found by Andersen and Cermak, 2018 and was used to evaluate FLC detection from satellite.

2.2.2 Preparing ERA5 data for composites

All downloaded ERA5 data was clipped spatially to show the south African subcontinent with some surrounding, and was clipped temporally to the dates with fog / low cloud occurrences at CM. The data was clipped to 5 UTC because that is the time of maximum FLC cover at the station during the daily cycle and it is the center of the time slot used for assessing the FLC dates.

The ERA5 data was split into fog days and low cloud days according to the station data. In each pixel, the median value of each variable was calculated from all fog days to get a median composite. The same was done for the low-cloud days. The median composites were produced for the whole time span as well as separately for the different seasons. The median was chosen as a safe measure for also non-gaussian distributions because it equals the mean in the case of Gaussian anyway but is also suitable to describe skewed distributions. Among the distributions of low cloud versus fog day values in each pixel and for different seasons, it is likely that not all distributions are Gaussian. This advantage was considered more important than possible distortion due to small sample size. While one could visually seek for differences between the fog and low cloud composite, it is convenient to calculate the difference by subtracting one composite from the other. The low cloud image is subtracted from the fog image, which means that high values in the resulting image mean higher variable values on fog days and lower values on low cloud days. Low differences mean lower variable values on fog days.

To further extract the difference between fog and low cloud situations, the annual variation in the data can be eliminated. Each pixel is therefore standardised with its climatological monthly mean and standard deviation of the years 2004 - 2020 (Equ. 2.1, Code-Sn. 2.1), including clear, cloudy and fog days to ensure a representation of the annual variation of the variables. The data is first standardised and then aggregated into seasons and split into the fog and low-cloud days. The standardised value x in each pixel was calculated

$$x_{standardised} = \frac{(x_i - \bar{x})}{std(x)} \tag{2.1}$$

with x_i, one pixel value within the timeseries and std, the standard deviation.

```python
import xarray as xr

def standardise_months(ds):
    climatology_mean = ds.groupby("time.month").mean("time")
    climatology_std = ds.groupby("time.month").std("time")
    anomalies = xr.apply_ufunc(
        lambda x, m, s: (x - m) / s,
        ds.groupby("time.month"),
        climatology_mean,
        climatology_std,
    )
    return anomalies
```

Listing 2.1: function to standardise each pixel per month.

Fog days occur mostly within a few months in the austral winter (AMJ with 10 - 18 fog days per month) and occur with often less than 5 days in the other months. So, when calculating a median composite of fog days over the year, the median will be driven mostly by the winter season. Because of the pronounced seasonality, it cannot

be excluded that the few fog days in summer are caused by a different constellation of pressure, temperature, etc and we miss this different pattern in the yearly median. The patterns found in the overall median can thus not help to distinguish fog and clouds in summer, which is the goal of this study, as we want to make better predictions than a mere guess from climatology. To find differences between fog and low cloud situations that persist throughout the year, separate composites for the seasons are computed.

Chapter 3

Methods

In this chapter, the methods used to tackle the two hypotheses are treated. The first hypothesis concerns the conceptual understanding of fog or low clouds occurrence at Coastal Met. In the second hypothesis, the conceptual understanding is captured in features which are used to train classification models. Figure **3.1** summarises the workflow that is explained further in the following subsections.

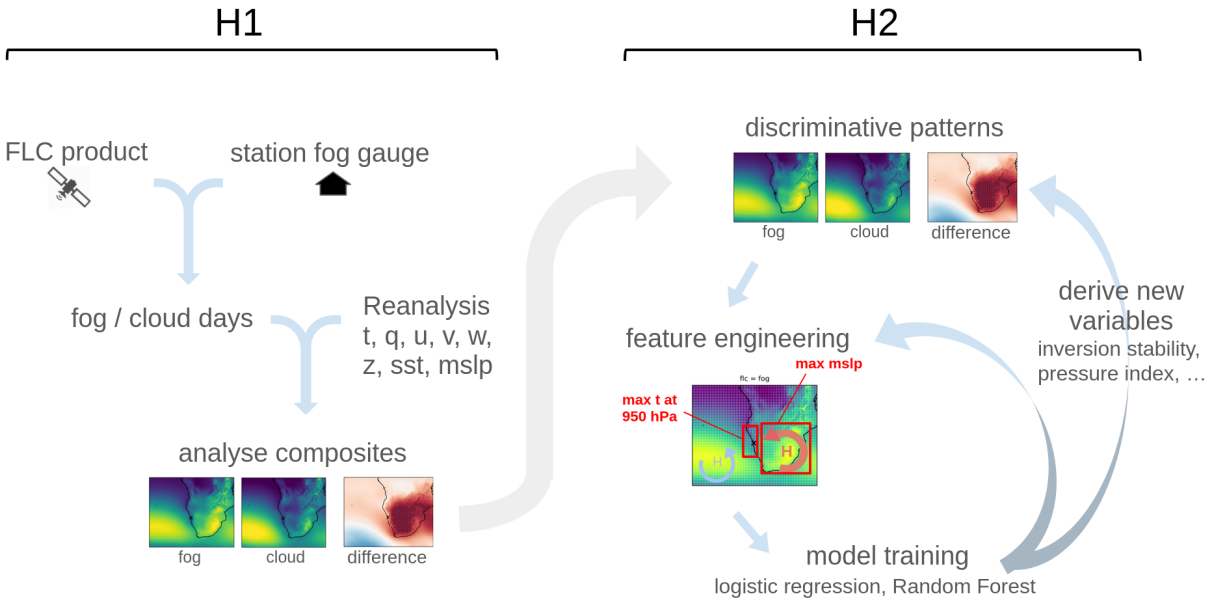


Figure 3.1: Workflow for treating hypothesis 1 and 2 (H1 and H2, respectively). FLC = fog and low cloud. cloud = low cloud. Reanalysis variables are temperature, specific humidity, wind components east, north and up, geopotential, sea surface temperature and mean sea level pressure.

3.1 H1: Composite analysis and case study

The first hypothesis is treated by visually examining the ERA5 composites (whose preparation is explained in subsection **2.2.2**) for relevant patterns. The understanding is deep-

ened by a case study of a specific fog night. To take up the research of Spirig et al., 2019, an exemplary fog event was chosen within the time frame of their intensive field campaign in September 2017. For that period, ceilometer data is available for CM that can be used to determine the cloud base height. The ceilometer backscatter reveals a comparably long interval of intensive near-ground backscatter during the night 22./23.09.2017 (fig. **3.2**), which was chosen for the case study.

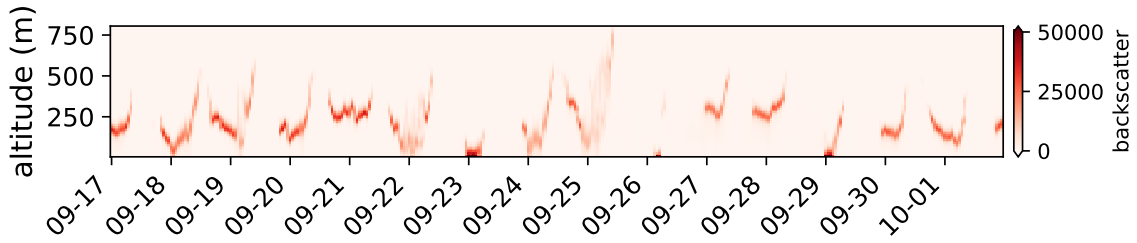


Figure 3.2: Ceilometer backscatter indicating stratus elevation at Coastal Met, Namibia, in September 2017. The fog event in the early morning of 2017-09-23 is chosen for a case study.

3.2 H2: Classification

Hypothesis 2 concerns the ability of machine learning models to predict a given FLC event as fog or low cloud. For that purpose, three model types along a gradient of rising model complexity are trained on the fog and low cloud data. The first, most simple classification model is the logistic regression (M1). The second is the ensemble classifier Random Forest (M2). The third, 'black box' model is a Convolutional Neural Network (M3).

3.2.1 Feature engineering

Both logistic regression (logreg) and Random Forest (RF) can handle many but 1-dimensional predictor variables. That means that in our context of spatial predictors, like pressure patterns, the map (= matrix) of the pressure cannot be given to the model as one predictor, though the spatial pattern may contain the relevant information on fog or low cloud occurrence.

Deploying the deepened understanding of fog and low cloud drivers gained in H1, the spatial patterns are in this study reduced to single meaningful values, e.g. by averaging over a certain region. This approach of capturing complex information in one value is called feature engineering. The features are used as input to the statistical models. Alternatively, spatial patterns can be fed to statistical models by entering each pixel as an independent variable. This was done with MSLP by Andersen et al., 2020 to classify

FLC versus clear days over the central Namib and yielded a hit rate of 94 % and 17 % false alarm rate.

The inversion limiting the stratus vertical height can be captured with a stability measure. The lower tropospheric stability (LTS) expresses the inversion strength between the 700 hPa level and the surface by comparing the potential temperatures of the two levels (Klein and Hartmann, 1993). Based on the findings in 4.1.1, I use the 925 hPa level to capture the inversion at the Namibian coast and call it Namib Lower Tropospheric Stability (NLTS):

$$NLTS = \Theta_{925\,\mathrm{hPa}} - \Theta_{surface} \tag{3.1}$$

with Θ being the potential temperature.

A more elaborate measure of inversion strength is the profile-based estimated inversion strength EISp, which has been shown to be a better proxy for low cloud cover than LTS and EIS (Z. Wang et al., 2023). Within the bottom 5 km of the atmosphere, the layer with the strongest inversion (m) and its lower and upper neighbour (base and top) are used in the calculation

$$EISp = \Theta_{top} - \Theta_{base} - \Gamma_m(z_{top} - z_{base}) \tag{3.2}$$

where Θ is the potential temperature and z is the geopotential height. Though it is recommended to calculate EISp with the full available pressure levels of ERA5 data (up to 1 hPa), the inversion at the Namibian coast is located below 800 hPa so I refrained from downloading all pressure levels but used 500 hPa downwards.

The synoptic situation, mainly characterised by the SAH and CH, influences FLC occurrence (Andersen et al., 2020). To get a grip on the position of each high-pressure cell, the High Pressure Index (HPI) was calculated, which is the mean longitude of a pressure cell weighted by the pressure values. This index was successfully applied by Jia et al., 2015 to characterise the Siberian High in relation to Atmospheric Optical Depth (AOD) and is formulated

$$HPI = \frac{\sum P_i \cdot L_i}{\sum P_i} \tag{3.3}$$

with P being the pressure and L being the longitude at each gridpoint i within a defined area of $P > p_{border}$. The border pressure delineating the high has to be chosen specifically for the high of interest.

A suitable isobar for the Siberian High was determined to be 1023 in Jia et al., 2015 but no suitable isobar has yet been determined for the SAH or CH. A low isobar includes too much area while a high isobar may be missing on some days. Based on visual expectation of plotted HPIs with different isobars (6.6), 1017 hPa was chosen for the SAH and 1021 hPa for the CH.

The SST has been shown to have a significant impact on fog events at Luderitz, a

coastal city south of CM (Olivier and Stockton, 1989). To emphasise the SST connection to CM, the difference between each pixel to a coastal pixel near the station was calculated. A ridge regression with all pixels as independent variables was then performed to identify areas that impact the fog occurrence at CM most. The ridge regression is a regression model where overfitting is prevented by controlling the coefficients via ridge regularisation (L2-norm). See details about regularisation in **3.2.5**. The ridge regression is adapted to classification tasks by converting the labels into -1, 1.

3.2.2 Train-test split of time series

When the model were trained on the whole available dataset, there would be no unseen data left to test if the model has learned general rules applicable to unseen data instead of learning the training data by heart (overfitting). Therefore, a small share of the data is held back from training and used as a test set. The predictions based on unseen data are then a measure for the generalisation power of the model.

The train-test split is usually done at the very beginning before preprocessing to ensure that the test set is truly unseen. The same preprocessing steps must then be carried out on the train and test set separately, however at some points with care. If the data shall be standardised during preprocessing, the mean and variance in the standardisation must be the same for the train and test set to ensure that the values that are passed to the model mean the same.

The test set must be independent of the training set and is therefore in most cases randomly drawn. With time series, a random draw would not result in an independent set, as one randomly drawn day in the test set would very likely resemble the day before or after in the training set (temporal autocorrelation) and thus be very similar to the training and easy to predict. That would lead to an overestimated, not representative test accuracy.

Time series data is therefore usually split into two blocks with a gap in between to avoid placing consecutive (and highly correlated) days into the train and test set. Intuitively the future is predicted from the past so the first big block is assigned for training and the later short block for testing. However, as we assume no trend in the data (3.3) and are interested in the basic patterns and underlying mechanisms of fog and low cloud occurrence in the Namib, the order of train and test should be ignorable.

With a pronounced fog and low cloud season, the test block has to cover all seasons to give a representative evaluation of the model performance. This is easily illustrated: If the model is biased towards fog (meaning it often confuses a cloud for a fog day) and is tested with test data from the fog season, it will score better than it actually is - while tested on cloud data it will score worse.

Bringing together the blockwise split and the requirement to cover all seasons leads

to holding back the last year of the available 6.5 years of data, which is similar to the commonly used 4/5 - 1/5 random split in other applications. Because this last year could be somehow exceptional, it would be nice to use also other years for testing. This issue of accidentally choosing an exceptional test set is addressed by cross-validation. In cross-validation, the data is divided into k folds, e.g. 5 folds. Each fold is held back as a test set once, while a model is trained on the other four folds. The mean test accuracy from the five models is more robust and representative.

As said before, the explanatory goal of this study allows for splits of the time series that do not follow the order train (past) - test (future). A cross-validation for time series is therefore applied, cutting out each whole year from the time series once, training the model on the rest and testing it on the cut-out year, in the end averaging the performances. A gap of seven days is inserted around the cut-out test year to rule out temporal correlations of the train and test set.

This split is currently not implemented in sci-kit learn (sklearn), the commonly used Python library for machine learning tasks. Available splits are the random split (k-fold split of shuffled data), k-fold split, and time-series splits that are designed for models predicting the future from the current state (see fig.3.3). The offered k-fold split splits

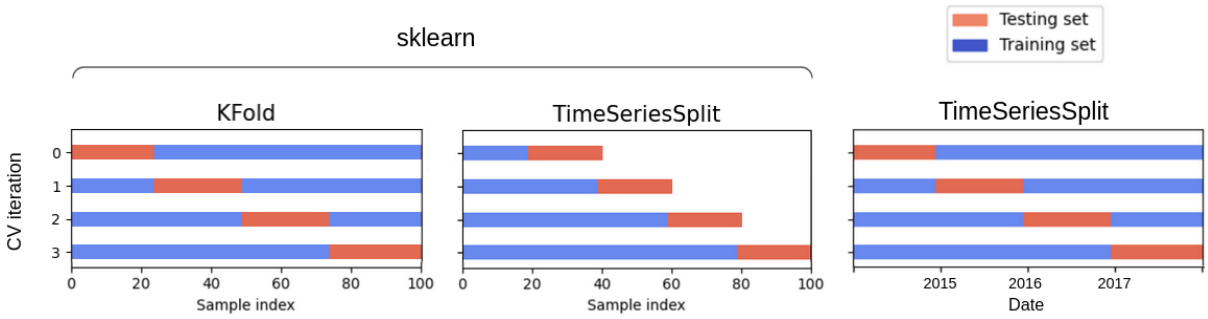


Figure 3.3: K-fold split and time series split available in sci-kit learn compared to the time series split implemented for fog and low cloud classification at Coastal Met, Namibia.

the samples by index, not by date and is therefore not capable of splitting an irregular time series in regular time intervals. With the sklearn time series split the mean test accuracy is calculated from models that were trained with different sizes of training data. The desired split is the sklearn k-fold split, but splitting by date.

The following function performs the desired split and can be wrapped in a cross-validation procedure:

```
1
2 import pandas as pd
3 def split_train_val(df, startday, space = 7):
4     ''' df is a features dataframe. startday is the startdate of the
    year to exclude as test year. space is added (and dropped from
    analysis) before and after the val set (in days). '''
5     # load sample dates
```

```
    dates = pd.read_csv('path-to-csv-with-sample-dates')
    # Days to exclude before and after test year
    excl = (dates > (startday - pd.Timedelta(space))) & (dates <
   startday+pd.Timedelta(days = 365 + space))
    # one year from startday on is for validation
    valdates = (dates > startday) & (dates < startday + pd.Timedelta(
   days = 365))
    # select the samples for a validation and training dataset
    val = df.loc[valdates]
    train = df.loc[-excl]
    return train, val
```

Listing 3.1: Function to split timeseries data into one test year (anywhere within the timeseries) and the rest (as training data) with optional space before and after the test year.

The cross-validation procedures in sklearn are implemented as classes with a method delivering the train and test set for the currently calculated model. The advantage of this architecture is that the split can be included among other steps within a pipeline. To make use of that advantage I created a resembling class for the k-fold time series split.

```
class my_cv:
def __init__(self, startdays=['2014-07-01', '2014-10-01',
    '2015-01-01', '2015-04-01', ...,  '2019-01-01']):
    self.startdays = startdays

def split(self, X, y, groups=None):
    df = X.copy()
    df['fog_01'] = y # add the labels before splitting
    for day in self.startdays:
        train, val = split_train_val(df, day, space = 7)
        # add further preprocessing here, e.g. balancing
        trainind = train.index.to_series() # get index
        valind = val.index.to_series()
        yield trainind, valind

def get_n_splits(self, X, y, groups=None):
    return len(self.startdays)
```

Listing 3.2: Cross validation (cv) class with k-fold time series split.

When the predictive ability of the model has been evaluated with test data and the model is good enough to be used for predictions, it is trained again with all available data. This procedure is based on the assumption that the final model will be as good as or better than the tested models.

3.2.3 Preprocessing

The preprocessing for machine learning includes several techniques, of which not all are required by every model. The following is an overview of the techniques, while the sections introducing the models explain which technique is used for which model.

Some models require complete input data without missing values in any predictor. This can be achieved by removing the data points with missing values, thereby losing information. The here applied option is filling the gaps by the mean of the predictor.

Different predictors usually come in different units and ranges that would make the model focus on predictors with larger values (Castillo-Botón et al., 2022). To avoid this, each predictor can be standardised by subtracting the mean and dividing by the standard deviation (sd). The standardisation conserves the distribution of the values. Some of the variation in the data comes from the seasonality which can also be removed before model training. This "deseasoning" is done by standardsising with the monthly mean and sd of all training data. The same mean and sd are used to standardise the test data of which no information must leak into the training.

In contrast to the deseasoning for the composites in **2.2.2**, clear days are not considered for calculating the mean and sd. The composites are for deducing knowledge on the mechanisms, therefore a comparison to clear days is plausible, while the deseasoning here is meant to crystallise differences between fog and low cloud days and make them easily digestible for the models.

3.2.4 M1: logistic regression

On a gradient of transparency and complexity, the most simple classification algorithm for a binary classification into fog and low clouds is the logistic regression (logreg), introduced by Cox, 1958. It resembles the linear regression but uses not a linear function of the form $y(X) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + ... + \beta_n x_n$ but a sigmoid function that yields either 0 or after a sharp rise 1 as surrogate for the one or the other class (Hosmer Jr et al., 2013). The logistic regression is fitted by maximum likelihood estimation (MLE) instead of the ordinary least squares (OLS) used in linear regression. The sigmoid function,

$$f(x) = \frac{1}{1 + e^{-x}} \tag{3.4}$$

can be fitted to the training data by MLE and its outcome can be interpreted as a class probability:

$$p(X) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 x_1 + \beta_1 x_1 + ... + \beta_n x_n)}} \tag{3.5}$$

3.2.5 Regularisation

When the data is noisy and there are many explaining variables compared to sample size, the basic form of the logistic regression is prone to overfit. That means, the coefficients are closely fit to the data (including noise) and thus adapt a wide range of values. To fight the meaningless heterogeneity of the coefficients, regularisation techniques can be used and are implemented by default in the logistic regression classifier in sklearn (Pedregosa et al., 2011). A regularisation is done by a penalty term that is added to the basic formulation in Eq. 3.5. The penalty term contains the sum of the coefficients and, during optimisation, forces the model to keep the sum of the coefficients small. With the absolute sum of the coefficients (L1-regularisation, "lasso", Hoerl and Kennard, 1970), the model is forced to only include the most important predictors, setting the coefficients of all other predictors to 0. With the sum of squares (L2-regression, "ridge", Tibshirani, 1996), the coefficients β are forced to a low range, thus yielding a balanced set of small coefficients:

$$\lambda \sum_{i=1}^{p} \beta_i^2 \tag{3.6}$$

The weight of the penalty λ determines how rigorous the model is forced to ignore noise in the data. A suitable value can be found automatically by bootstrapping. While the L1 regularisation discards all but one from a group of correlated predictors and thus throws away meaningful variables, potentially puzzling the user, the L2 method keeps all predictors including the irrelevant (Akalin, 2020, Friedman et al., 2010). The elastic net method (Zou and Hastie, 2005) combines both penalty terms but is computationally more expensive. I use the default L2-regularisation within the logistic regression and in the ridge regression for the SST feature (3.2.1).

3.2.6 M2: Random Forest

Random Forest (RF) is a popular machine learning technique that ensembles decision trees (Breiman, 2001). For each of the M trees to build, a random share of size a_n is drawn from the total of the data in a bootstrap manner (Biau and Scornet, 2016). The random draw is implemented with replacement in sklearn (Pedregosa et al., 2011), offering a data point to eventually occur multiple times in the sample.

