GOETHE UNIVERSITÄT FRANKFURT AM MAIN

Master Thesis

M.Sc. Atmospheric and Climate Sciences

# Climatology of Amazonian Squall Lines derived by Algorithmic Detection and Tracking of Coherent Features

Frederik Luis Vieira Fischer

19.01.2026

*Supervisors:*
Prof. Anna Possner
Dr. Lianet Hernándes Pardo

Institute for Atmospheric and Environmental Sciences,
Goethe University Frankfurt

# Declaration

I hereby declare that I have written this thesis independently and have not used any sources other than those listed in the bibliography. I have adhered to the rules of good scientific practice.

All passages that are taken verbatim or in substance from published or unpublished sources are identified as such.

The drawings or illustrations in this paper have been created by me or are accompanied by appropriate references.

The use of artificial intelligence (AI), specifically of large language models (LLMs), was limited to the improvement of phrasing.

This thesis has not been submitted in the same or a similar form to any other examination authority.

| | |
|---|---|
| Place, Date | Signature |

# Acknowledgements

I would like to thank Dr. Lianet Hernándes Pardo for supervising my Master thesis, for all the productive meetings, the encouragement, and for taking the time to manually label an entire year of squall lines.

I would like to thank Prof. Anna Possner for giving me the opportunity to write my master's thesis in her group. I am thankful for the guidance and challenging questions, which encouraged me to examine problems more critically.

I would like to thank the Cloud Physics in the Climate System group as a whole for their valuable input, encouragement and engaging discussions.

Finally, I would like to thank my family for supporting me in my scientific endeavours and for helping me with the proofreading of this thesis.

# Abstract

The Amazon River basin provides an abundance of moisture, a warm climate and a high tropopause, which are favourable conditions for convective storms. Through manual detection it has been established that over 100 squall lines form yearly at the coast near the Amazon River delta. Half of them dissipate at the coast, while the other half propagates into the Amazon basin and a fifth cover distances over 400 km and some even exceed travel distances of $2 \cdot 10^3$ km.

In this study I develop an algorithm to automatically detect and track squall lines based on top of the atmosphere brightness temperature data from the geostationary satellite GOES-16. The algorithm relies on a threshold based feature extraction and density based segmentation with DBSCAN, complemented by subdividing with watershed segmentation. Tracking was performed with a maximum overlap criterion and followed up by squall line classification with Gaussian kernel density functions, which were created with a small manually labelled dataset.

The resulting multi-year track climatology of squall lines and mesoscale convective systems includes a plethora of system properties like length, area, lifetime, and growth rate. The tracking shows that during wet season the basin occurring squall lines have a strong link to coastal squall lines, suggesting direct succession in many cases. The occurrence of coastal squall lines is much lower during dry season, with most squall lines occurring in the basin. Squall line regeneration after partial dissipation are mostly located in the central basin with close proximity to regions with squall line initiation.

# Contents

# List of Figures

# List of Tables

# Part 1
# Introduction

Squall lines (SLs) are large line shaped convective storms, home to both the tropics and mid-latitudes, where they can cause severe and disastrous weather. In their weaker appearances they still cause significant amounts of rain and thunder. Their size ranges from hundreds to often more than a thousand kilometres in length and their lifetime can range from hours to several days (Greco et al., 1990). Exhibiting directional movement, SLs are self-sustaining and propagate hundreds to thousands of kilometres, leading them to cover swaths which may far exceed a thousand by a thousand kilometres in size.[1]

Mid-latitude SLs have been subject to a lot of research, since they can be embedded with supercells and cause strong tornados (Markowski and Richardson, 2010). Their tropical counterpart is less well studied, however the basic mechanisms that govern them are the same. Driven by low level wind shear balanced with a cold pool, their theoretic basis has been explained by the RKW-Theory (see Section 3.1.1). The formation and forcing is region specific and differ significantly between the tropics and mid-latitudes, but also within each climate.[1]

Mid-latitude SLs can, for example, form embedded within a low pressure system and be driven by its cold front, propagating eastwards (Markowski and Richardson, 2010). Tropical Sahelian SLs form scattered over land and propagate eastwards onto the Atlantic ocean, if their lifetime permits it (Desbois et al., 1988). Amazonian SLs usually form at the coast and propagate inland over the Amazon River basin towards the Andes Mountain Range (Rickenbach, 2004).