Within the tree, the sample is split multiple times and each split is called a node. The split is done using the Gini impurity on all or a subset of the features (explaining variables) and selecting a specific criterion from one feature. As a specific example for the case of fog or low cloud classification, the feature could be the mean relative humidity at the coast and the criterion could be "> 0.7". After the split, the procedure is applied again and again to the growing number of subsamples. It stops at a subsample that consists entirely of items with the same value of a feature or when only one item is left

or when all features have been used for splitting. The depth of the tree is the number of iterations of the procedure or, synonymously, the number of nodes one would have to pass to get to one bottommost node, also called a leaf node.

The higher a feature appears within the tree the more impact it has on the final decision. The converse argument then says that the position of the feature in the tree can be used as a measure of feature importance, expressing how important that feature is to decide the class of the sample. The feature importance can be used to identify relevant drivers of the studied phenomenon.

One decision tree alone can already be used as a classifier. In a random forest, the predictions of many trees are combined to a final decision. The individual trees look different because they have been built on an individual, randomly selected portion of the whole data. They may also differ when trained on the same data sample because two features can get equally ranked for one node by the Gini impurity and then one is selected randomly. In a classification task, the final decision of the ensemble was originally found by a majority vote (Breiman, 2001). However, the current implementation in sklearn (Pedregosa et al., 2011) combines the final class probabilities of the trees.

Hyperparameters are parameters that specify how the model works and therefore are not learned during training. Finding suitable hyperparameters (tuning) can improve the model performance (Castillo-Botón et al., 2022). The trees can be cut to a certain depth (pruning) to avoid intensive computations. The best number of trees compromises between better model performance with more trees and less computation costs with fewer trees. At a certain number of trees the performance increase levels off. The default depth in sklearn is 100.

3.2.7 Evaluation metrics

The metric is meant to give an impression of how well the model can predict a sample and must be chosen according to the specific task to solve. Most metrics are calculated from the confusion matrix that is in our case (declaring fog as 1 and low clouds as 0):

	fog predicted	low cloud predicted
fog labeled	true positives (TP)	false negatives (FN)
low cloud labeled	false positives (FP)	true negatives (TN)

Table 3.1: Confusion matrix for binary classification of fog or low clouds in the Namib.

The widely used accuracy is defined as the percentage of true predicted values on all predictions: true predictions (TP + TN) / all predictions (TP + TN + FP + FN). As low clouds are more frequent than fog (higher fog day frequency 20-60 km inland than at the coast, Seely and Henschel, 1998), a high accuracy could be awarded to a model that assigns the majority class 'low cloud' to all samples and has no skill. One metric

especially designed for unbalanced data is the F1-score:

$$F1 = (2 \cdot (Precision) \cdot (Recall))/((Precision + Recall)) \tag{3.7}$$

with

$$Precision = \text{true predicted fog/all predicted fog} \tag{3.8}$$

$$Recall = \text{true predicted fog/all fog labeled} \tag{3.9}$$

Another strategy is to use rank-based metrics like the Receiver Operating Characteristic (ROC) where the True Positive Rate (= Recall) is plotted against the False Positive Rate (false predicted clouds / all labeled clouds) for different thresholds of classifying 'fog' or 'low cloud'. The area under the resulting curve (ROC AUC) is a measure of the model skill and should be maximal.

3.2.8 Recursive feature elimination

The recursive feature elimination (RFE) is a method to determine the most relevant features for the prediction task. First, the model is trained with all features. Using some measure of feature importance (e.g. the feature importance based on the Gini impurity in Random Forest), the least important feature is determined and discarded. This procedure is repeated until the desired number of most important features remains. Because of interactions between features, the most or least important feature varies depending on what features are present in the dataset. That makes RFE more than just selecting the most important features of a once-trained model.

3.2.9 M3: CNN

Analogies to image recognition

CNNs are extremely successful in image recognition tasks (Wu and Chen, 2015). These tasks were formerly solved by experts spending large amounts of time for feature engineering, to reduce spatial information on images to one-value features. The hand-crafted features are created in a try-and-error procedure and cannot be better than their designer. The big advantage of CNNs is that they learn patterns from the raw training data (end-to-end learning), thus being able to learn even patterns that are unrecognised by humans (Chollet, 2021). The patterns can be recognised independent of their location on the image which is a valuable skill for image recognition tasks but also for applications in nature science. A downside of the approach is, that the learned patterns are in some cases difficult to extract from the model and thus do not necessarily yield a deeper understanding of the modeled phenomenon.

The maps of MSLP or SST can be seen as images showing recognisable spatial patterns.

Like colour images combine three independent (not ordered) channels, three maps of meteorological variables could be combined. Thus, our classification can be redefined as classifying meteorological situations on images as fog situations or low cloud situations. For the feature engineering in **3.2.1**, the data within areas of large mean differences of fog and low cloud days are manually cut out and reduced e.g. by taking the average of that area. With a CNN the raw maps of e.g. MSLP can be used as input from which the CNN learns the relevant patterns automatically. Because CNNs recognise patterns location-invariant, the seasonality-driven varying position of e.g. the high-pressure cells would not dilute prediction skills.

Fundamentals

Deep learning models, as a subcategory of machine learning algorithms, are characterised by their structure, many stacked layers of basic blocks (a "deep" stack). Blocks are functions with parameterisable weights. The convolutional layers used in CNNs consist of a set of filters, that slide step by step over the image while performing reducing operations on the respective image part (convolution).

The weights of the layers are adjusted to the training data via optimisation. Starting with random values, the weights are updated via gradient descent on a loss function. The loss measures the difference between the true value of the sample and the value predicted by the model, and is defined by the user (e.g. cross entropy for binary classification). Instead of calculating the loss of each sample, the loss can be calculated for a batch of several samples, thereby accelerating training and avoiding large weight oscillations.

The training data is given to the model in an iterative process, i.e. one sample or batch passes the layers until a final prediction; the prediction is compared to the true value which yields a certain loss; the loss function gives the direction in which to adapt the weights of all the passed layers; the layers are adapted and the next sample or batch is taken. When the model has seen the whole training data, one epoch is over and a performance metric, e.g. accuracy or F1-score, can be calculated based on a validation set.

As each layer operates on the result of the former, the layers ground one over the other and contain hierarchical information. Bottom layers are sensitive to basic features, upper layers to complex abstract features. The bottom layers of a model can thus be reused for training the model on a completely different task (transfer learning).

Implementation

The CNN was built using the keras python library (Chollet et al., 2015) with tensor flow backend (Abadi et al., 2016). To train a network the data is split into three parts: training and validation for learning and test data to test the trained model. The training data is

used to adapt the weights in each layer. The validation data is used to calculate after each epoch how well the network performs on unseen data, i.e. its generalisation power. The metric used for training and validation data is the same (here accuracy and F1-score) but can differ from the loss (binary cross entropy). The loss function quantifies the loss in each input, the metric quantifies how well the network performs on validation or test data.

As the network optimises its weights to the training data, the loss will decrease continuously with each epoch. With enough capacity and epochs, the network can learn more and more characteristics of the individual training data, in the end learning each image by heart. Not every characteristic of every single image is helpful in predicting the target variable and the network loses generalisation power. This phenomenon is called overfitting and can be dealt with by several strategies.

The loss and the metric of validation and training data can be plotted during training. Overfitting starts in the epoch where validation loss or validation error starts to increase. It is good practice to cut the number of epochs in which the training metric still improves but the validation metric stalls or worsens (Chollet, 2021).

The best option to reduce overfitting is to add more and diverse training samples. The samples can also be preprocessed to hold only relevant information. For this purpose, the seasonality was subtracted from the data. A further option is to tune the hyperparameters of the model. A small capacity (e.g. fewer layers) forces the model to learn only the most important (general) rules to predict the target values instead of memorizing features that are characteristic of the single training images. One can start with a small net and increase the size until overfitting starts. Apart from that, L2-regularisation is a common tool to force the model to keep only relevant information.

Alternatively to a self-made model from scratch one can use the convolutional base of a pre-trained model and only train a self-made classifier on top. The convolutional base contains weights that are already trained on large samples to identify basic visual structures such as edges. These valuable weights should not be changed, so the base is frozen while the classifier is trained on the new training data. After the classifier is trained, the upper layers of the convolutional base can be unfrozen and trained together with the classifier (fine-tuning). At that point, the classifier must already be trained because random initial weights would cause large weight changes. These could harm the pre-trained base via backpropagation. For the same reason, the learning rate has to be small (Garcia-Garcia et al., 2017).

All available pre-trained networks in `keras` are trained on the ImageNet database which holds RGB images mostly of furniture and animals Chollet, 2021. As RGB images have the three channels red, green, blue, the CNNs are built to take input images with three channels. Though it is possible to adapt a CNN to other-dimensional input, for this simple potential assessment, three input variables are selected for training based

on the findings of H1. The self-made CNN is constructed to take multi-channel input. MobileNetV2 (Sandler et al., 2018) was chosen as pretrained network. Çevik, 2020 gives an overview of pre-trained CNNs and their prediction efficiency. MobileNetV2 is a lightweight network that was considered sufficient for the purpose of this study.

3.2.10 Baseline scenario

The goal of this study is to outperform fog predictions that are solely based on the well-known climatology. I thus calculate a baseline accuracy from the climatology alone as a reference for my further models. At first, the mean fog day frequency for each month is calculated from the available station data (2014-2020). These frequencies are used as fog probabilities. For each real FLC day in the station data period, a 'fog' or 'cloud' event is randomly drawn with the given fog probability of the respective month. Each real event is used as truth data and compared to the random draw to compute overall accuracy and confusion matrix elements. As the procedure contains the random draw as a stochastic element, it is repeated 1000 times to get a mean baseline accuracy.

3.3 Implicit assumptions

When taking several years of data and analysing them for basic patterns, we implicitly assume that there is no change in these patterns within the chosen time frame. In this study, we take five and a half years of data to characterise the synoptic circulation and SST in the low cloud situation versus the fog situation in Coastal Met by contrasting them. A poleward shifting of the SAH has been observed for 1980-2014 (Jury, 2018) and predicted for the future until 2100 (Jury, 2013). This was associated with a rotation to more easterly surface winds over Namibia from 1980 to 2014. However, I retaliate these linear trends by assuming that the difference between low cloud and fog situations is bigger than the temporal shift within the five and a half years.

Chapter 4

Results

To treat H1, the ERA5 data is analysed in median composites as well as in a case study to gain insights about the drivers of fog and low cloud occurrence in the central Namib, specifically at the station Coastal Met (CM). For H2, three models are trained on fog / low cloud occurrence at CM and the corresponding ERA5 data to assess the potential of machine learning for fog and low cloud differentiation.

4.1 H1: Discriminative patterns

Based on the literature, the ERA5 variables MSLP, t, q, RH, SST and wind components u, v, w were selected and investigated for discriminative patterns (find a definition of the variables in **2.1.3**). The median composites indicate general patterns related to fog and low clouds at CM, while the case study allows a detailed look at the interplay of the drivers during a specific fog night.

4.1.1 Composite analysis

The reanalysis variables MSLP, t, q, RH and wind components show distinct patterns on coastal fog days versus low cloud days. If not stated differently, patterns that are consistent over all seasons are treated. Because of the seasonal peak of fog occurrence in AMJ, the median fog composite of all data mainly depicts the situation in these peak months; while the median low cloud situation mainly shows the situation in the peak low cloud months SON. To account for seasonal differences the time series was grouped into seasons and median images of the four seasons were plotted separately. The three main identified drivers of fog and low cloud occurrence are the atmospheric inversion capping the marine boundary layer, the synoptic scale pressure, and the local wind system.

Pressure patterns

In fig. **4.1 a-c** the overall median of MSLP anomalies (anomalies compared to long-term monthly MSLP) is shown. The difference between fog anomaly and low cloud anomaly (fig. **4.1 a**) points to interesting areas where the fog and the low cloud situation differ. The two maps of the separate situations (fig. **4.1 b,c**) allow for a characterisation of each case.

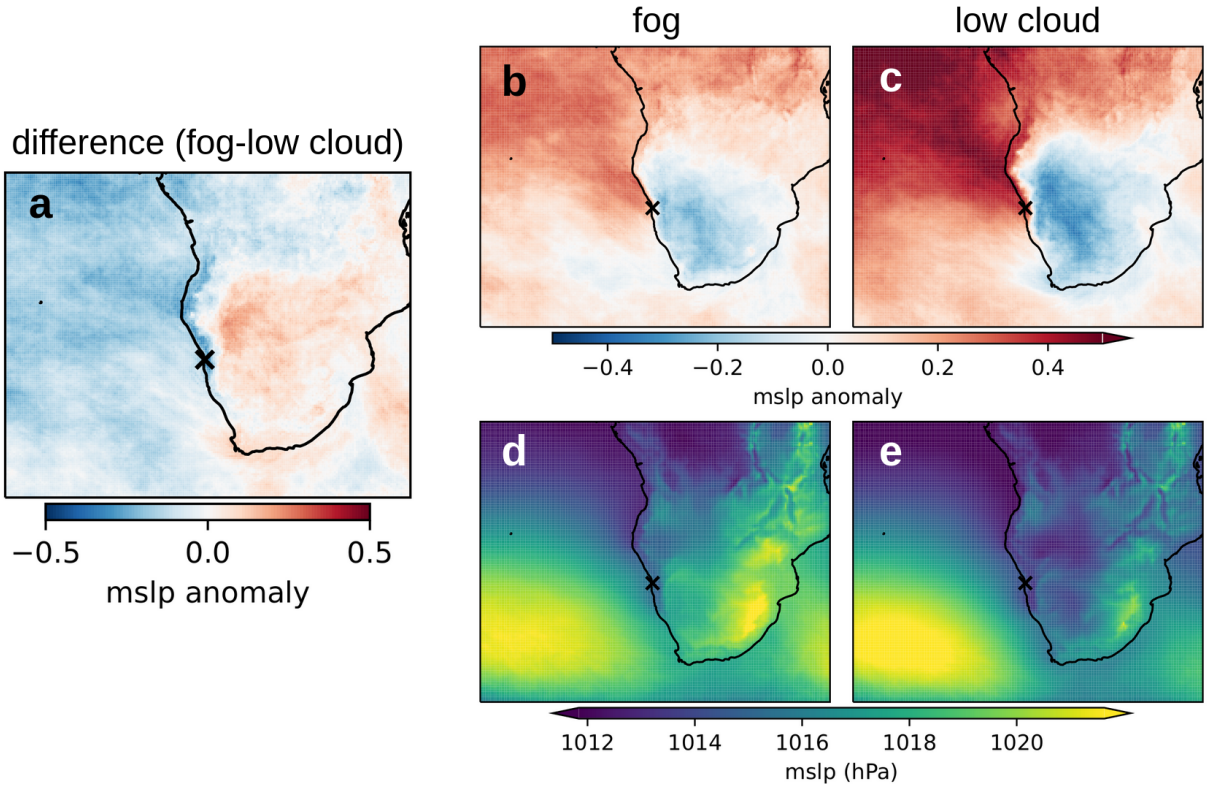


Figure 4.1: Median composites of mean sea level pressure (MSLP) at 5 UTC during fog and low cloud events at Coastal Met, Namibia. **a,b,c** anomalies to deseasoned longterm mean. **a** difference fog-cloud, **b,d** fog, **c,e** low cloud. x indicates Coastal Met station.

The fog and low cloud separate maps show a similar anomaly pattern that is stronger on low cloud days and weaker on fog days, thus indicating a common reason for the advection of the stratus deck. Characteristic feature of the common pattern is an anomalously low pressure (blue in fig. **4.1 b,c**) over the southern subcontinent compared to normal (i.e. the total of foggy, cloudy and clear days). The low inland pressure is more pronounced and also reaches further north and northeast on low cloud days.

Over the Atlantic Ocean, the low cloud situation shows stronger positive anomalies than the fog situation, indicating overall higher pressure on low cloud days. On fog days there is a blueish area over the Atlantic Ocean in fig. **4.1 b**. It appears in all seasons (Appendix fig. 6.1), though of varying position and extent. The location suggests a connection to the SAH that exists throughout the year, shifting north in austral winter and south in austral summer. In both fog and low cloud situations, the anomaly at the

location of the SAH is complemented by higher pressure over the northern parts of the ocean, thus indicating a lesser pressure gradient from the SAH to the equator than usual.

Another general difference in MSLP anomalies holding over all seasons can be seen within the regional surrounding of Coastal Met. In fog situations, the gradient between higher pressure on the ocean and lower inland pressure lies close to the coastline. In low cloud situations, the gradient is much sharper and set back inland towards the Great Escarpment (blue in fig. **4.1 a-c**).

To link the anomalies to real existing pressure features, maps of the median absolute values are provided (Fig. **4.1 d,e**). The high pressure on the Drakensberge and East African plateau (SE on the continent) is visible in both situations and all seasons, but stronger on the fog days. Compared to low cloud occurrence, the fog situation is characterised in all seasons either by a stronger CH or by a weaker SAH or both (fig. **4.1 d,e**). The pressure pattern in the Coastal Met region is less obvious but visible in both the fog and the low cloud median situation.

Inversion

To investigate the behaviour of the temperature inversion capping the marine boundary layer, latitudinal curtain plots through Coastal Met are provided (fig. **4.2**). The inversion can be seen in (fig. **4.2 d,e**) and is located where the cool air of the marine boundary layer (green) is topped by warmer air (yellow). It is inclined from the ocean towards the coast and is strongly developed on both fog and low cloud days, with an enhanced temperature contrast between the cool marine boundary layer (MBL) and the warm subsiding air from above (fig. **4.2 b,c**). The top of the MBL is in both situations exceptionally cool. This creates especially on fog days a gradient within the MBL, which is warmer than usual at the sea surface, indicating a greater moisture load and ability to produce a stratus when cooled. Both fog and low cloud situations show warmer temperatures over land (fig. **4.2 b,c**), going with the low-pressure anomaly seen before.

The difference between the fog and the low cloud situation is depicted in fig. **4.2 a**, and highlights the different position of the inversion. The inversion is lower on fog days than on low cloud days. It is not only marked by a temperature gradient but also by a gradient of specific humidity q (Appendix fig. 6.2). The exceptionally moist MBL is overlain by the exceptionally dry free troposphere, capping the stratus at different heights. Over the year the inversion capping the MBL varies in its westward extent. Most of the year it is over its full extent lower or higher in fog or low cloud situations, respectively. In DJFM the elevation difference appears only near the coast. The season DJFM is the season where the described patterns are mostly weaker than in the other seasons. E.g. the MBL is moister than usual in both fog and low cloud situations but not so much in DJFM. However, the dry anomaly above the MBL is more pronounced in DJFM, indicating that the moisture gradient is the relevant driver (fig. **4.3**).

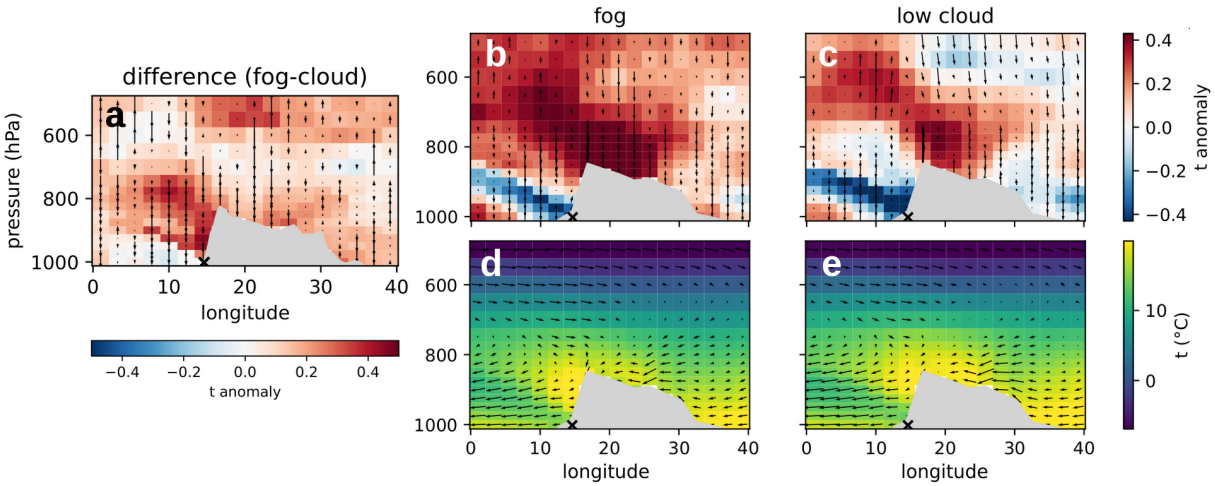


Figure 4.2: Latitudinal profiles of median temperature (t) at 5 UTC during fog and low cloud occurrence at Coastal Met, Namibia (x). Grey shaded area indicates the southern African landmass. **a** difference of fog and low cloud anomalies **b,c** anomalies during fog and low cloud events, respectively **d,e** median temperature during fog and low cloud events, respectively. Quivers denote wind (upward component multiplied by 20).

The latitudinal extent of the inversion turns out to be much wider than the local surrounding of Coastal Met and the Namib (blue in fig. **4.4**). On higher pressure levels the inversion stretches out and mimics the shape of the continent. The shallow inversion near the coast is more restricted in its extent and does not completely match the coastline; instead, it follows the cold anomalies in sea surface temperature (fig. **4.5**). Meanwhile, the SST does not show consistent differences between fog and low cloud days.

Besides the inversion height, fog and low cloud days also differ in the behaviour of the extra warm and dry air subsiding over the continent. On fog days the warm anomaly is spread over the continent and reaches further to the eastern coast than on low cloud days (fig. **4.2 a-c**), matching the pressure anomalies.

Wind

The wind vectors show that the subsidence of the warm continental air is much greater on fog and low cloud days than usual, but extremely so on fog days (fig. **4.2 a**). It is complemented by an upward motion at the beginning and top of the Great Escarpment, which is also stronger on fog days. The latter could be attributed to the local land-sea-wind system, where a thermally induced upward movement extends the dynamically developed low pressure over the southern subcontinent. The winds responsible for the inland advection are not visible on this scale.

The advection can be seen in the anomalies of the horizontal wind components (quivers in fig. **4.4**). In low cloud situations, the wind blows more from the NW than from SW like in fog situations, this is visible from the surface to about 900 hPA. The inland advection is on fog days already visible at 1000 hPa, while on low cloud days it becomes apparent

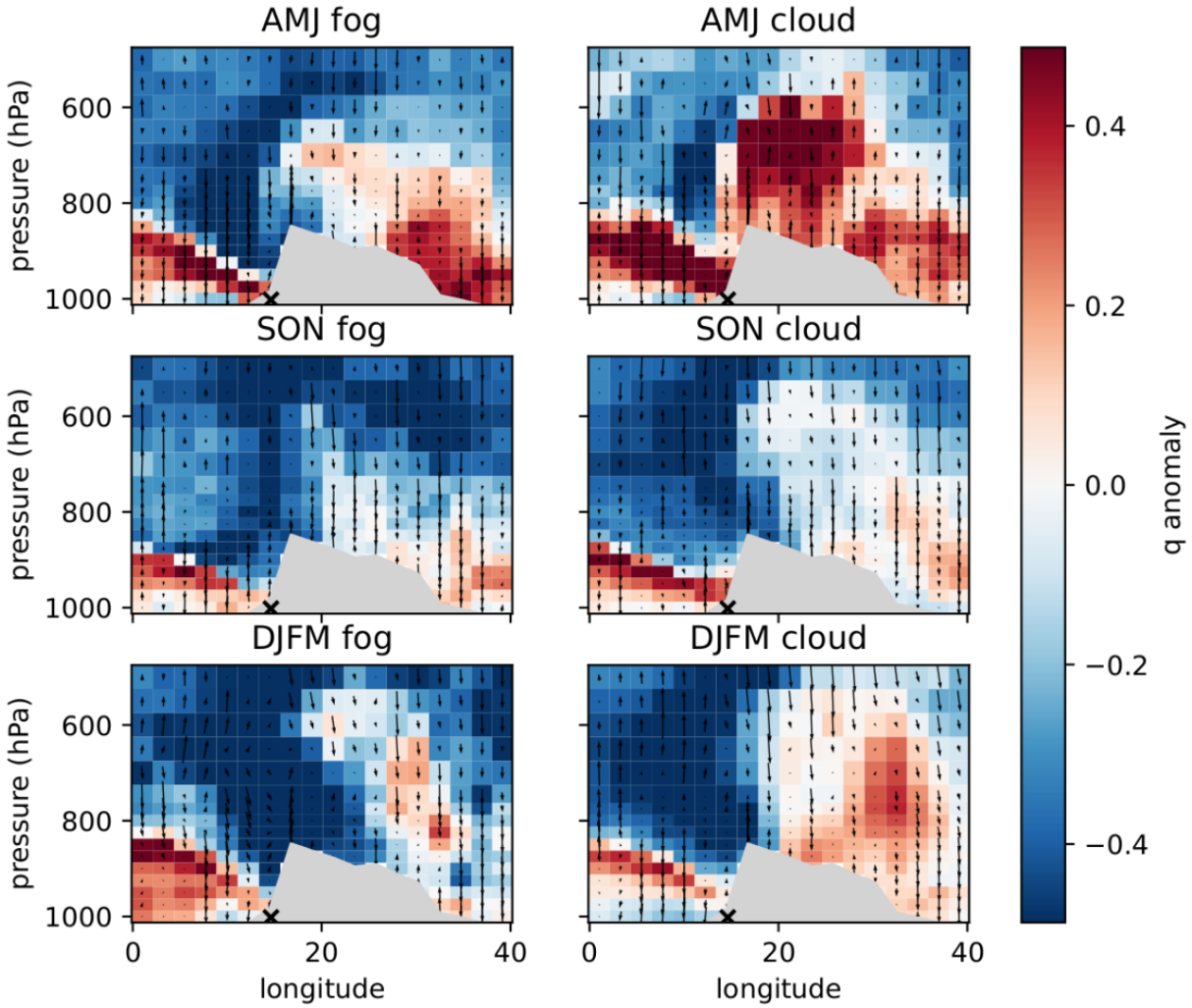


Figure 4.3: Latitudinal curtains of specific humidity (q) anomalies during fog and low cloud events at Coastal Met, Namibia (x) in different seasons. AMJ = April, May, June. SON = September, October, November. DJFM = December, January, February, March. Grey shaded area indicates the southern African landmass. Quivers denote wind (upward component multiplied by 20).

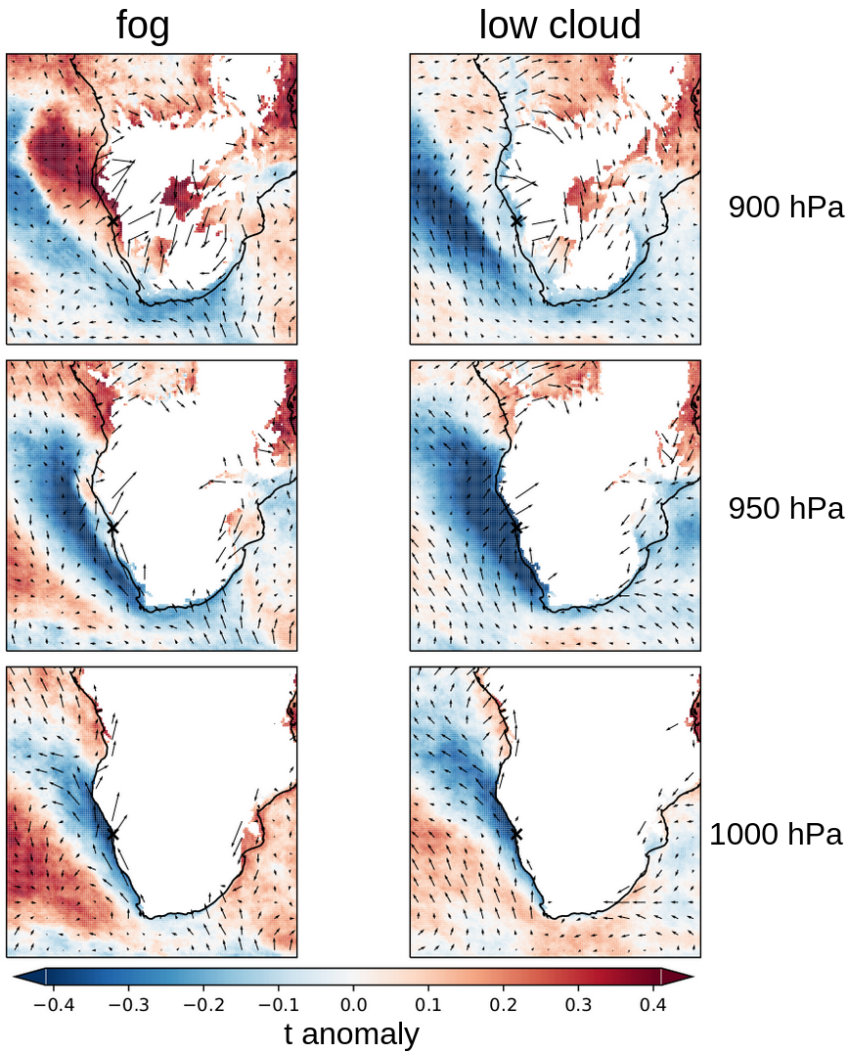


Figure 4.4: Spatial extent of the temperature inversion at different pressure levels. Left plots are the median temperature (t) anomaly during fog occurrence at Coastal Met, Namibia (x) at 5 UTC, right plots during low cloud occurrence. White are areas beneath the land surface, quivers indicate wind.

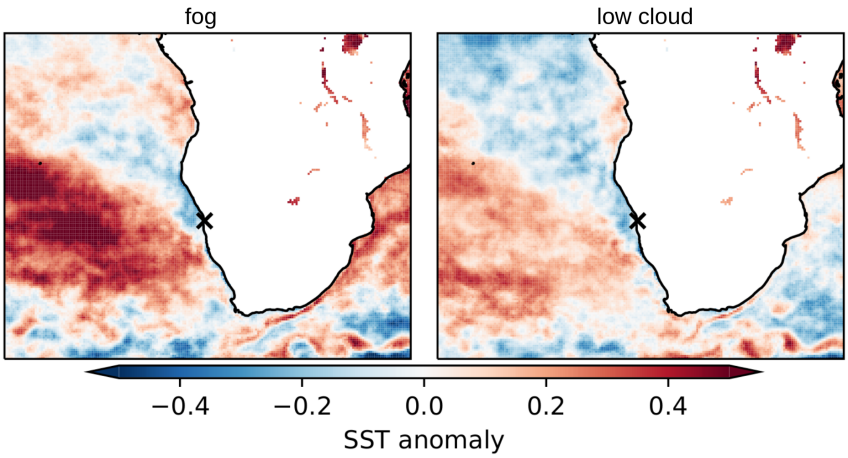


Figure 4.5: Median sea surface temperature (SST) anomaly during fog (left) and low cloud (right) events at Coastal Met, Namibia, at 5 UTC. x indicates Coastal Met station.

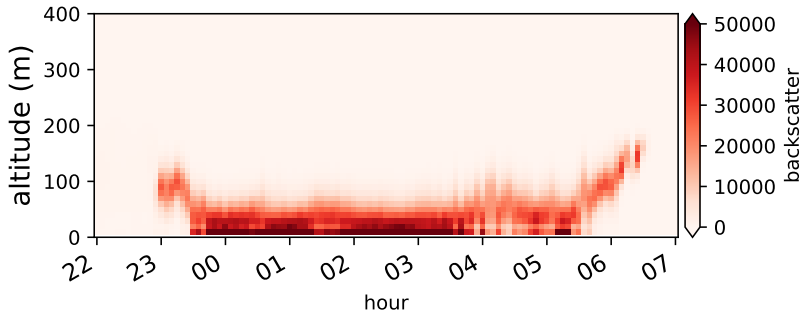


Figure 4.6: Ceilometer backscatter during the fog night 22./23.09.2017 at Coastal Met, Namibia.

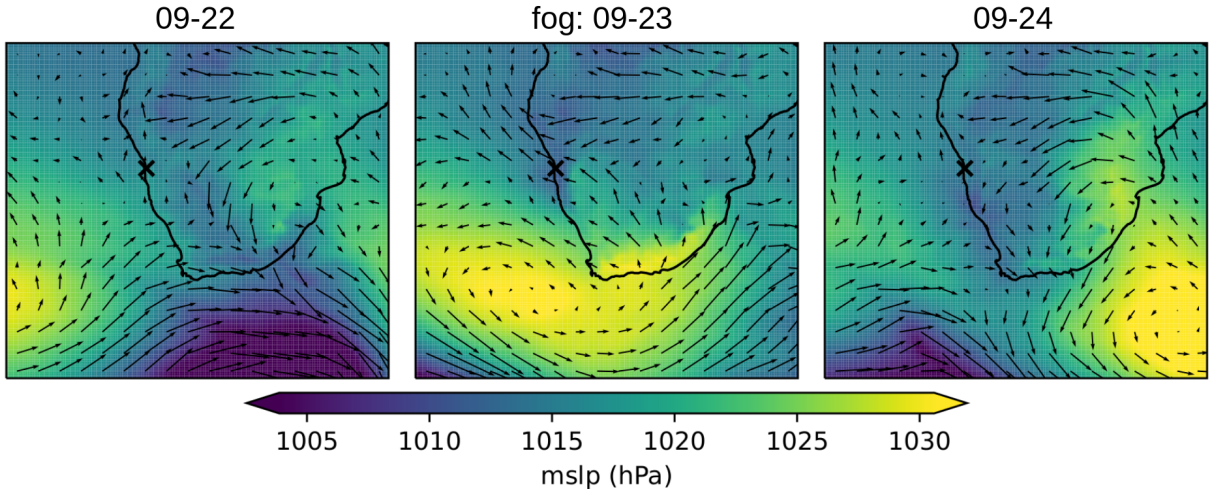


Figure 4.7: Mean sea level pressure (MSLP) at 03 UTC during fog occurrence on 2017-09-23 and one night before and after. Quivers denote the 850 hPa wind.

from 950 hPa upwards.

4.2 Case study

The general findings of the above section can be held against a specific fog event at CM to gain more detailed knowledge about the interplay of the proposed drivers and the temporal development of a fog event. In the night 22./23.09.2017 the ceilometer backscatter indicates fog from about 23 to 6:30 UTC (fig. **4.6**).

The large-scale MSLP confirms the pattern found in the previous analysis. In contrast to the nights before and after, there is a pronounced high pressure over the southern tip of Africa between 3 and 6 UTC, sending winds towards the Namibian coast (fig. **4.7**). The shown winds at 850 hPa are just high enough to overcome the Great Escarpment and push down the stratus. A closer look (fig. **4.8**) shows the local pressure and wind system that is driven by the land-sea contrasts. A high pressure over the Great Escarpment sends winds downward to the coast. Down at Coastal Met it is calm and allows the fog

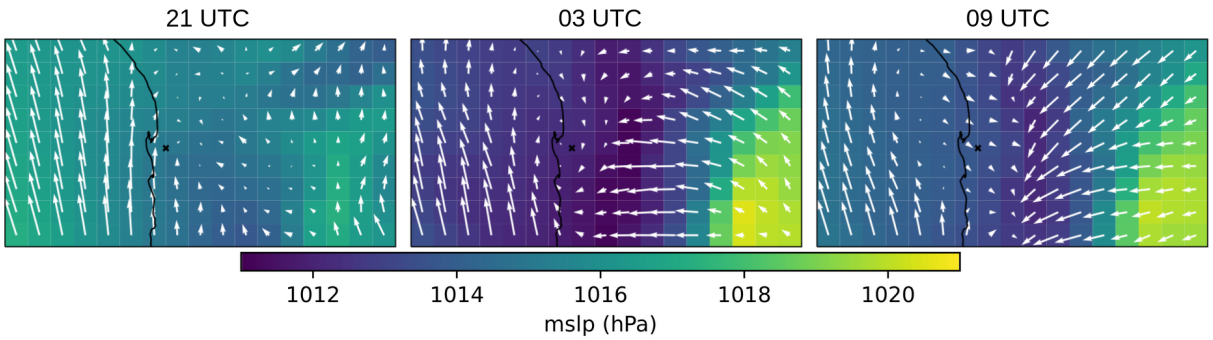


Figure 4.8: Mean sea level pressure (MSLP) during the fog night 22./23.09.2017. Quivers denote the surface wind at 10 m.

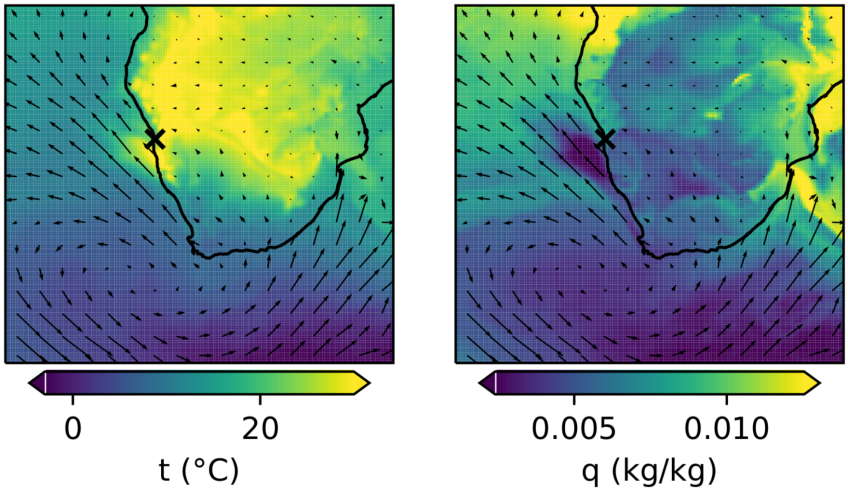


Figure 4.9: Temperature (t) and relative humidity (RH) at 950 hPa during the fog night 22./23.09.2017 at 3 UTC with corresponding wind (quivers).

to settle (10m wind is shown). The pressure gradient strengthens over night and faints before noon. While the onset of the fog occurrence around 23:30 UTC is well coordinated with the appearance of the pressure gradient, the dissipation happens at 4:50 UTC (fig. **4.6**) about half an hour after sunrise while the pressure gradient is still present.

The temperature inversion is located at Coastal Met and stretches to the south of the station. It is accompanied by an even larger area of comparably dry air (fig. **4.9**).

The vertical distribution of t and q corresponds to air movement implied by the wind quivers in (fig. **4.10**, detailed figure of q showing advection in Appendix 6.4). At 19 UTC, before fog occurrence at CM, the air is still rising above the Great Escarpment, probably a remainder of the earlier thermal uplifting. The air coming from the continent is carried away by this upward stream and comes down further away roughly over the coastline. The upward air movement then weakens and its location retails inland during the night until ceasing completely (Appendix fig. 6.3). Later in the night (from about 0 UTC on) the wind blows down almost parallel to the gently sloped land surface towards the ocean but is dismissed by the cold and moist layer covering the ground. The winds warm on their descent and intensify the gradient between the cold ground layer and warmer air

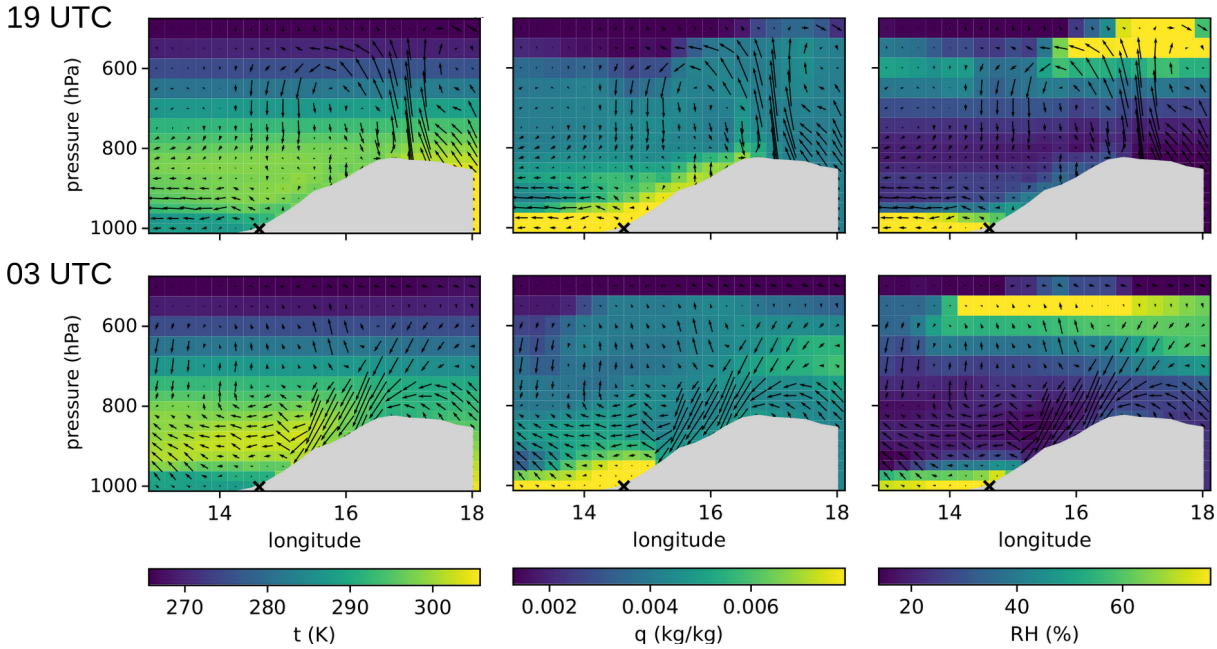


Figure 4.10: Temperature (t), specific humidity (q) and relative humidity (RH) at 19 UTC (before fog event, upper, row) and 3 UTC (during fog event, lower row) at a latitudianl slice through Coastal Met, Namibia, with corresponding wind (quivers, w is multiplied by 20).

above. They also seem to push back the layer, as there is much less moisture inland at 3 UTC than at 19 UTC. Moisture and temperature make up a thick layer of high relative humidity in the evening, situated mostly over the ocean and barely touching the land. At night, this layer has become thinner and has swapped over on the land, yet not reaching inland stations like Gobabeb at about 15°E.

4.3 H2: Classification

In this section, the potential of machine learning for the discrimination of fog and low clouds is assessed. First, some introductory statistics are shown on the information content of the features that were designed as input for the logreg and RF. After a comparison of all three models, logreg and RF are presented together with the best-rated features followed by the Convolutional Neural Network (CNN) and the baseline.

4.3.1 Feature engineering

With the understanding gained in the foregoing analysis, the drivers of fog and low cloud occurrence shall be packed into meaningful variables that serve as input for logreg and RF. Besides several simple features that were created by calculating the mean, maximum or minimum of a variable in a certain area, special focus was put on the inversion and the synoptic pressure situation. Additionally, a feature for the SST was designed.