The Amazonian SLs, which I am focussing on in this study, show a broad range of lifetimes and propagation distances (Alcântara et al., 2011). Their coastal formation is usually caused by the sea breeze and once detached from the coast, these SLs have been observed to pass over the central Amazon basin. Here they have been observed to cause nocturnal precipitation, leading to a secondary precipitation maximum during the night (Rickenbach, 2004). Obsevations

[1] Paragraph taken from Vieira Fischer (2025) with some adaptations.

also revealed, that SLs may undergo regeneration events in the Amazon basin. Two types of regeneration events have been observed, where a SL seemingly "disappears" and later re-forms. Firstly, a SL may have weak convection, a loss of organisation or a reduction in size, leading to a partial or full dissipation during the night or early morning. Shortly followed by reorganisation into a fully formed SL (Cohen, 1989; Rickenbach, 2004). The second type of regeneration occurs when a SL fully dissipates during the night and a new SL forms the following day with the onset of the afternoon convection. This new SL originates approximately where the old one would have been at that time of day, had it not dissipated (Kousky, 1980; Alcântara et al., 2011). Neither of these regeneration events has been studied in depth, although Anselmo et al. (2020) links the first type to the onset or strengthening of the Amazonian low level jet in the early morning hours.[1]

Because of discontinuities in their structure and other types of organised convection in the tropics, tropical SLs are notoriously hard to detect and track with widely adopted automatic algorithms. Similarities to other large convective systems may even cause disagreeing classifications when performed manually on the basis of individual satellite pictures. Additional context in the form of a time series containing the evolution or lifecycle of a SL is beneficial for its identification. In the Amazonas region many studies either do not differentiate between different types of organised mesoscale convection (Anselmo et al., 2021), or study mesoscale convective clusters with SLs being sub-group thereof (Greco et al., 1990; Rickenbach, 2004; Oliveira and Oyama, 2019; Anselmo et al., 2020; Sousa et al., 2021). These studies apply subjective manual identification of the SLs, which is not particularly viable for long periods of time with high temporal resolution. Manual identification also limits the data retrieved from each SL case, which can be circumvented by hybrid approaches.[1]

It seems that there is currently no method to automatically detect and track SLs, like there is for other cloud types (see Section 3.3). The scientific community knows from case studies that squall lines propagate from the coast to the basin, but statistical analyses focus more on the coastal region and omit the central basin. Regeneration has similarly been observed in case studies, but there is limited knowledge about the geographic distribution of regeneration events.

[1] Paragraph taken from Vieira Fischer (2025) with some adaptations.

# Part 2
# Objectives

This study is motivated by the current state of the research about Amazonian SLs and their regeneration, as described above. Its primary focus is to generate an extensive climatological dataset of SLs spanning from the coastal regions into the central Amazon basin and its analysis. The thesis will address the following two key objectives:[1]

I. A new algorithm, specifically tailored for the detection and tracking of tropical SLs, is needed to generate the proposed dataset. Such an algorithm hast to be tested for detection performance and accuracy and its results have to match the scientific consensus about tropical SLs. With many competing tracking algorithms, the new algorithm should prove the necessity of its existence by significantly outperforming other algorithms when detecting and tracking SLs.[1]

II. To gain a better understanding of Amazonian SLs, a climatological analysis will be performed with a special focus on the life cycle of SLs. Within this large dataset of SLs, I hope to observe several instances of both regeneration events. I expect to observe basin-occurring SLs with a link to prior SLs near the coast and I expect to observe SLs with stages during their life cycle, which would not be categorised as a SL if observed without context.[1]

[1] Paragraph taken from Vieira Fischer (2025) with some adaptations.

Part 3

# Theoretical Background

## 3.1 Squall Lines

Squall lines (SLs) can differ vastly in appearance, both between different occurrences in the same region and between different regions. The main commonality between all different types of SLs is their linear spacial distribution of convective cores, although this line can be curved and distorted (Markowski and Richardson, 2010). The length of this line often exceeds 1000 km and can reach 3500 km with propagation speeds ranging 10–40 $\mathrm{km\,h^{-1}}$ (Greco et al., 1990). Other sources report upper speed limits for mesoscale convective systems (MCSs) of 45 $\mathrm{km\,h^{-1}}$ (L. A. T. Machado et al., 1998) or for clouds in the tropics of 80 $\mathrm{km\,h^{-1}}$ (Mathon and Laurent, 2001).[1]

SLs usually consist of series of convective cores, which exhibit a mostly two-dimensional structure, contrary to i.e. super cells, which have a complex three-dimensional structure. This means that a SL does not necessitate air movement or interactions along its linear structure, only needing air movement and interactions horizontally and vertically across the SL which is covered by the RKW-Theory (see Section 3.1.1). Individual cross sections of the SL are therefore mostly independent of each other. Each cross section consists of an updraft, which "rides" the edge of a cold pool. However, a SL can also have supercells embedded within its structure, which need airflow and interactions along the SL (Fovell and Tan, 1998).[1]