Pressure patterns

The two subtropical pressure cells SAH and CH have been identified as potential drivers of fog or low cloud occurrence at CM in 4.1.1. To express the pressure cells in one variable, the HPI is calculated. The HPI combines the location and strength of a pressure cell by calculating the pressure-weighted mean location of the high-pressure cell, delineating it by a border isobar. For all considered isobars from 1016 to 1023 the HPIs on fog days differed significantly from the HPIs on low cloud days (Kruskal-Wallis test $p << 0.01$) for both SAH and CH. Though the HPI shows considerable variation for both pressure cells when plotted to a map (Appendix fig. 6.6), the variation cannot be directly attributed to fog or low cloud days from sight. The centers of fog day and low cloud day HPIs overlap and the overall pattern does not differ from clear day HPI (not shown). Broken down to the seasons (see Appendix fig. 6.7 and 6.8), only DJFM and AMJ latitudes of the SAH differed significantly (Kruskal-Wallis $p < 0.05$) on fog versus low cloud days, and SON longitudes of the CH. The CH shows a latitudinal position shifting during austral winter

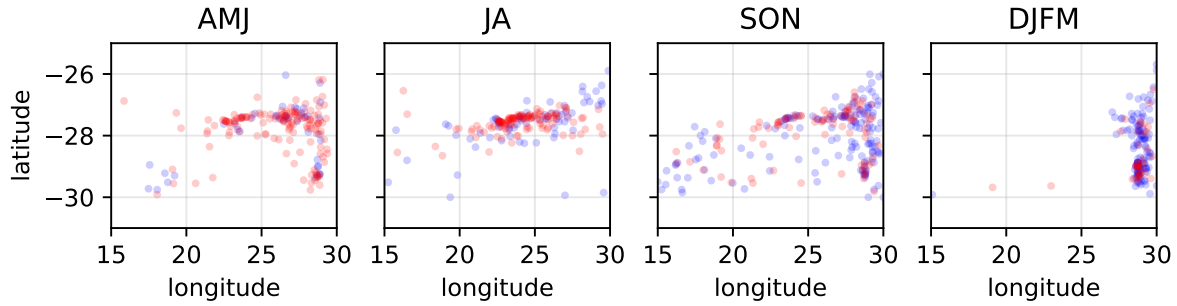


Figure 4.11: Continental high pressure index (CHPI) with border isobar 1021 hPA in different seasons during fog (red) and low cloud (blue) occurrence at Coastal Met, Namibia.

(JA in fig. **4.11**) and a longitudinal shifting in austral summer (DJFM). As fog days occur mostly in austral winter, the slight westerly bias of fog day CHPIs is probably an artifact caused by seasonality. Corresponding histograms to fig. **4.11** showed that the distributions of HPI longitude and latitude vary over the seasons but are similar for fog and low cloud days within the seasons (Appendix fig. 6.8).

Temperature inversion

A characteristic driver of the stratus elevation at CM is the inversion in front of the Namibian coast. Both Namib lower tropospheric stability (NLTS) and profile-based estimated inversion strength (EISp) are stability measures, the NLTS calculated from two pressure levels (925 hPa and surface) and the EISp from the strongest layer within the whole profile up to 500 hPA.

Both variables indicate a stable atmosphere on fog and low cloud days over the ocean northwest of the central Namib, compared to the greater surrounding. Especially on

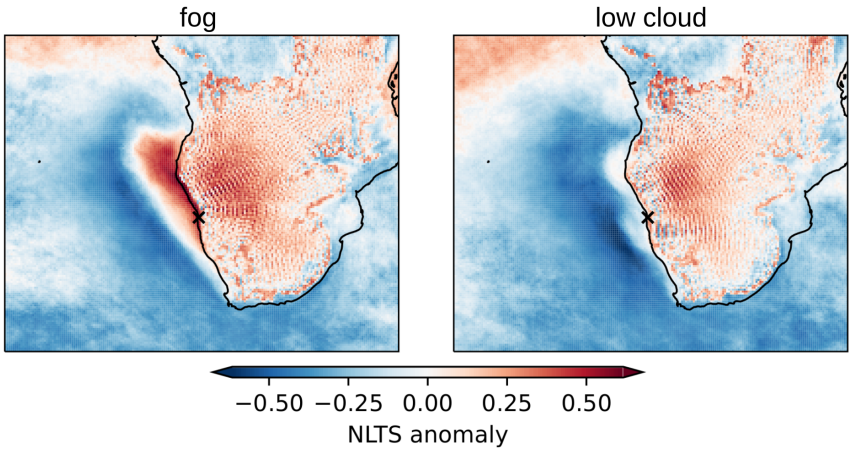


Figure 4.12: Namib Lower Tropospheric Stability (NLTS) median anomalies over the year on fog days and on low cloud days. The values on the inner continent cannot be safely interpreted because of the elevated terrain.

fog days, the atmosphere is more stable over a broad region northwest of the continent. The NLTS draws a pattern very similar to the temperature maps (fig. **4.13**). Despite the deseasoning there are considerable differences between the seasons in the broad scale anomalies (Appendix fig. 6.5).

The pEIS is stronger NW from CM at fog days but the amplitude and the exact location of the stronger pEIS varies. In DJF the anomalies are even reverse in some areas (fig. **4.13**).

Sea surface temperature

The SST was shown to impact fog occurrence at Lüderitz, a costal station south of CM (Olivier and Stockton, 1989), so SST was not completely discarded from analysis, though the composite analysis for CM has not yielded distinct patterns for fog versus low cloud days. A ridge regression with deseasoned SST indicates that the area south from CM is relevant for the fog/low cloud distinction at CM (fig. **4.14**). A second relevant area appears in the East of the subcontinent. Both are positively related to fog occurrence.

4.3.2 Logreg, RF, feature selection

To predict a fog / low cloud event at 5 UTC at Coastal Met station, logreg, RF and CNN models were used. The best performance was achieved by the logistic regression with an F1-score of 69.24 % and accuracy of 67.69 %, followed by Random Forest with F1 of 62.72 % and accuracy of 64.21 % (fig. **4.16**). The CNN did not predict better than the baseline, which was implicated as random draws from seasonal probabilities.

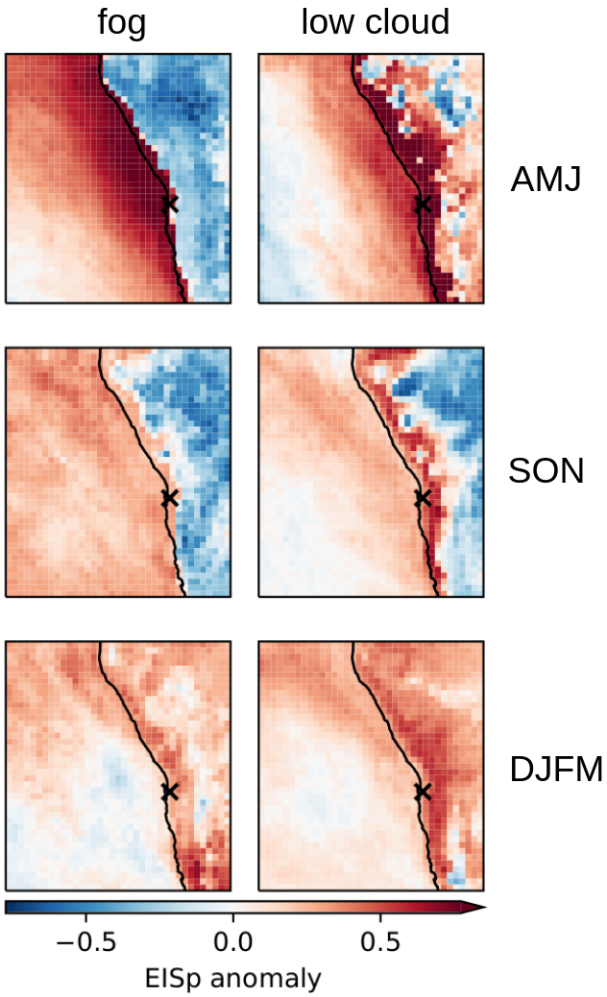


Figure 4.13: Profile-based estimated inversion strength (EISp) median anomalies during fog (left) and low cloud (right) events at Coastal Met, Namibia, in different seasons. AMJ = April, May, June, SON = September, October, November, DJFM = December, January, February, March. x indicates Coastal Met station.

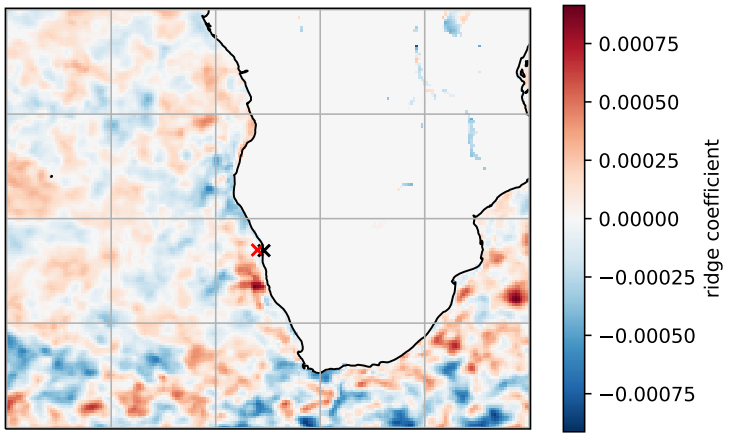


Figure 4.14: Coefficients from a ridge regression of sea surface temperature (SST) difference to a pixel (red x) near Coastal Met, Namibia (black x). Target factor is fog or low cloud occurrence at Coastal Met, $\lambda = 100\,000$.

Logistic regression

The logreg is the most simple method for supervised binary classification. In cross-validation with the custom time series split (3.2.2), the best mean prediction result was an F1-score of 69.24 %, achieved by standardising the data during preprocessing (65.51 % without standardisation). The F1-score balances between precision and recall, thus assuring that the prediction skill for both classes is assessed.

The standardisation is known to improve prediction results because arbitrary range differences between predictors (e.g. due to different units) are removed (Ait Ouadil et al., 2023). To crystallise the meteorological drivers from the seasonality, the data was, alternatively to the overall standardisation, standardised by the monthly mean and standard deviation of the training data. This yields a lower cross-validated mean F1-score below 60 %. There is no bias towards fog or low clouds.

In fig. **4.16** the test scores of the excluded test years during the cross-validation are shown. The lowest scores occur at the end of the time series with test year 2020, indicating exceptional conditions in that year. A periodical variation appears in the test scores when the SST feature is included.

Random Forest

With randomised search cross-validation (`RandomizedSearchCV` in sklearn) the best hyperparameters for the RF were determined in a cross-validation procedure, yielding a maximal depth of 21 and a number of trees of 500. With the best RF model, the following confusion matrix (CM) was achieved for testing with 2019 and 2020: which means

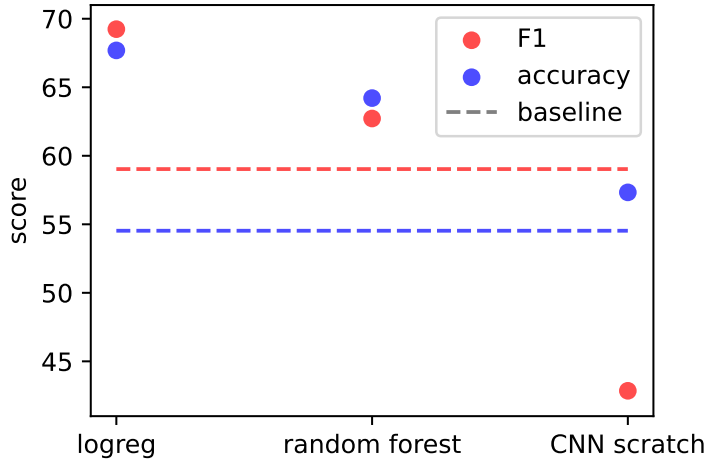


Figure 4.15: Best performances of the three model types logistic regression, Random Forest and Convolutional Neural Network built from scratch, on distinguishing fog and low cloud occurrence at Coastal Met, Namibia, compared to baseline (random draw with monthly probabilities).

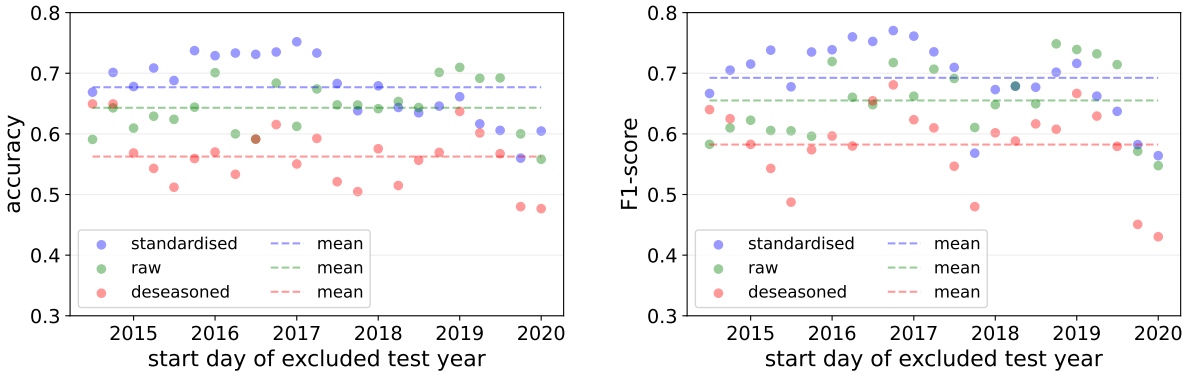
	fog predicted	low cloud predicted
fog labeled	true fog 90	false clouds 48
low cloud labeled	false fog 59	true clouds 102

Table 4.1: Confusion matrix for binary classification of fog and low clouds at Coastal Met, Namibia.

an accuracy of 64.21 %, precision of 65.21 % and recall of 60.40 %, making an F1-score of 62.72 %. Running the cross-validation procedure with F1 or ROC AUC as scoring yielded no better performing model.

Feature importance

According to feature importance, the 5 most important features of the best RF model are the HPI longitude of the SAH, the NLTS, the SST relative to a pixel near CM in the area east from the subcontinent, the HPI latitudes of the SAH and the MSLP (fig. **4.17**). These features were also the chosen features when applying recursive feature elimination (RFE) with RF estimator. With logreg estimator, the 5 most important features were NLTS, AHPI_lat, eastern relative SST, western relative SST and EISp. A detailed feature description is given in table **4.2**.



(a) accuracy (b) F1-score

Figure 4.16: Logistic regression with all features, compared to including standardisation in preprocessing and standardisation with monthly mean and standard deviation of training data.

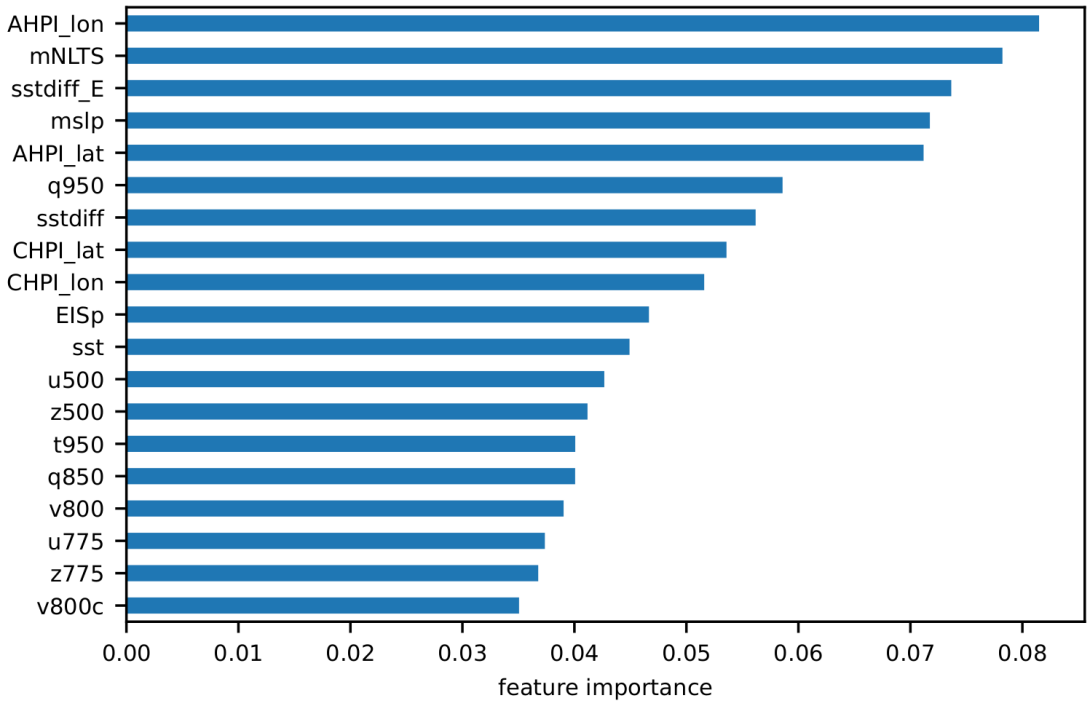


Figure 4.17: Feature importance of the best Random Forest model for classifying fog and low cloud events at Coastal Met, Namibia. Find a detailed description of the features in table **4.2**.

feature	logreg rank	RF rank	description	ERA5 raw input	extent	reduce method
mNLTS	1	1	Namib Lower Tropospheric Stability	t(925, 1000 hPa)	lon 7, 14, lat -14, -27	mean
AHPI_lat	1	1	HPI latitude of SAH (1017 hPa)	MSLP	-	HPI calculation
sstdiff_E	1	1	SST relative to coastal pixel near CM	SST	lon 38.5, 39, lat -27.5, -28	mean
sstdiff	1	3	SST relative to coastal pixel near CM	SST	lon 13.5, 14, lat -26.25, -26.75	mean
EISp	1	6	EISp	t, RH, z	lon 7, 14, lat -14, -27	mean
AHPI_lon	2	1	HPI longitude of SAH (1017 hPa)	MSLP	-	HPI calculation
CHPI_lat	3	4	HPI latitude of CH (2021 hPa)	MSLP	-	HPI calculation
CHPI_lon	4	5	HPI longitude of CH (2021 hPa)	MSLP	-	HPI calculation
u500	5	8	u wind component at 500 hPa	u	lon -10, 12, lat -28, -38	max
v800c	6	15	v wind component at 800 hPa	v	lon 23, 33, lat -12, -32	min
t950	7	10	t at 950 hPa	t	lon 8, 15, lat -15, -28	max
sst	8	7	SST	SST	lon 7, 13, lat -5, -18	min
u775	9	13	u wind component at 775 hPa	u	lon 0, 25, lat -5, -18	max
v800	10	12	v wind component at 800 hPa	v	lon -10, 18, lat -20, -38	max
MSLP	11	1	MSLP	MSLP	lon 15, 18, lat -12, -35	max
z500	12	9	geopotential on 500 hPa	z	lon -10, 5, lat -35, -40	max
z775	13	14	geopotential at 775 hPa	z	lon 3, 25, lat 3, 25	max
q950	14	2	specific humidity at 950 hPa	q	lon 8, 15, lat -15, -28	min
q850	15	11	humidity at 850 hPa	q	lon -10, 15, lat -30, -38	max

Table 4.2: Features to classify fog and low cloud events at Coastal Met, Namibia, with logistic regression and Random Forest. Prepared with ERA5 data. Rank is calculated by recursive feature elimination for 5 target features.

4.3.3 CNN

Two CNNs were trained on the maps of meteorological variables (instead of the features), one from scratch and one pre-trained CNN. For training a CNN from scratch, the best result was a test accuracy of 56.89 % and F1-score of 47.57 % achieved with a CNN consisting of one block (fig. **4.18**) and trained with deseasoned data (optimizer: adam, learning rate: 0.001, stride: 3). During training, accuracy and loss of the training data behave like expected (accuracy going up, loss going down) while validation accuracy is barely going up (**4.19**). Based on the curves, the CNN was trained for 5 epochs with which the named performance was achieved. Predicting the training samples after the initial 15 epochs yielded perfect prediction of fog and low clouds, indicating together with the low test accuracy that the model was overfit. From the raw, not standardised data, the CNN was not able to identify any relevant pattern as it predicts clouds for all samples of training and test data independent of how many layers are included.

```
_________________________________________________________________
Layer (type)                 Output Shape              Param #
=================================================================
conv2d_10 (Conv2D)           (None, 53, 53, 16)        4416

conv2d_11 (Conv2D)           (None, 49, 49, 16)        6416

max_pooling2d_5 (MaxPooling  (None, 24, 24, 16)        0
2D)

flatten_4 (Flatten)          (None, 9216)              0

dense_5 (Dense)              (None, 1)                 9217

=================================================================
Total params: 20,049
Trainable params: 20,049
Non-trainable params: 0
```

Figure 4.18: Architecture of the convolutional neural network built for a potential analysis concerning the binary classification of fog and low clouds in the Namib Desert.

The pre-trained CNN (mobileNetV2) takes 3-channel images, so three variables were selected. Due to their high feature importance in the RF and logreg, the NLTS, MSLP and SST were chosen. The best learning curves of the classifier on top of the frozen base were achieved with deseasoned data. However, validation accuracy did not exceed 55 %, and also training accuracy stalled below 60 %. Defreezing the top 50 layers of the base and fine-tuning them yielded no improvement.

4.3.4 Baseline scenario

The baseline scenario was computed by drawing random events from the monthly climatologic probability of a fog day and comparing them to the real events of a test year. When excluding different years for testing the mean baseline accuracy of 1000 trials ranged from 56.82 % for 2020 to 61.46 % for 2017 (see exemplary Fig. **4.20**). The total range of

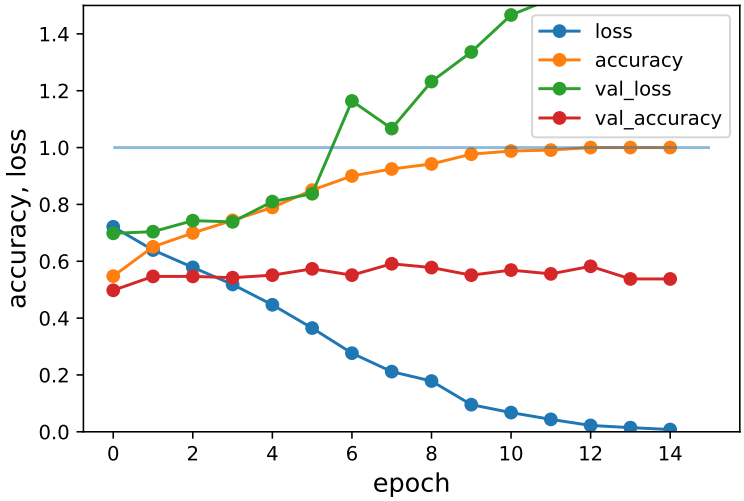


Figure 4.19: Learning curves of a 1-block CNN built from scratch and trained with deseasoned ERA5 variable maps to classify fog and low cloud events at Coastal Met, Namibia.

accuracies is 45.62 % to 72.67 %. The mean baseline F1-score of 54.53 % was lower than

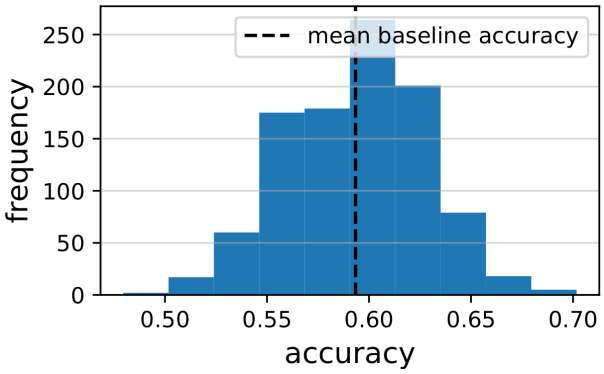


Figure 4.20: Baseline accuracies of fog classification computed from long-term monthly fog probability. Probabilities are the mean monthly share of fog days on all fog and low cloud days from August 2014 to December 2020, while excluding 2018 as the test year.

the mean accuracy of 59.03 % and it ranged in total from 23.70 % to 73.02 %.

Chapter 5

Discussion

The overall goal of this study was to deepen the understanding of mechanisms that lead to fog and low cloud occurrence in the central Namib (H1) and to explore the potential of machine learning to classify a stratus as fog or low cloud, based on meteorological variables (H2). Both hypotheses could be confirmed, with the only limitation that the CNN did not outperform the simpler algorithms. So far, the hypotheses have been treated independently. However, during the evaluation of the machine learning algorithm performances, further insights about relevant drivers emerged, as some features were ranked more important than others for predicting fog or low clouds. In the following, an extensive description of processes leading to fog and low cloud occurrence will be given (H1), with comparisons to literature and eventual consultation of the results from feature engineering and modeling. Thereafter, the procedure of feature engineering and the performance of the different machine learning algorithms (H2) will be discussed focusing on the modeling aspects.

5.1 New perspectives on the formation of fog and low clouds: conceptual framework

The fog and low cloud occurrence at Coastal Met turned out to be driven by four main drivers: the inversion capping the marine boundary layer (MBL), the synoptic pressure pattern, the local wind and the sea surface temperature (SST). The pressure and the inversion appeared in the composite analysis, the wind system was highlighted by the case study and the SST was recognised as an important feature by machine learning. The most direct driver of fog and low clouds is the height of the inversion layer that keeps the advected stratus down. The height of the inversion layer is on a synoptic scale determined by the constellation of the two subtropical high-pressure cells, the SAH and the CH. The large-scale pressure pattern is complemented by the local wind system that is governed mainly by daily temperature variation. Figure **5.1** broadly visualises the new concept of

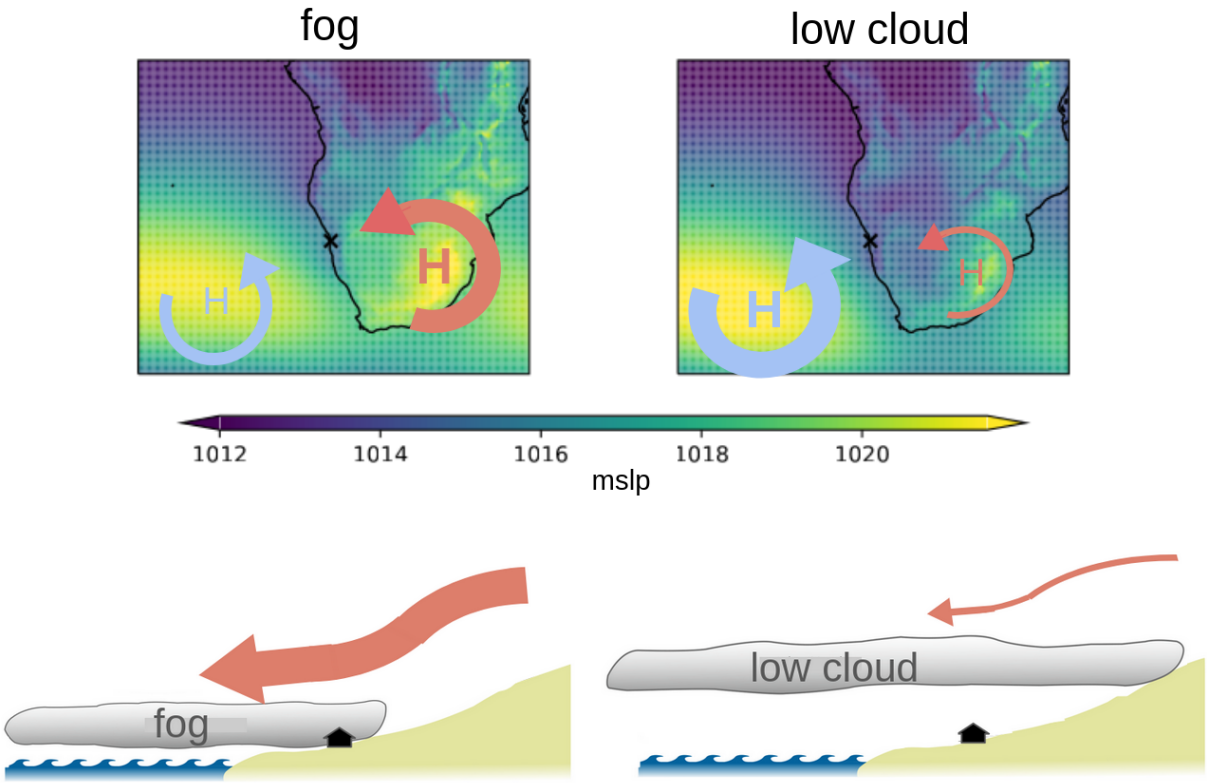


Figure 5.1: Conceptual visualisation of how synoptic pressure drives fog and low cloud occurrence in the central Namib. Figure modified from Andersen et al., 2019. The arrow symbolises warm and dry wind.

fog and low cloud situations in the Namib. The SST is apparently essential as well but its specific role needs further study.

5.1.1 Pressure

Beginning on a large scale, fog and low cloud situations are characterised by similar anomalies from the usual pressure. The anomaly patterns are more pronounced on low cloud days, indicating that low clouds are meteorologically an intensified version of fog. This insight is new to the fog and low cloud research in the Namib, as the different meteorological situations of fog and low clouds have been described extensively only on a local scale, based on station observations (Seely and Henschel, 1998, Lancaster et al., 1984). Large-scale research exists but has focused on the joint category of fog and low clouds (FLC) (Veloso et al., 2024, Andersen and Cermak, 2018). The finding is plausible when assuming that both fog and low clouds are advected stratus with the difference that low clouds are situated higher and are transported further inland. Thus the results also emphasise the common advective nature of both phenomena, in contrast to dissenting findings from isotope analysis (Kaseke et al., 2017, Kaseke et al., 2018).

In low cloud situations, a low-pressure anomaly on the land is sharply separated from the anomalous high-pressure coastal strip by the Great Escarpment. The trade winds,

that blow along the coast and over the cold Benguela current with its moisture-loaden marine boundary layer, are probably pulled inland by the low pressure and touch Coastal Met especially because of the thermal low south of the station. The role of this local low pressure was already highlighted by Olivier and Stockton, 1989.

Over the ocean, the median composites revealed a weaker pressure gradient from the SAH to the ITC than usual, so that the trade winds are possibly easier deflected inland. The local inland deflection of the wind is a known feature of the local climate of the central Namib (Weischet and Endlicher, 2000, Preston-Whyte and Tyson, 1988), but differences in its manifestation have not yet been linked to large-scale synoptics and fog and low cloud occurrence. At the same time the SAH turned out to be stronger in low cloud than in fog situations, probably producing strong winds that reach up higher in the atmosphere and also further inland.

The pressure over the southeastern plateau was found to be anomalously low in all seasons, especially in austral winter when there is normally a thermal high. This indicates that a pronounced continental high hinders low cloud occurrence. The continental low-pressure anomaly has already been shown to be related to FLC occurrence in comparison to clear conditions (Andersen et al., 2020). As the majority of FLC events are low cloud events, these findings can, cautiously, be compared. Andersen et al., 2020 interpret synoptic-scale disturbances as the responsible element for FLC occurrence, triggered by a large-scale gradient of low pressure on the continent and high pressure over the ocean. Possibly, the weak high pressure sends only weak winds to the ITC in winter or catches the trade winds in summer by a strong thermal low at the Congo-air-confluence, thus allowing the stratus deck at the Namibian coast to rise and mount the coast to travel further inland. Or in other words, a pronounced high continental pressure sends airmasses over the continent to the central Namib that pushes down the stratus and prevents it from mounting further inland. The impact of continental air on fog and low clouds in the Namib was also found by Veloso et al., 2024. They saw in ERA5 data, that the easterly airmasses that arrive in the Namib are moister during the low cloud season, and suggest that these airmasses are less subsiding than the dry ones during the fog season.

In fog situations the inland low-pressure anomaly appears to be less pronounced than on low cloud days and in fog peak months AMJ and JA the border between high-pressure coast and low-pressure inland is closer to the coastline. The weaker (or nearer to the ocean) inland low around Coastal Met can probably pull the stratus deck of the MBL not as far inland as in the low cloud situation. In all seasons except for DJFM, the pressure of the CH and also of the Indic High is higher than in the low cloud situation. Complementing the discussed low cloud situation, the southeastern high pressure sends air masses over the continent, that dry and warm on their way. These air masses probably push down the moist cool stratus after they get over the barrier of the Great Escarpment. Supporting this idea, the same explanation was also used by Veloso et al., 2024 to explain

FLC occurrence in the fog season AMJ. However, the present study is the first one to apply the conceptual understanding to disentangle the processes leading to fog and low clouds.

5.1.2 Inversion

A pronounced temperature inversion capping the marine boundary layer (MBL) was apparent in the ERA5 composites of low cloud and fog situations, accompanied by a sharp gradient in specific humidity. As described by former studies, the warm and dry air over the cold and moist MBL hinders convective ascent and thus determines the height of the stratus (Logan, 1960, Eckardt et al., 2013). The inversion could be shown to be lower in fog situations and higher in low cloud situations, confirming that only the higher inversion allows the stratus to mount the Namib inland. This has, up to now, only been shown indirectly by the annual variation of the MBL height. The MBL is thin during the fog season in early austral winter and thickens during summer (Veloso et al., 2024).

The inversion is not only a local phenomenon, rather it rises from the coast over a latitudinal extent of more than 1500 km to about 800 hPa over the ocean. The composite analyses revealed that the inversion is higher over its full extent in low cloud situations, except during the summer months DJFM. This underlines that it is driven at least partially by large-scale synoptics. Also, the inversion still mimics the shape of the coastline at higher elevations, again suggesting that it is not only driven by the spatially limited cold upwellings of the Benguela current.

An anomalously warm and dry free troposphere over the moist MBL was found on both fog and low cloud days, especially in the austral summer months DJFM and especially in fog situations. This confirms earlier findings of a dry anomaly associated with FLC cover derived from satellite imagery (Andersen and Cermak, 2018) and was apparent in ERA5 data (Veloso et al., 2024). Dry air above the MBL lets the MBL emit longwave radiation which can lead to a cooling of the MBL top (Adebiyi and Zuidema, 2018). The cooled MBL top compared to a warm MBL base could be shown in the composites (4.2). The gradient within the MBL was more pronounced in the fog case, probably leading to increased condensation within the MBL. Veloso et al., 2024 find that the difference between the potential temperature of the MBL and the SST is larger in July (2.4 K) than in January (0.5 K) which supports the idea of increased condensation near the ocean surface in fog situations.

Bringing together the findings for temperature and specific humidity, the troposphere above the MBL is drier on low cloud days while it is warmer on fog days. This potentially indicates a different origin of the air masses that, however, cause the same capping effect. The resulting gradient in relative humidity and the condensation process might thus be driven primarily by a temperature gradient on low cloud days and by a humidity gradient

on fog days. The interplay of temperature and specific humidity was also emphasized by the case study, as it was shown spots of exceptionally warm air are not congruent with spots of exceptionally dry air, both in horizontal (fig. **4.9**) and vertical extent (fig. **4.10**).

The two relevant elements recognised in composite analysis, the pressure and the inversion, were given to machine learning algorithms in the form of features. The features related to the SAH and to the inversion were among the 5 most important features, thus confirming our qualitative explanation for fog and low cloud occurrence at CM because they capture the impact of the SAH and the local temperature inversion. The SAH features express the position of the SAH weighted by the pressure and both latitudinal and longitudinal variations contributed to fog and low cloud occurrence. The high-ranked feature for the inversion describes the inversion strength between 925 hPa and the surface.

5.1.3 Wind system and case study

The presented results highlight the role of the local wind system especially in the case study of the 22./23.09.2017 fog night. During the daytime, the mountains of the Great Escarpment warm up and cause low pressure that deflects some of the cool and moist air of the trade winds blowing parallel to the coast. In the case study, the inland advection started in the morning, which is in line with descriptions of the local climate (Seely and Henschel, 1998, Spirig et al., 2019).

In the case study, the inland air motion ceased at nightfall and it became calm at Coastal Met around midnight. Between 22 and 23 UTC a temperature-induced pressure gradient between the mountains of the Great Escarpment and the maritime coastal plains becomes apparent. The mountain regions cool down fast after sunset and a thermal high pressure forms. The subsiding air warms on its descent towards the coast and settles above the cool and moist air from the ocean that was blown in earlier. The downward air movement from the cooling mountains establishes a shallow inversion and is most probably the cause for the low elevation of the stratus, occurring at Coastal Met as fog.

Favourable synoptics may promote the local winds that force down the inversion. The interplay between the synoptic scale and the local scale was found by Andersen et al., 2020 in a ridge regression with MSLP to predict FLC. Regional scale MSLP yielded the best results, because, according to the author, the regional scale bridges between the local and synoptic scale. To my understanding, the high pressure over the southern tip of Africa sends winds over the continent towards the coast that dry down while traveling over the land. When the local subsidence starts at night, the existing air over the mountains is already especially dry because of its origin. Station measurements confirm easterly winds in the early morning (Spirig et al., 2019 during September 2017). The easterly wind blowing from the continent was visible in our composites at 850 hPa, which is the same pressure level that was identified by Veloso et al., 2024 to cause low cloud cover in the

central Namib during the fog season March - August. The large-scale subsidence over the Namib has been found for FLC during the fog season by other studies (Andersen et al., 2020, Veloso et al., 2024) but the composites presented here showed that it is greater on fog days throughout the year. Together with the case study, the results reveal that the fog events outside the peak season are caused by the same meteorological processes as within the season.

At the coastline, the anomaly composites showed that the wind blows more from SW in fog nights and more from NW in low cloud nights. This is in line with early assumptions about the origin of the stratus: That a southwesterly wind touches the cold ocean surface and brings fog to the Namib while the higher stratus deck of the MBL is advected inland by the local sea breeze as low clouds (Seely and Henschel, 1998). However, Spirig et al., 2019 noted northerly winds at the station during fog events at Coastal Met and NNW winds during fog events at the inland station Gobabeb (which corresponds to a low cloud event in the present study). The wind direction at an inland station might, though, not represent the overall advection wind. The turning of wind directions from NW to NNE at inland stations during fog events is discussed in detail in Seely and Henschel, 1998. In a case study of a low cloud event during the night of 2017-09-27 / 28, Spirig et al., 2019 identified stratus advection from the NW which supports the presented findings.

Fog dissipation started in the presented fog night at about 5:30 UTC. Advection of marine air started again between 9 and 10 UTC but the pressure gradient already weakened between 5 and 6 UTC, corresponding to the time of fog dissipation. The dissipation is most likely timed by the rising sun at 4:51 UTC (6:51 AM local time). The data show a lagged warming in the local temperature at 2 m, probably because some solar energy goes into the process of fog dissipation first.

5.1.4 SST

Another important element for fog and low cloud occurrence was the sea surface temperature (SST). It was one of the 5 most important features in logreg and RF, which confirms earlier indications of its essential role for fog and low cloud occurrence in the Namib (Olivier and Stockton, 1989, Andersen et al., 2020). This role was, however, not further determined during the composite analysis, because there was no consistent pattern over all seasons. Two important areas of SST, though, were revealed by ridge regression, one near the coast south of CM, and one near the eastern coast of southern Africa. In both areas, SST is positively associated with fog compared to low clouds.

The positive relation near CM is in line with findings from Andersen et al., 2020. Similarly, they found a positive relation of SST near the coast to FLC occurrence only for AMJ, which is the fog season. Contrasting findings come from Veloso et al., 2024, who find that SST and MBL water vapor content is little related, especially in AMJ.

While many authors have emphasized the role of the cold Benguela upwelling in its role to lower the inversion directly at the coast (Seely and Henschel, 1998, Eckardt et al., 2013), warmer SST further south may also have a considerable impact on fog and low cloud occurrence. Following the argumentation of Olivier and Stockton, 1989 and Andersen et al., 2020, warm ocean water south of the station might be necessary to create a temperature gradient towards the station. The air descending within the SAH is dry and needs to take up moisture on its way to CM. Coming from the south it has to pass warm water to take up moisture, followed by a cool ocean surface to make the moisture condensate. Possibly, a steeper gradient of ocean surface temperatures causes the air to descend even more and lowers the inversion layer. This explanation fits the SW wind direction anomaly that appeared in the fog composites.

The SST on the eastern coast of southern Africa has not yet been related to fog and low clouds in the Namib. The warm waters of the Agulhas ocean current might cause the positive relation to fog occurrence. As a first guess, a warmer SST might act as an energy source and enhance the discussed processes leading to fog occurrence. More specifically, the warm SST could steepen the thermal gradient between land and sea, thus adding to the impact of the CH.

5.2 The potential of machine learning

5.2.1 Feature design and seasonality

To serve as model input for logreg and RF, the 3-dimensional ERA5 data had to be reduced to single-value features. Based on literature and insights from the composites, three elements were focused on: the temperature inversion which was captured by the Namib lower tropospheric stability (NLTS) and profile-based estimated inversion strength (EISp); the synoptic pressure captured by the SAHPI and CHPI latitudes and longitudes; and the SST captured by the mean difference of two different areas to a coastal pixel near CM. Several other variables were given to the models as mean, minimum or maximum of specific areas determined visually from the composites (table 4.2). The designed features turned out to be generally more important than the means of raw variables in specific areas, indicating higher information content.

The comparably simple measure of inversion strength, NLTS, was ranked more important than the more elaborated EISp. The NLTS was adapted from the commonly used LTS by calculating the temperature gradient of the 925 hPa instead of 700 hPa to the surface. The EISp captures the inversion strength based on the steepest gradient within a temperature profile up to 500 hPa. EISp has shown better performance than LTS in global applications on low cloud cover (Z. Wang et al., 2023) and the LTS was shown to be strong also in clear conditions offshore of the Namib (Veloso et al., 2024). The superiority

of the NLTS in the present study shows that the pressure level of 925 hPa, compared to the original LTS with 700 hPa, was a suitable choice together with the area focused by the feature. Here the simple measure, cautiously adapted to the use case, outperformed the more complex, universal measure.

Though the SST showed no distinct patterns in the composite analysis, it appeared among the most important features in logreg and RF. The special part of this feature was the idea to relate all pixels to a coastal pixel near CM. The high importance of the feature emphasises the need to work with gradients and relative measures. The applied ridge regression has proven a suitable tool to determine areas of interest for drivers that evade examination by simple median composites. Ridge regression has shown good performance also in classifying FLC and clear days in the Namib from MSLP (Andersen et al., 2020). Though the discrimination of fog and low clouds is probably more difficult than the discrimination of FLC and clear days, ridge regression is a promising tool for further research on fog and low clouds in the Namib.

The HPI was more important for logreg and RF than expected from statistics, and also simple MSLP was ranked comparably high. This highlights the importance of the synoptic pressure situation for the occurrence of fog and low clouds in the central Namib. However, the HPI relies on delineating single pressure cells, which was a successful approach for the Siberian High (Jia et al., 2015) but might not be the best concept to apply in the subtropical high-pressure belt. Images of single fog events revealed the frequent presence of a ridging high when the SAH and the IH merge. The ridging high has been identified as typical element of austral winter climate (Weischet and Endlicher, 2000) and spoils the delineation of two discrete high-pressure cells. I suggest treating the high-pressure belt near the subcontinent as one entire area for HPI calculation in further studies.

Both the HPI features and the SST revealed their importance for fog and low cloud occurrence only when used in classification models. In contrast to the simple median composites of each variable, the models can take into account interactions between the predictors. This suggests that the interaction between the drivers is crucial for fog and low cloud occurrence. The position of the SAH and CH as well as SST are thus probably indirectly related to fog and low clouds and their impact relies on other co-occurring factors. To further design meaningful features for the differentiation of fog and low clouds, a sound understanding of the driving processes is needed that might not be achievable by simple statistics.