SLs have large stratiform regions which are advected from their convective cores. These stratiform regions can also differ significantly between SLs (Parker and Johnson, 2000). Depending on the strength and direction of the wind strength and shear, SLs can form trailing, leading, and parallel stratiform regions, as depicted in Figure 3.1. The trailing stratiform type is the most common with strong winds from the front to the rear and stratiform rain falling behind the SL (Parker and Johnson, 2004).[1]

[1] Paragraph taken from Vieira Fischer (2025) with some adaptations.

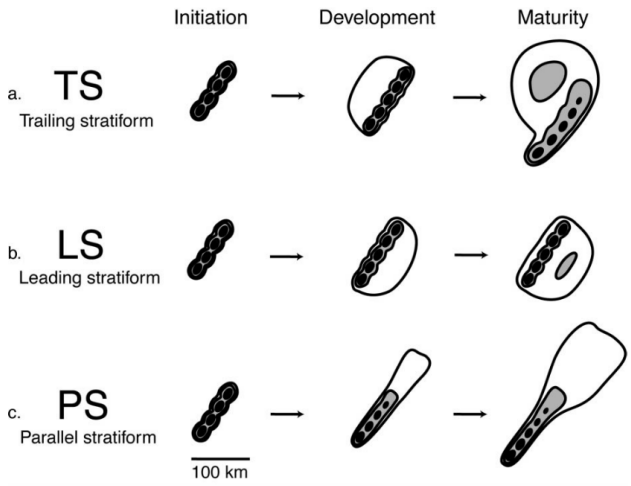


Figure 3.1: Schematic reflectivity drawing of idealized life cycles for three linear MCS archetypes: (a) TS, (b) LS, and (c) PS. (Figure and Caption adapted from Figure 4 in Parker and Johnson, 2000)

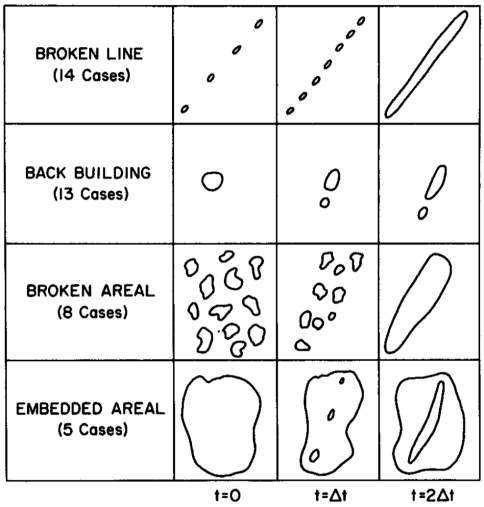


Figure 3.2: Idealized depiction of squall-line formation. (Figure and Caption adapted from Figure 1 in Bluestein and Jain, 1985)

SLs can form in different ways, as depicted in Figure 3.2 (Bluestein and Jain, 1985). They may originate from a broken line of almost simultaneously formed cells, with new cells forming in between old ones as the SL strengthens. SLs may also form as a single cell, which forms more numerous and line-shaped cells as they propagate forward, thus building back upon each other. Lastly, they may also form from or within an area of broken up convection or embedded within an area of stratiform precipitation.[1]

### 3.1.1 RKW-Theory

The RKW-Theory, named after Rotunno, Klemp and Weisman (1988) in its revised form by Weisman and Rotunno (2004), represents the current understanding of the basic mechanisms that govern SLs. According to this theory long-lived SLs form and persist at the shear-ward flank of a surface cold pool (Rotunno et al., 1988). The vorticity generated by the density current at the front between the cold pool and the moist-warm ambient air is balanced by the vorticity from the low level wind shear. This balance leads to lifting and generation of convective cells, which is shown in Figure 3.3. Although these convective cells may each be short-lived, the interaction between cold pool and wind shear creates new cells on the shear-ward side of the old cell, resulting in an overall long-lived multicellular convective storm. The subsidence caused by

[1] Paragraph taken from Vieira Fischer (2025) with some adaptations.