Capturing the conceptual findings from H1 in useful features for modeling turned out to be challenged by seasonality. The temperature inversion was displayed in the stability measures NLTS and EISp which showed differences between fog and low cloud days. However, the differences were not uniform over the seasons despite standardising the data by the monthly mean and standard deviation. Although the same general pattern could be recognised in all seasons, the anomalies varied in their strength and the regions of strong

anomalies varied in their location. Ideally, the values of the feature should be higher in one group (e.g. fog) than in all seasons of the other group. Otherwise, the model will hardly be able to distinguish the two groups without information on the season. The varying location and strength of the inversion over the seasons complicates the preparation of a feature that is meaningful in all seasons.

Especially in DJFM the differences between fog and low cloud anomalies of pEIS were weak or in some regions reversed, which was also the case for other variables, though less pronounced. Some studies have shown that the LTS and EIS explain only 4 % of near-time LCC variations when the monthly mean is subtracted compared to 12 % with seasonality (Klein, 1997, De Szoeke et al., 2016). This shows that separating causal relations from coincident co-occurrence due to seasonal variation is difficult in low cloud research. Accompanied by our study, the general impression forms that the mechanistic drivers of low clouds are to date only partially understood and little more than can be deduced from seasonality. This study is a first approach to disentangle seasonality from causal mechanisms. DJFM is the season with the fewest fog occurrences, so median images might be less reliable than in other seasons. To get a grip on the meteorological drivers of fog and low clouds, further study has to focus on fog events within this season.

A periodical variation became visible in the test years scores (see fig. **4.16**) when including the SST feature in the logreg. The variation is not annual but a little more than two years. This indicates that the SST and its driving forces, the ocean currents, might follow their own seasonality that is not directly related to the annual movement of the ITC (as is e.g. the pressure pattern). Such variation would not be caught by monthly standardisation. The SST is partially driven by the El Nino Southern Oscillation (ENSO) but no relation was found between the irregularly occurring ENSO phases and SST or fog and low cloud days (ENSO time series from Alizadeh, 2024). According to Veloso et al., 2024, the summerly (DJF) moisture transport from the continent to the free troposphere of the Namib region strengthens during the cold phase of the ENSO. As the ENSO events occur in irregular intervals of one to several years (Alizadeh, 2024), a longer time series is needed to study its impact on fog and low clouds in the Namib.

Synthesising the gained conceptual understanding and the reflections on feature engineering yields some ideas for features in further research. As discussed, the height of the temperature inversion is the most important element in determining the height of the advected stratus. Both NLTS and EISp measure the strength of the inversion rather than its height. This may be adequate in studies focusing on FLC as a joint category (Veloso et al., 2024, Z. Wang et al., 2023). For the differentiation of fog and low clouds, the two features might not capture the desired information. In the calculation of the EISp, the ERA5 pressure level of the steepest gradient of the potential temperature is found and used to calculate the inversion strength. In further research, not the inversion strength but the pressure level of the inversion could be used as a feature, potentially capturing

the relevant information. As the MBL is capped by a gradient in humidity as well as temperature, a feature that combines both variables could inspire further research.

Pre-fog data was not included in this study to keep it a simple potential analysis. The LTS has been shown to forego LCC change about 24 - 36 hours (Mauger and Norris, 2010). In the Namib, Andersen et al., 2020 found in a ridge regression with MSLP that FLC were predicted best with the actual or one-day-before MSLP compared to longer time lags. However, they attributed the good performance of the pressure of the foregoing day to temporal autocorrelation of synoptics and FLC events. Conceptually, the fog and low cloud occurrence might be more related to the meteorological conditions in the foregoing evening than in the early morning when the fog or the low cloud is already present. This is especially plausible as we assume an advection process. Especially the highlighted large-scale synoptics may have delayed local effects.

5.2.2 Comparison of machine learning algorithms

In order to assess the potential of machine learning, three different algorithms were trained to distinguish fog from low clouds at CM. The skills of the models were compared to a baseline scenario that represents guessing from season. Logreg and RF showed better prediction skills than the baseline and will be discussed first, followed by the CNN that yielded an F1-score below 45 %.

With better predictions than the baseline, logreg and RF showed that some relevant mechanistic information was contained in the features given to the models. It opens the potential to classify fog and low clouds from meteorological data which is a thrilling perspective, given that research so far had to rely on the punctual and temporally limited station data. Large-scale studies have treated fog and low clouds as one joint category and the responsible drivers for the two situations could only be outlined by characterising the peak seasons (Andersen et al., 2019, Veloso et al., 2024). With classification by meteorological parameters, fog occurrence over the last 80 years could be traced back, because the ERA5 data used in this study is available back to 1940. This could contribute to filling the knowledge gap around the past and future of the Namib as a fog-dependent ecosystem, which could so far only be treated by modeling at coarse resolution (Haensler et al., 2011) or by repeated photography (Rohde et al., 2019).

The achieved model prediction skills of 67 % accuracy and 69 % F1-score (logreg) and 64 % accuracy and 63 % F1-score (RF) are promising in the context of potential analysis. The differentiation of fog and low clouds with meteorological variables at the time of occurrence has not been attempted so far (Lakra and Avishek, 2022), so that there is no benchmark to compare the scores. In fact, the results of the present study create a benchmark for further studies concerning the differentiation of fog and low clouds. Most studies that classify fog come from now-casting at airports, where the fog event

is predicted from a time series of foregoing data (J. Park et al., 2022, Fabbian et al., 2007). In remote sensing, fog and low clouds are jointly treated as FLC (Andersen and Cermak, 2018, Cermak, 2018) or as low cloud cover (LCC) compared to higher clouds or clear conditions. Recently, the fog precipitation in the Namib could be approximated by CALIPSO Lidar data with an R^2 of 0.85 (Qiao et al., 2022). The R^2 is a metric for a regression task and cannot be directly compared to classification metrics like accuracy. However, the presented 67 % accuracy is considered a good performance, considering the complexity of the meteorological processes.

On a gradient of model complexity, the most simple model (logreg) yielded the best performance. It is usually outperformed by RF, as has been shown in a meta-analysis of binary classification tasks (RF better in 70 % of cases, Couronné et al., 2018). The logreg can be better than RF when the number of features is small compared to the number of observations. (Couronné et al., 2018). It yields higher accuracies when applied to datasets with high variations in the predictors (Kirasich et al., 2018) and can deal better with imbalanced datasets which was shown for storm prediction (Ruiz and Villa, 2008). Random forest is sensitive to imbalanced data because the Gini impurity favours the majority class (Cieslak and Chawla, 2008). The number of features here given to the models is with 15 comparably small to what RF models can handle, which may be one reason why RF did not outperform the logreg. However, the slight imbalance of the dataset, consisting of 623 low cloud samples and 504 fog samples might have caused the lower prediction skill. In the original approach of this study, only FLC-positive days were included and split into fog and low clouds by station fog measurements. The obtained dataset was heavily imbalanced towards clouds with about twice as many low cloud samples than fog samples, which yielded zero-skilled RF models. This indicates a strong sensitivity of RF to imbalanced data and highlights the challenges associated with modeling meteorological phenomena with pronounced seasonality.

The logreg showed better skill on generally standardised data than on data standardised with the monthly climatological mean and sd. The standardisation method can have a significant impact on the model skill (Ait Ouadil et al., 2023). Here, the monthly standardisation was meant to eliminate seasonality from the data and focus on the meteorological mechanisms. This made predictions worsen, indicating that the chosen features did not represent the meteorological mechanisms alone and that the remaining seasonality was important to predict the class. It can be concluded that, along with all technical problems of feature engineering, the meteorological drivers of fog and low cloud occurrence in the Namib are to date not entirely understood and need further study. However, including seasonality enabled the model to distinguish between fog and low clouds with an F1-score of 66.81 % which highlights the potential of predicting fog and low clouds from meteorological data.

Taking the end of the time series (last year 2020) as test data yielded especially low

prediction scores of the logreg compared to testing on other years. The low prediction scores are possibly related to exceptional weather in that period. The numbers of fog and low cloud days show an unusual fog day peak in January 2020 and little fog during the year, accompanied by few low cloud days around the turn of the year (2019/2020). The predictions of the logreg considering seasonality are thus possibly put off by the unusual fog and low cloud seasonality in 2020, which might be caused by unusual weather conditions. Hence we interpret the low prediction skill of the meteorology-focused logreg (with monthly standardisation) to anomalous weather conditions during these periods, blurring the mechanisms that usually lead to fog or low cloud events. This is even more plausible having in mind the morphological heterogeneity among the fog events. As the ceilometer data for the case study demonstrates, fog events can look different and the fog precipitation is not quantitatively related to the duration of the fog event or the visibility (as was also shown by Spirig et al., 2019). Those are co-determined by the droplet size or the density of the stratus etc. (Spirig, 2022).

The CNN yielded low accuracy of about 57 %, which was above baseline (54 %), but a low F1-score of 43 % (baseline 59 %). The F1-score combines recall and precision and can therefore, in contrast to accuracy, not be fooled by only predicting the majority class. The low F1-score was caused by overestimating low cloud occurrence. Deep learning models are extremely complex model architectures that are known to require large amounts of training data. A CNN was outperformed by simpler machine learning models until a sample size of 1500 on classifying spectroscopy data (Ng et al., 2020). The benchmark dataset for image recognition, MNIST handwritten digits, contains 70 000 images (Wu and Chen, 2015). With 1127 samples of which the minority class, fog, comprises 504 samples, the here used data set is comparably small. Only 112 fog days fall into the low cloud season SON and 71 into NDJ, further reducing the number of fog and low cloud samples that are available for direct comparison during the same meteorological setting. The lack of model skill can thus be attributed to the lack of enough training data.

Compared to logreg and RF, the CNN was given the raw information of the variables instead of condensed information in the form of features. The overload of explanatory variables may have caused the instant overfitting of the model. CNNs, like RF, are sensitive to class imbalance which has initiated research around different dealing strategies (Pouyanfar et al., 2018, Chen et al., 2021). The sensitivity comes from optimising over all predictions which drags the model towards predicting the majority class. Possibly, the slight imbalance in the fog and low cloud dataset affected the CNN even more than the RF. Unstandardised data yielded completely useless predictions which shows that the CNN is sensitive to noise. The results of machine learning algorithms can be largely affected by standardising or normalising the input data (Castillo-Botón et al., 2022).

Two CNN models were trained, a CNN built from scratch and a classifier on top of the pretrained mobileNetV2. The best CNN trained from scratch showed instant

overfitting. In contrast, fine-tuning the pre-trained CNN mobileNetV2 yielded a stalling training and validation accuracy below 60 % and was therefore not further tested. The different behaviour of the scratch CNN and the finetuned CNN can probably be attributed to the number of input variables given. While the self-made CNN was created to take eleven variables as input, the pre-trained CNN was given only three variables due to the pretraining on 3-channel images.

Training the scratch CNN on eleven variables probably allowed the net to adapt its weights closely to the seen samples, resulting in overfitting. In contrast, the pre-trained CNN apparently got too few variables to adequately adapt its weights, as it yielded only low cloud predictions even on the training data. The three variables were chosen based on the feature importance assessed with RF and logreg: raw SST, MSLP and NLTS. They did apparently not hold enough information to predict the training samples. This indicates that more variables are needed to describe fog occurrence in the Namib adequately and highlights the complexity of the meteorological processes.

The low performance of the CNNs encourages to question the proposed concept of treating meteorological situations as an image recognition task. Possibly the transfer is not as straightforward as expected, as recognising an object or pattern within an image may be in some ways different from recognising a meteorological situation. E.g., different meteorological conditions could lead to the same phenomenon, like condensation can occur due to cooling or due to pressure increase. This was thought to be analogue to recognising an object of different appearances. However, with the limited process understanding of the current state, there is no knowledge about how diverse the conditions differentiating fog and low clouds actually are. Also, the absolute positions of e.g. the pressure cells compared to the continent may indeed matter, while the unique skill of CNNs is to find patterns in images independent of their location.

There is still great potential in transferring the concept of image recognition to the classification of atmospheric phenomena by meteorological variables, as only limited time was put into building the CNN and tuning its hyperparameters. The hyperparameters have a significant influence on the learning process and thus on the final classification skill. The number of adjustable hyperparameters is large, like the structure and capacity of the model, learning rate, optimiser, loss, to name a few, and they cannot be found other than by try-and-error (Ahmed et al., 2020, Radiuk, 2017). Considering the sensitivity of complex models to imbalanced or noisy datasets and the variety of possible hyperparameter combinations, more effort has to be invested in the preprocessing and model tuning to assess the potential of deep learning for distinguishing fog and low clouds in the Namib. As a potential improvement of the conceptual frame, the latitudinal curtain plots of the temperature inversion or humidity gradients could be used as input, as they probably contain more information compared to noise.

5.3 Other aspects

The drivers of fog occurrence have been examined in this study under the assumption that they have been stationary within the examination timeframe. However, under climate change, some drivers are supposed to change which could alter the fog occurrence. The constellation of the two quasi-permanent high-pressure cells over the South Atlantic and over the southern continent was shown to be an important factor for the elevation of the stratus and thus for the spatial pattern of fog occurrence in the Namib. For an A1B emission scenario until the year 2100, Jury, 2013 gets a poleward drift of the subtropic high-pressure cells together with a sea level rise of about +1.5 mm/year. Kaseke et al., 2017 considered an already changing water cycle system when isotope analysis claimed more than half of the fog events at Gobabeb to originate from evaporated groundwater and local cooling instead of transported ocean water in 2017.

Chapter 6

Conclusion

This study is the first attempt to explicitly differentiate fog and low clouds in the central Namib using large-scale meteorological variables. Fragmented knowledge on the processes that drive fog and low cloud occurrence comes from local observations at weather stations or from attributing meteorological characteristics of the peak seasons AMJ to fog occurrence and SON to low cloud occurrence. Two hypotheses guided this work, the first one concerning the drivers that lead to fog and low cloud occurrence and the second one addressing the potential of machine learning algorithms to differentiate between the two cases.

Specifically, the first hypothesis states that fog occurrence is characterised by different meteorological conditions than low cloud occurrence. This could be confirmed by median composites of a chosen set of ERA5 variables. Consistent patterns of anomalies from climatology were found throughout all seasons and were confirmed within a case study. Three main drivers could be recognised: The synoptic pressure, especially the SAH and the CH; the height of the temperature inversion at the Namibian coast; and the local wind system, which is probably enhanced by the synoptic pressure.

The second hypothesis states that machine learning can, based on the relevant meteorological drivers, differentiate fog from low cloud events. The hypothesis was confirmed, as the two machine learning algorithms logistic regression and Random Forest yielded prediction accuracies of 67 % and 64 %, respectively, and F1-scores of 69 % and 63 %, which exceeds guessing from seasonality. As a second part of the hypothesis, deep learning was expected to outperform the named models. This could not be confirmed, as the chosen model architecture, a Convolutional Neural Network, could not be sufficiently trained with the available small dataset.

The presented work founds a new process understanding of fog and low cloud occurrence. During low cloud occurrence, the SAH is strong and the CH is weak, which allows the marine boundary layer to extend higher up. This constellation also allows the local sea breeze to transport the offshore stratus far inland. During fog events, a stronger CH sends winds over the continent that subside over the Namib and, merged with the local

wind, push down the temperature inversion capping the marine boundary layer. Ridge regression of the SST revealed that it could be another large-scale driver of fog and low cloud occurrence. Besides a spot south of Coastal Met, logreg and RF gave high feature importance to SST in a limited region in front of the eastern coast of southern Africa for which there is currently no qualitative explanation. This highlights the complexity and interactive nature of the processes that lead to fog or low clouds in the central Namib.

A major challenge was to detangle causal relationships from co-occurrence due to seasonality. The median composites were prepared separately for the seasons and checked for common patterns. However, the applied deseasoning by standardisation with monthly means does not account for positional shifts of e.g. the pressure cells during the year, as it is applied in each pixel independently. The varying location of spatial patterns also complicated the feature engineering for logreg and RF. Concerning modeling, the slight imbalance of the dataset due to more frequent low cloud events was probably the reason why the RF performed not as well as the simple logreg.

During this work, some specific aspects caught attention that may inspire further research. Apparently, gradients may be more important than the raw variables at specific locations. A horizontal gradient in SST from the ocean to CM as well as vertical temperature and moisture gradients within the MBL and from the MBL towards the free troposphere have been recognised. They appear persistently over annual variations and may thus be an inspiration for further feature engineering. For fog to actually materialise, the exact ratio of temperature and specific humidity is crucial. The case study has shown that there can be high q reaching far inland, but without the matching temperature, no fog appears. So, while it is clear that low temperature and high q make fog occurrence likely, the specific spatial and temporal interplay of the two, driven by air masses of different nature, wait for further study. The inversion that is relevant for fog and low clouds at Coastal Met has been shown to extend from the tip of the subcontinent to the Angolan coast. More studies are needed to investigate if this location of the inversion produces fog at the whole coastal strip of the Namib or if coastal fog at, say, Lüderitz is linked to a different spatial pattern of stable conditions.

The presented results highlight the potential of machine learning for classifying fog and low clouds from given meteorological variables. They open the perspective to model the Namib fog regime of the past and future which is highly desired information as fog is the only water source for many organisms in the desert and can also be used to generate irrigation and drinking water for human purposes. Deep learning can still, with more data and intensified hyperparameter tuning, hold a great potential for the differentiation of fog and low clouds in the Namib.

Bibliography

Abadi, M., Barham, P., Chen, J., Chen, Z., Davis, A., Dean, J., Devin, M., Ghemawat, S., Irving, G., Isard, M., et al. (2016). TensorFlow: A system for large-scale machine learning. *12th USENIX symposium on operating systems design and implementation (OSDI 16)*, 265–283.

Adebiyi, A. A., & Zuidema, P. (2018). Low cloud cover sensitivity to biomass-burning aerosols and meteorology over the southeast Atlantic. *Journal of Climate*, *31*(11), 4329–4346.

Adhikari, B., & Wang, L. (2020). The potential contribution of soil moisture to fog formation in the Namib Desert. *Journal of Hydrology*, *591*, 125326.

Ahmed, W. S., et al. (2020). The impact of filter size and number of filters on classification accuracy in CNN. *2020 International conference on computer science and software engineering (CSASE)*, 88–93.

Ait Ouadil, K., Idbraim, S., Bouhsine, T., Carla Bouaynaya, N., Alfergani, H., & Cliff Johnson, C. (2023). Atmospheric visibility estimation: A review of deep learning approach. *Multimedia Tools and Applications*, 1–26.

Akalin, A. (2020). *Computational genomics with r*. CRC Press.

Alizadeh, O. (2024). A review of ENSO teleconnections at present and under future global warming. *Wiley Interdisciplinary Reviews: Climate Change*, *15*(1), e861.

Andersen, H., & Cermak, J. (2015). How thermodynamic environments control stratocumulus microphysics and interactions with aerosols. *Environmental Research Letters*, *10*(2), 024004.

Andersen, H., & Cermak, J. (2018). First fully diurnal fog and low cloud satellite detection reveals life cycle in the Namib. *Atmospheric Measurement Techniques*, *11*(10), 5461–5470.

Andersen, H., Cermak, J., Fuchs, J., Knippertz, P., Gaetani, M., Quinting, J., Sippel, S., & Vogt, R. (2020). Synoptic-scale controls of fog and low-cloud variability in the Namib Desert. *Atmospheric Chemistry and Physics*, *20*(6), 3415–3438.

Andersen, H., Cermak, J., Fuchs, J., Knutti, R., & Lohmann, U. (2017). Understanding the drivers of marine liquid-water cloud occurrence and properties with global observations using neural networks. *Atmospheric Chemistry and Physics*, *17*(15), 9535–9546.

Andersen, H., Cermak, J., Solodovnik, I., Lelli, L., & Vogt, R. (2019). Spatiotemporal dynamics of fog and low clouds in the Namib unveiled with ground-and space-based observations. *Atmospheric Chemistry and Physics*, *19*(7), 4383–4392.

Bari, D., Bergot, T., & El Khlifi, M. (2015). Numerical study of a coastal fog event over Casablanca, Morocco. *Quarterly Journal of the Royal Meteorological Society*, *141*(690), 1894–1905.

Baumhauer, R. (2023). Der Geofaktor Klima. In *Die physische geographie afrikas* (pp. 77–123). Springer.

Bendix, J., Thies, B., Cermak, J., & Nauß, T. (2005). Ground fog detection from space based on MODIS daytime data—a feasibility study. *Weather and Forecasting*, *20*(6), 989–1005.

Biau, G., & Scornet, E. (2016). A random forest guided tour. *Test*, *25*, 197–227.

Breiman, L. (2001). Random forests. *Machine learning*, *45*, 5–32.

Burke, A. (2007). Plant endemism in the central Namib Desert. *EVOLUTIONARY ECOLOGY RESEARCH*, *9*(2), 283–297.

Castillo-Botón, C., Casillas-Pérez, D., Casanova-Mateo, C., Ghimire, S., Cerro-Prada, E., Gutierrez, P., Deo, R., & Salcedo-Sanz, S. (2022). Machine learning regression and classification methods for fog events prediction. *Atmospheric Research*, *272*, 106157.

Cermak, J. (2006). *Sofos: A new satellite-based operational fog observation scheme* [Doctoral dissertation, Marburg, Univ., Diss., 2006].

Cermak, J. (2018). Fog and low cloud frequency and properties from active-sensor satellite data. *Remote Sensing*, *10*(8), 1209.

Cermak, J., & Bendix, J. (2011). Detecting ground fog from space–a microphysics-based approach. *International Journal of Remote Sensing*, *32*(12), 3345–3371.

Çevik, K. (2020). Deep learning based real-time body condition score classification system. *IEEE Access*, *8*, 213950–213957.

Chen, Z., Duan, J., Kang, L., & Qiu, G. (2021). Class-imbalanced deep learning via a class-balanced ensemble. *IEEE transactions on neural networks and learning systems*, *33*(10), 5626–5640.

Chollet, F., et al. (2015). Keras.

Chollet, F. (2021). *Deep learning with Python*. Simon; Schuster.

Cieslak, D. A., & Chawla, N. V. (2008). Learning decision trees for unbalanced data. *Machine Learning and Knowledge Discovery in Databases: European Conference, ECML PKDD 2008, Antwerp, Belgium, September 15-19, 2008, Proceedings, Part I 19*, 241–256.

Couronné, R., Probst, P., & Boulesteix, A.-L. (2018). Random forest versus logistic regression: A large-scale benchmark experiment. *BMC bioinformatics*, *19*, 1–14.

Cox, D. R. (1958). The regression analysis of binary sequences. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, *20*(2), 215–232.

De Szoeke, S. P., Verlinden, K. L., Yuter, S. E., & Mechem, D. B. (2016). The time scales of variability of marine low clouds. *Journal of climate*, *29*(18), 6463–6481.

Ebner, M., Miranda, T., & Roth-Nebelsick, A. (2011). Efficient fog harvesting by Stipagrostis sabulicola (Namib dune bushman grass). *Journal of arid environments*, *75*(6), 524–531.

Eckardt, F., Soderberg, K., Coop, L., Muller, A., Vickery, K., Grandin, R., Jack, C., Kapalanga, T., & Henschel, J. (2013). The nature of moisture at Gobabeb, in the central Namib Desert. *Journal of arid environments*, *93*, 7–19.

Ellrod, G. P. (1995). Advances in the detection and analysis of fog at night using GOES multispectral infrared imagery. *Weather and Forecasting*, *10*(3), 606–619.

Evans, S. E., Dueker, M. E., Logan, J. R., & Weathers, K. C. (2019). The biology of fog: Results from coastal maine and Namib Desert reveal common drivers of fog microbial composition. *Science of the Total Environment*, *647*, 1547–1556.

Fabbian, D., De Dear, R., & Lellyett, S. (2007). Application of artificial neural network forecasts to predict fog at Canberra International Airport. *Weather and forecasting*, *22*(2), 372–381.

Friedman, J., Hastie, T., & Tibshirani, R. (2010). Regularization paths for generalized linear models via coordinate descent. *Journal of statistical software*, *33*(1), 1.

Garcia-Garcia, A., Orts-Escolano, S., Oprea, S., Villena-Martinez, V., & Garcia-Rodriguez, J. (2017). A review on deep learning techniques applied to semantic segmentation. *arXiv preprint arXiv:1704.06857*.

Gerber, H. (1981). Microstructure of a radiation fog. *Journal of atmospheric Sciences*, *38*(2), 454–458.

Gottlieb, T. R., Eckardt, F. D., Venter, Z. S., & Cramer, M. D. (2019). The contribution of fog to water and nutrient supply to arthraerua leubnitziae in the central Namib Desert, Namibia. *Journal of Arid Environments*, *161*, 35–46. https://doi.org/https://doi.org/10.1016/j.jaridenv.2018.11.002

Hachfeld, B., Jürgens, N., et al. (2000). Climate patterns and their impact on the vegetation in a fog driven desert: The central Namib Desert in Namibia. *Phytocoenologia*, *30*(3/4), 567–589.

Haensler, A., Cermak, J., Hagemann, S., & Jacob, D. (2011). Will the southern African west coast fog be affected by future climate change? Results of an initial fog projection using a regional climate model. *Erdkunde*, 261–275.

Henschel, J. R., Wassenaar, T. D., Kanandjembo, A., Louw, M. K., Neef, G., Shuuya, T., & Soderberg, K. (2019). Roots point to water sources of Welwitschia mirabilis in a hyperarid desert. *Ecohydrology*, *12*(1), e2039.

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., Nicolas, J., Peubey, C., Radu, R., Schepers, D., et al. (2020). The ERA5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society*, *146*(730), 1999–2049.

Hoerl, A. E., & Kennard, R. W. (1970). Ridge regression: Applications to nonorthogonal problems. *Technometrics*, *12*(1), 69–82.

Hosmer Jr, D. W., Lemeshow, S., & Sturdivant, R. X. (2013). *Applied logistic regression* (Vol. 398). John Wiley & Sons.

Hoyer, S., & Hamman, J. (2017). Xarray: ND labeled arrays and datasets in Python. *Journal of Open Research Software*, *5*(1).

Hutchison, K. (2002). The retrieval of cloud base heights from MODIS and three-dimensional cloud fields from NASA's EOS aqua mission. *International Journal of Remote Sensing*, *23*(24), 5249–5265.

Jia, B., Wang, Y., Yao, Y., & Xie, Y. (2015). A new indicator on the impact of large-scale circulation on wintertime particulate matter pollution over china. *Atmospheric Chemistry and Physics*, *15*(20), 11919–11929.

Juergens, N., Oldeland, J., Hachfeld, B., Erb, E., & Schultz, C. (2013). Ecology and spatial patterns of large-scale vegetation units within the central Namib Desert. *Journal of arid environments*, *93*, 59–79.

Jury, M. R. (2013). Climate trends in southern Africa. *South African Journal of Science*, *109*(1), 1–11.

Jury, M. R. (2018). Climate trends across south Africa since 1980. *Water SA*, *44*(2), 297–307.

Juvik, J. O., & Nullet, D. (1995). Comments on "A proposed standard fog collector for use in high-elevation regions". *Journal of Applied Meteorology (1988-2005)*, *34*(9), 2108–2110. Retrieved February 13, 2024, from http://www.jstor.org/stable/26187431

Kaseke, K. F., Tian, C., Wang, L., Seely, M., Vogt, R., Wassenaar, T., & Mushi, R. (2018). Fog spatial distributions over the central Namib Desert-an isotope approach.

Kaseke, K. F., & Wang, L. (2022). Reconciling the isotope-based fog classification with meteorological conditions of different fog types. *Journal of Hydrology*, *605*, 127321.

Kaseke, K. F., Wang, L., & Seely, M. K. (2017). Nonrainfall water origins and formation mechanisms. *Science Advances*, *3*(3), e1603131.

Kirasich, K., Smith, T., & Sadler, B. (2018). Random forest vs logistic regression: Binary classification for heterogeneous datasets. *SMU Data Science Review*, *1*(3), 9.

Klein, S. A. (1997). Synoptic variability of low-cloud properties and meteorological parameters in the subtropical trade wind boundary layer. *Journal of climate*, *10*(8), 2018–2039.

Klein, S. A., & Hartmann, D. L. (1993). The seasonal cycle of low stratiform clouds. *Journal of Climate*, *6*(8), 1587–1606.

Lakra, K., & Avishek, K. (2022). A review on factors influencing fog formation, classification, forecasting, detection and impacts. *Rendiconti Lincei. Scienze Fisiche e Naturali*, *33*(2), 319–353.

Lancaster, J., Lancaster, N., & Seely, M. (1984). Climate of the central Namib Desert. *Madoqua*, *1984*(1), 5–61.

Lange, O. L., Green, T. A., Meyer, A., & Zellner, H. (2007). Water relations and carbon dioxide exchange of epiphytic lichens in the Namib fog desert. *Flora-Morphology, Distribution, Functional Ecology of Plants*, *202*(6), 479–487.

Lei, J., & Guo, Z. (2020). A fog-collecting surface mimicking the Namib beetle: Its water collection efficiency and influencing factors. *Nanoscale*, *12*(13), 6921–6936.

Li, B., Wang, L., Kaseke, K. F., Vogt, R., Li, L., & Seely, M. K. (2018). The impact of fog on soil moisture dynamics in the Namib Desert. *Advances in Water resources*, *113*, 23–29.

Logan, R. F. (1960). *The central Namib Desert, south west Africa*. National Academy of Sciences, National Research Council.

Loots, S., Ritz, C. M., Schwager, M., Sehic, J., Herklotz, V., Garkava-Gustavsson, L., & Nybom, H. E. (2019). Distribution, habitat profile and genetic variability of Namibian succulent lithops ruschiorum. *Bothalia-African Biodiversity & Conservation*, *49*(1), 1–18.

Mauger, G. S., & Norris, J. R. (2010). Assessing the impact of meteorological history on subtropical cloud fraction. *Journal of climate*, *23*(11), 2926–2940.

Mitchell, D., Henschel, J. R., Hetem, R. S., Wassenaar, T. D., Strauss, W. M., Hanrahan, S. A., & Seely, M. K. (2020). Fog and fauna of the Namib Desert: Past and future. *Ecosphere*, *11*(1), e02996.

Mupambwa, H. A., Hausiku, M. K., Nciizah, A. D., & Dube, E. (2019). The unique Namib desert-coastal region and its opportunities for climate smart agriculture: A review. *Cogent Food & Agriculture*, *5*(1), 1645258.

Ng, W., Minasny, B., Mendes, W. d. S., & Demattê, J. A. M. (2020). The influence of training sample size on the accuracy of deep learning models for the prediction of soil properties with near-infrared spectroscopy data. *Soil*, *6*(2), 565–578.

Norris, J. R., Allen, R. J., Evan, A. T., Zelinka, M. D., O'Dell, C. W., & Klein, S. A. (2016). Evidence for climate change in the satellite cloud record. *Nature*, *536*(7614), 72–75.

Olivier, J., & Stockton, P. (1989). The influence of upwelling extent upon fog incidence at lüderitz, southern Africa. *International Journal of Climatology*, *9*(1), 69–75.

Park, J., Lee, Y. J., Jo, Y., Kim, J., Han, J. H., Kim, K. J., Kim, Y. T., & Kim, S. B. (2022). Spatio-temporal network for sea fog forecasting. *Sustainability*, *14*(23), 16163.

Park, J. K., & Kim, S. (2019). Three-dimensionally structured flexible fog harvesting surfaces inspired by Namib Desert beetles. *Micromachines*, *10*(3), 201.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, E. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, *12*, 2825–2830.

Pouyanfar, S., Tao, Y., Mohan, A., Tian, H., Kaseb, A. S., Gauen, K., Dailey, R., Aghajanzadeh, S., Lu, Y.-H., Chen, S.-C., et al. (2018). Dynamic sampling in convolutional neural networks for imbalanced data classification. *2018 IEEE conference on multimedia information processing and retrieval (MIPR)*, 112–117.

Preston-Whyte, R., & Tyson, P. (1988). *The atmosphere and weather of southern Africa*. Oxford University Press.

Qiao, N., Wang, L., Marais, E., & Li, F. (2022). Fog detection and estimation using CALIPSO lidar observations. *Geophysical Research Letters*, *49*(24), e2022GL101375.

Qiao, N., Zhang, L., Huang, C., Jiao, W., Maggs-Kölling, G., Marais, E., & Wang, L. (2020). Satellite observed positive impacts of fog on vegetation. *Geophysical Research Letters*, *47*(12), e2020GL088428.

Radiuk, P. M. (2017). Impact of training set batch size on the performance of convolutional neural networks for diverse datasets.

Rohde, R. F., Hoffman, M. T., Durbach, I., Venter, Z., & Jack, S. (2019). Vegetation and climate change in the pro-Namib and Namib Desert based on repeat photography: Insights into climate trends. *Journal of Arid Environments*, *165*, 119–131.

Roth-Nebelsick, A., Ebner, M., Miranda, T., Gottschalk, V., Voigt, D., Gorb, S., Stegmaier, T., Sarsour, J., Linke, M., & Konrad, W. (2012). Leaf surface structures enable the endemic Namib desert grass Stipagrostis sabulicola to irrigate itself with fog water. *Journal of the Royal Society interface*, *9*(73), 1965–1974.

Ruiz, A., & Villa, N. (2008). Storms prediction: Logistic regression vs random forest for unbalanced data. *arXiv preprint arXiv:0804.0650*.

Sandler, M., Howard, A., Zhu, M., Zhmoginov, A., & Chen, L.-C. (2018). Mobilenetv2: Inverted residuals and linear bottlenecks. *Proceedings of the IEEE conference on computer vision and pattern recognition*, 4510–4520.

Schachtschneider, K., & February, E. C. (2010). The relationship between fog, floods, groundwater and tree growth along the lower kuiseb river in the hyperarid Namib. *Journal of Arid Environments*, *74*(12), 1632–1637.

Schüller, L., Bennartz, R., Fischer, J., & Brenguier, J.-L. (2005). An algorithm for the retrieval of droplet number concentration and geometrical thickness of stratiform

marine boundary layer clouds applied to MODIS radiometric observations. *Journal of Applied Meteorology and Climatology*, *44*(1), 28–38.

Seely, M. K., & Henschel, J. R. (1998). The climatology of Namib fog. *Proceedings of the First International Conference on Fog and Fog Collection*, 19–24.

Shanyengana, E., Henschel, J., Seely, M., & Sanderson, R. (2002). Exploring fog as a supplementary water source in Namibia. *Atmospheric Research*, *64*(1-4), 251–259.

Skakun, S., Wevers, J., Brockmann, C., Doxani, G., Aleksandrov, M., Batič, M., Frantz, D., Gascon, F., Gómez-Chova, L., Hagolle, O., et al. (2022). Cloud mask intercomparison exercise (cmix): An evaluation of cloud masking algorithms for landsat 8 and sentinel-2. *Remote Sensing of Environment*, *274*, 112990.

Spirig, R. (2022). *Fog in the Namib - occurrence dynamics properties* [Doctoral dissertation, University_of_Basel].

Spirig, R., Vogt, R., Larsen, J. A., Feigenwinter, C., Wicki, A., Franceschi, J., Parlow, E., Adler, B., Kalthoff, N., Cermak, J., et al. (2019). Probing the fog life cycles in the Namib Desert. *Bulletin of the American Meteorological Society*, *100*(12), 2491–2507.

Talukdar, S., Singha, P., Mahato, S., Pal, S., Liou, Y.-A., & Rahman, A. (2020). Land-use land-cover classification by machine learning classifiers for satellite observations—a review. *Remote Sensing*, *12*(7), 1135.

Thuiller, W., Midgley, G. F., Hughes, G. O., Bomhard, B., Drew, G., Rutherford, M. C., & Woodward, F. I. (2006). Endemic species and ecosystem sensitivity to climate change in Namibia. *Global Change Biology*, *12*(5), 759–776.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, *58*(1), 267–288.

Veloso, J. V., Böhm, C., Schween, J. H., Löhnert, U., & Crewell, S. (2024). A comparative study of the atmospheric water vapor in the Atacama and Namib Desert. *Global and Planetary Change*, *232*, 104320.

Wang, L., Kaseke, K. F., Ravi, S., Jiao, W., Mushi, R., Shuuya, T., & Maggs-Kölling, G. (2019). Convergent vegetation fog and dew water use in the Namib Desert. *Ecohydrology*, *12*(7), e2130.

Wang, X., Hu, K., Wu, Y., & Zhou, W. (2023). A survey of deep learning-based lightning prediction. *Atmosphere*, *14*(11), 1698.

Wang, Z., Yuan, J., Wood, R., Chen, Y., & Tong, T. (2023). Profile-based estimated inversion strength. *Atmospheric Chemistry and Physics*, *23*(5), 3247–3266. https://doi.org/10.5194/acp-23-3247-2023

Warren-Rhodes, K. A., McKay, C. P., Boyle, L. N., Wing, M. R., Kiekebusch, E. M., Cowan, D. A., Stomeo, F., Pointing, S. B., Kaseke, K. F., Eckardt, F., et al. (2013). Physical ecology of hypolithic communities in the central Namib Desert:

The role of fog, rain, rock habitat, and light. *Journal of Geophysical Research: Biogeosciences*, *118*(4), 1451–1460.

Weischet, W., & Endlicher, W. (2000). *Regionale Klimatologie: Teil 2: Die alte Welt, Europa, Afrika, Asien*. Springer.

Wiegner, M., Mattis, I., Pattantyús-Ábrahám, M., Bravo-Aranda, J. A., Poltera, Y., Haefele, A., Hervo, M., Görsdorf, U., Leinweber, R., Gasteiger, J., et al. (2019). Aerosol backscatter profiles from ceilometers: Validation of water vapor correction in the framework of CeiLinEx2015. *Atmospheric Measurement Techniques*, *12*(1), 471–490.

Wu, M., & Chen, L. (2015). Image recognition based on deep learning. *2015 Chinese automation congress (CAC)*, 542–546.

Yu, Z., Zhang, H., Huang, J., Li, S., Zhang, S., Cheng, Y., Mao, J., Dong, X., Gao, S., Wang, S., et al. (2021). Namib desert beetle inspired special patterned fabric with programmable and gradient wettability for efficient fog harvesting. *Journal of Materials Science & Technology*, *61*, 85–92.

Yue, H., Zeng, Q., Huang, J., Guo, Z., & Liu, W. (2022). Fog collection behavior of bionic surface and large fog collector: A review. *Advances in Colloid and Interface Science*, *300*, 102583.

Zhong, L., Zhu, H., Wu, Y., & Guo, Z. (2018). Understanding how surface chemistry and topography enhance fog harvesting based on the superwetting surface with patterned hemispherical bulges. *Journal of colloid and interface science*, *525*, 234–242.

Zou, H., & Hastie, T. (2005). Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, *67*(2), 301–320.

Appendix

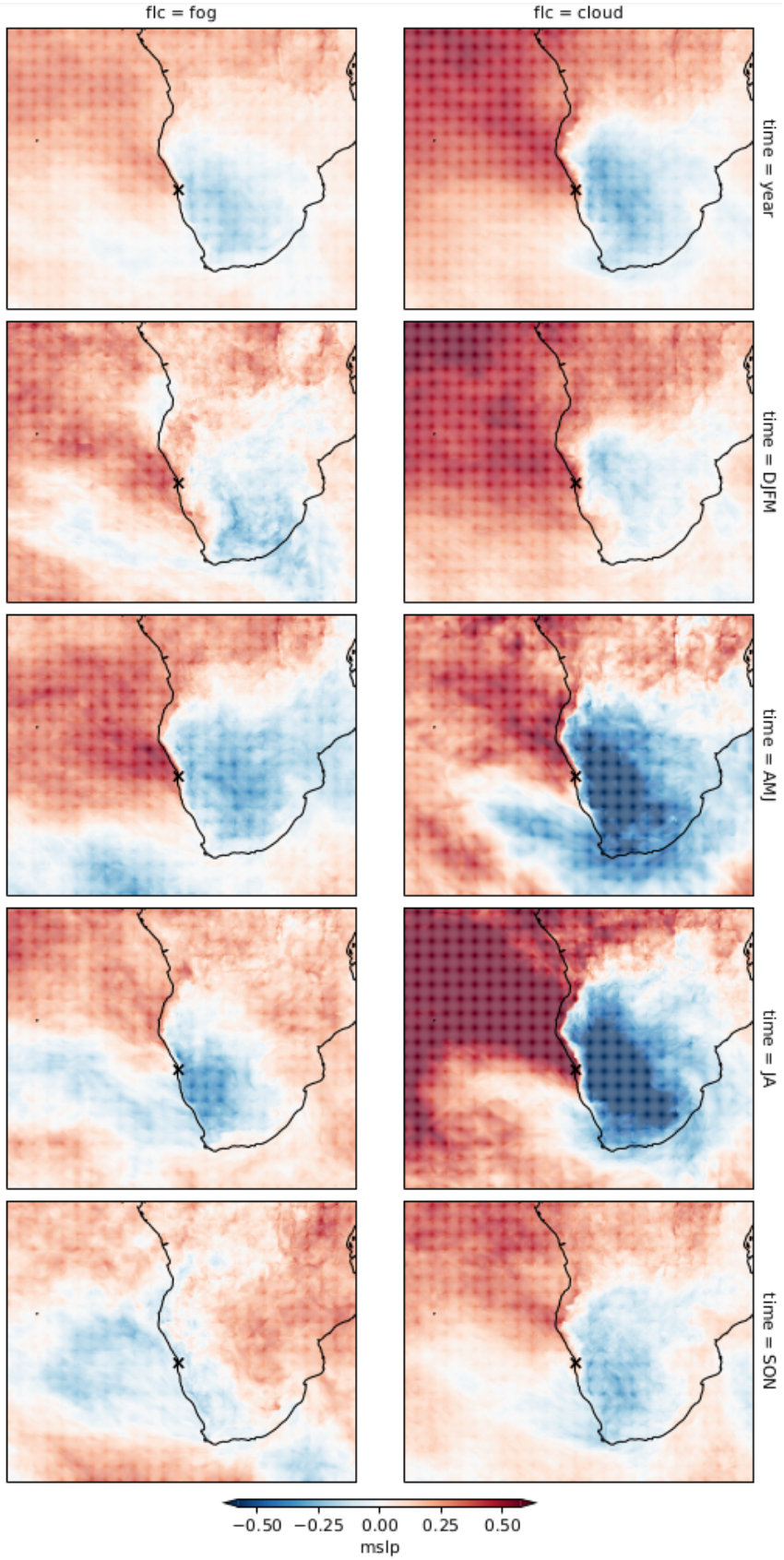


Figure 6.1: Median composites of ERA5 median sea level pressure anomaly (MSLP) on fog days (left) and low cloud days (right), over the whole available time series (2014-07 to 2020-12) and broken down to seasons (upper case letters denote the month, e.g. AMJ = April, May, June). x indicates Coastal Met, Namibia, where the fog or low cloud occurred.

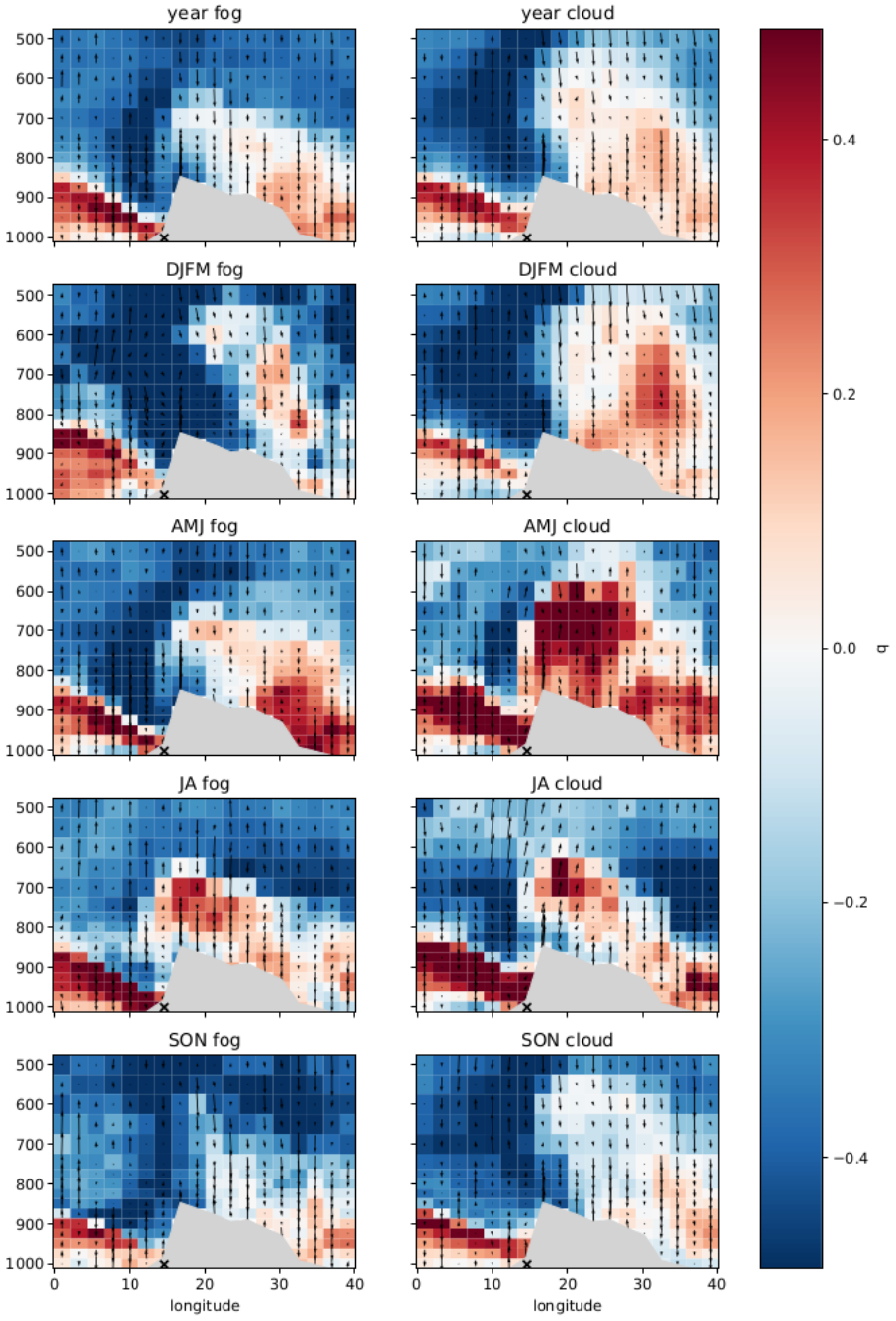


Figure 6.2: Curtain plots of median specific humidity anomaly (q) on fog days (left) and low cloud days (rigth), over the whole available time series (2014-07 to 2020-12) and broken down to seasons (upper case letters denote the month, e.g. AMJ = April, May, June). The y-axis is elevation in pressure levels (hPa). x indicates Coastal Met, Namibia, where the fog or low cloud occurred. Grey area indicates the landmass.

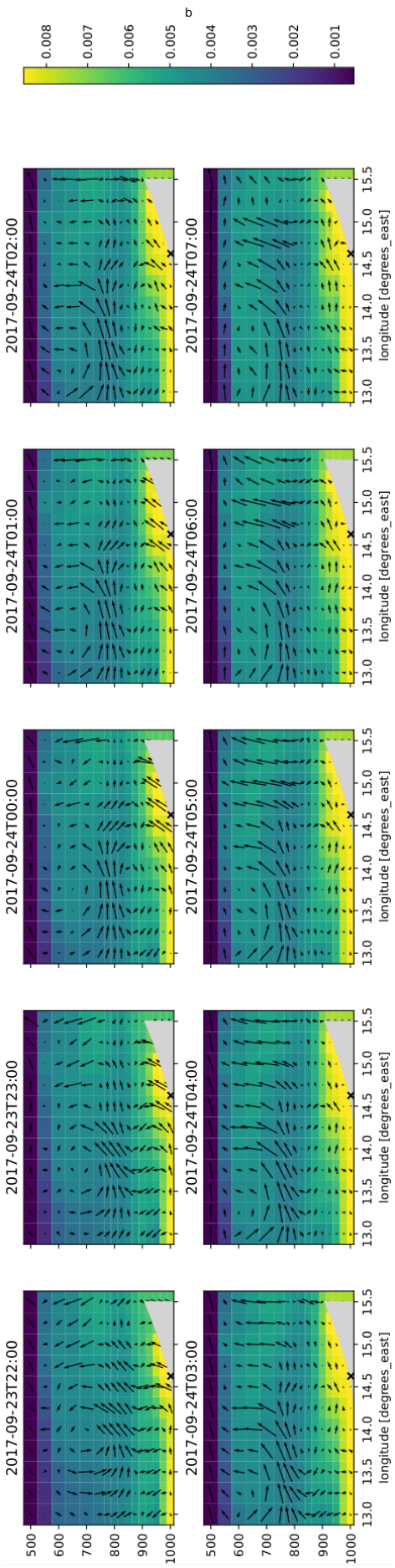


Figure 6.3: Curtain plots of median specific humidity (q in kg/kg) during the fog event 2017-09-22 / 23 at Coastal Met, Namibia (x), showing the surface wind advecting stratus clouds to the coast. The y-axis is elevation in pressure levels (hPa). Grey area indicates the landmass, quivers denote the wind (vertical component multiplied by 20).

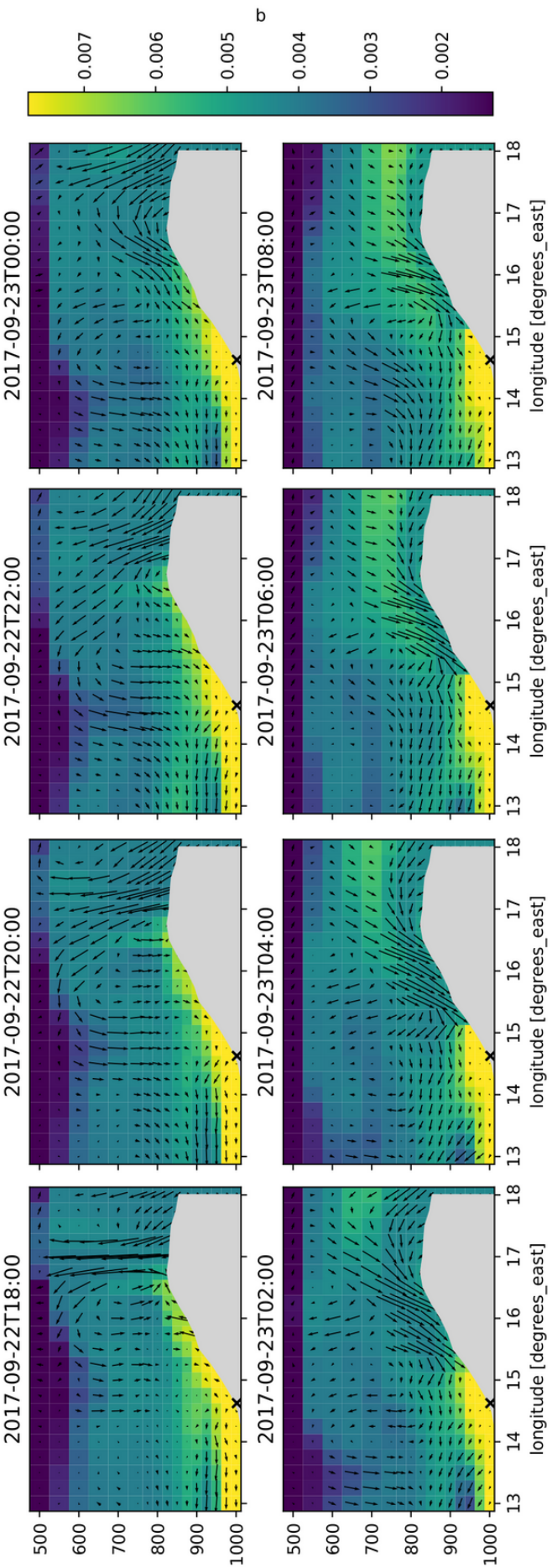


Figure 6.4: Curtain plots of median specific humidity (q in kg/kg) during the fog event 2017-09-22 / 23 at Coastal Met, Namibia (x), showing the thermal air uplift over the Great Escarpment retreating inland. The y-axis is elevation in pressure levels (hPa). Grey area indicates the landmass, quivers denote the wind (vertical component multiplied by 20).

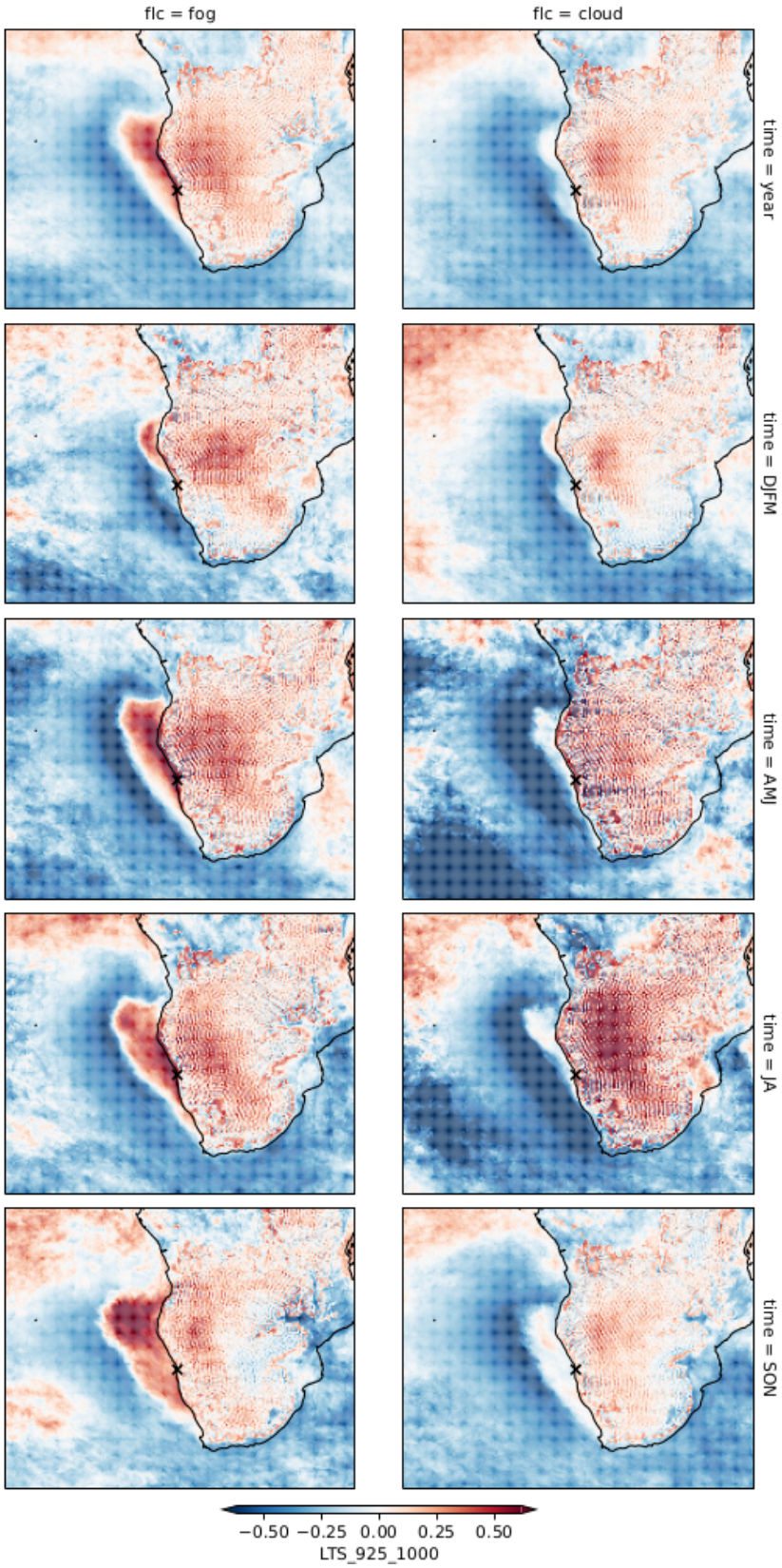


Figure 6.5: Median anomaly composites of the lower tropospheric stability (LTS) calculated with pressure levels 925 and 1000 hPa (Namib lower tropospheric stability, NLTS), on fog days (left) and low cloud days (right) and for the whole data and broken down to seasons (upper case letters denote the month, e.g. AMJ = April, May, June). x indicates Coastal Met, Namibia, where the fog or low cloud occurred.

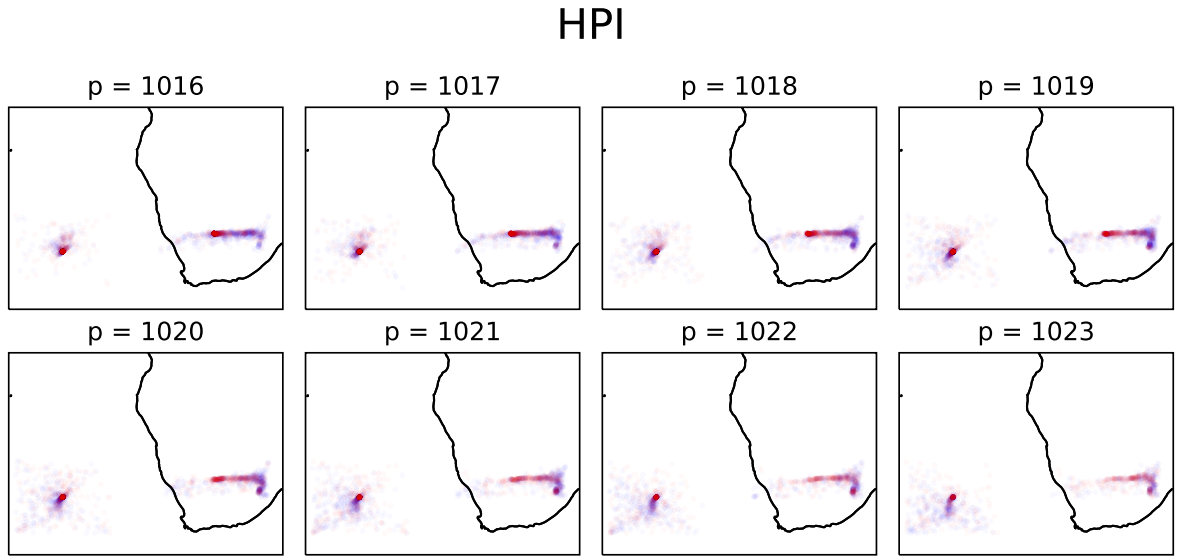


Figure 6.6: High Pressure Index (HPI) calculated with different border isobars, for South Atlantic High and Continental High, during fog nights (red) and low cloud nights (blue) at Coastal Met, Namibia.

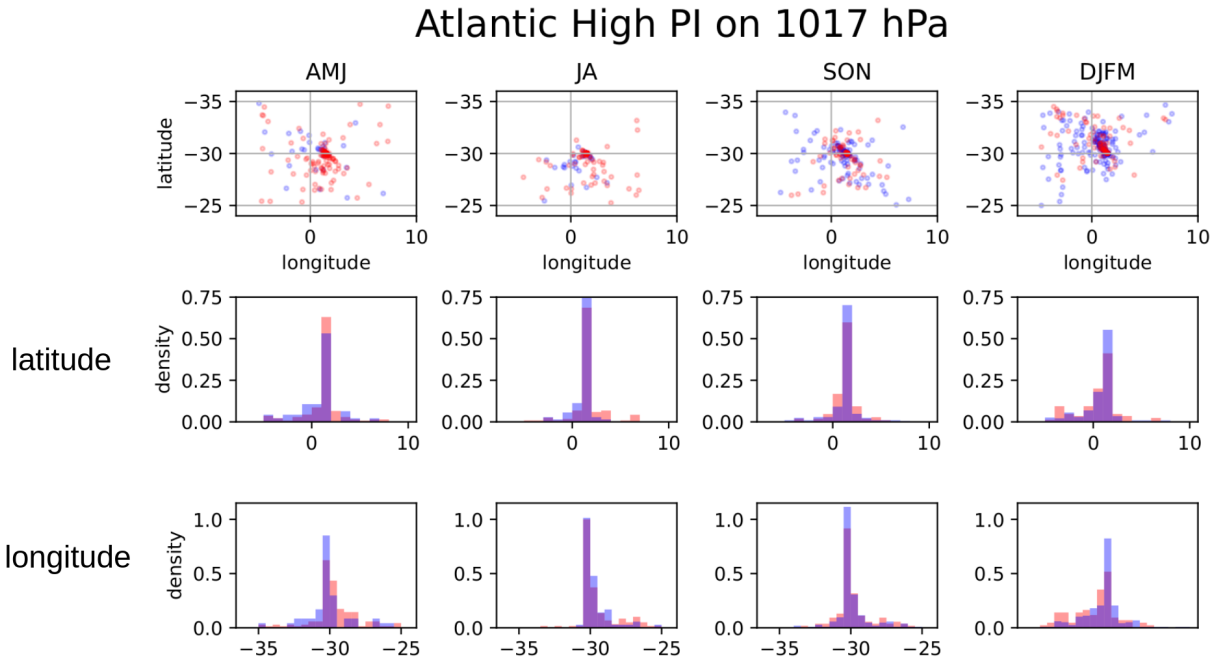


Figure 6.7: South Atlantic High Pressure Index (SAHPI) calculated with border isobar 1021 hPa during fog nights (red) and low cloud nights (blue) at Coastal Met, Namibia. Histograms show the distributions of longitudes and latitudes.

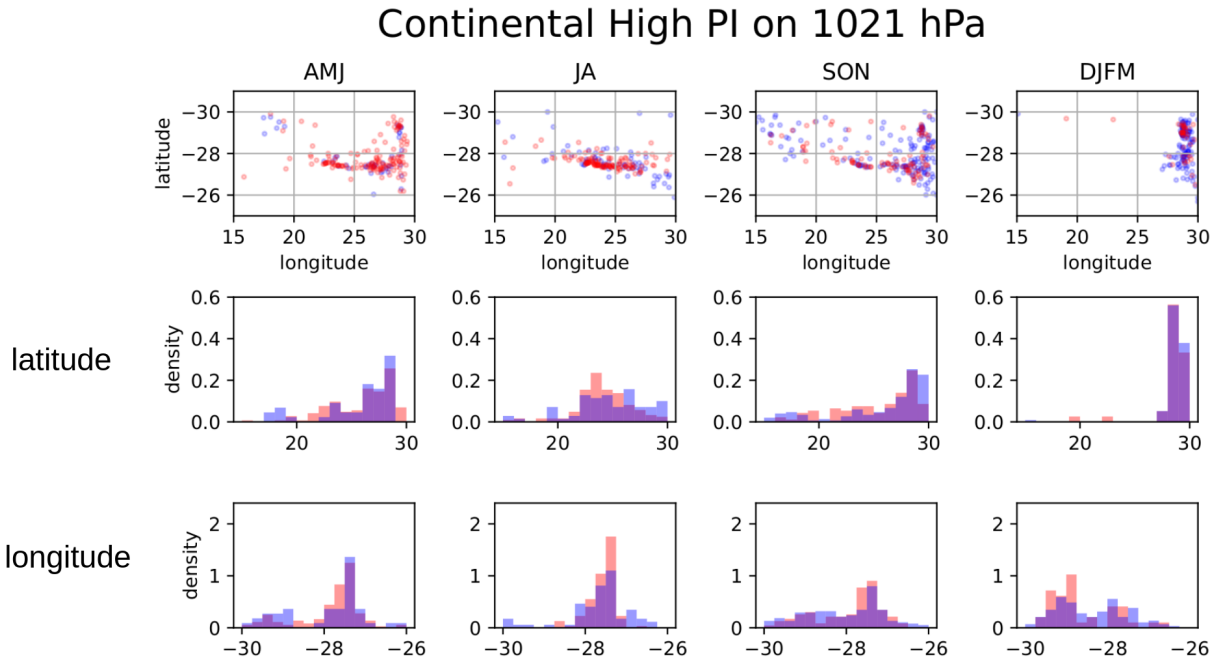


Figure 6.8: High Pressure Index calculated with different border isobars, for South Atlantic High and Continental High, during fog nights (red) and low cloud nights (blue) at Coastal Met, Namibia.