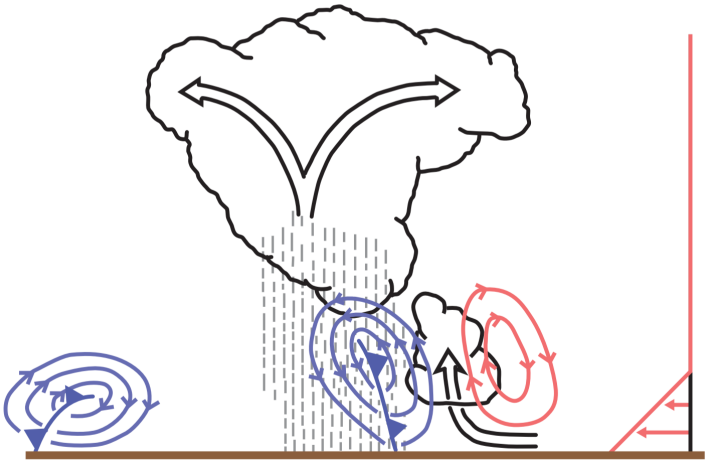

Figure 3.3: Interaction between a cold pool (blue) and low level wind shear (red) and their corresponding vorticities (Figure and Caption adapted from Figure 3 in Bryan et al., 2012).

each cell and its precipitation feeds further into the cold pool, sustaining it. As a result, SLs do not move solely propelled by the wind, but also propagate forward by forming new cells.[1]

In an environment with a weaker low level wind shear a slow-moving multicellular SL will form. With higher wind shear a faster-moving unicellular SL will form (Fovell and Ogura, 1989). In the latter case, newly forming cells in front of the SL will merge with the old cell and keep it alive. A rear inflow current helps sustaining the cold pool in faster-moving SLs and further balances the vorticity. Shear in higher levels of the atmosphere may be a deciding factor if the SL will be of a trailing, leading, or parallel stratiform nature (Fovell and Tan, 1998). In the case of the leading stratiform SL, multicellularity is of especial importance for longevity (Parker and Johnson, 2004).[1]

### 3.1.2 Mid-latitude and tropical Squall Lines

SLs in the mid-latitudes have generally been well studied. These often result from forced convergence by a frontal type system and usually exhibit a contiguous line of strong convection (Markowski and Richardson, 2010). As part of the mesoscale weather, they generally move eastwards. In the case of bow and comma echoes and line echo wave patterns, mid-latitude SLs are associated with tornados. These shapes form under the influenced of the significant effect the Coriolis force has in the mid-latitudes.[1]

[1] Paragraph taken from Vieira Fischer (2025) with some adaptations.

In contrast, tropical SLs have been studied a lot less. They are usually discontiguous due to less strong forcing, and travel westwards as part of the trade wind-dominated mesoscale weather. As the tropopause is higher in the tropics and the air has an abundance of moisture and thermal energy, the tropical SL can be of similar strength as their mid-latitude counterparts. Tropical SLs can be triggered by forced convection due to orography or airmass advection, for example the sea breeze front (Kousky, 1980; Anselmo et al., 2020). There are rare occasions where cold fronts reach the tropics and cause eastbound SLs with a high destructive potential (Negrón-Juárez et al., 2010).[1]

## 3.2 The Amazon Basin

The Amazon basin is home to Earth's largest tropical rainforest, exhibiting warm and moist atmospheric conditions. The convection shows a strong diurnal cycle with a maximum during the early afternoon hours and a minimum during the late morning, due to the nocturnal stabilisation of the boundary layer (L. A. T. Machado et al., 2002; Oliveira and Oyama, 2019). In the east the region is semi-open to the Atlantic ocean, and to the west it is enclosed by the Andes Mountain Range. The local wind systems include a low level jet over the Amazon basin, which originates at the coast in the region of the Amazon River delta and blows inland, with a maximum over the central Amazon as can be seen in Figure 3.4. When the Amazonian low level jet occurs, it is stronger during the night, with its maximum in the early morning, and weaker during the day. This nocturnal strengthening of the jet can be linked to an increase in the density of cloud cluster area (Anselmo et al., 2020).

Generated by the seas breeze, over 100 SLs form every year at the north-eastern coast around the Amazon River delta, although about half of them dissipate when still in the coastal region. The other half travels west-south-west, with a maximum during March, April, and May and minimum during October, November, and December (Cohen, 1989; Alcântara et al., 2011; Sousa et al., 2021). About 20–30 % of overall SL cases cover distances greater than 400 km, with some SLs having been reported to travel far into the Amazon basin, dissipating 2000–3000 km away from where they initiated (Cohen et al., 1995; Rickenbach, 2004). The latter type is observable throughout the whole year (Cohen, 1989). SLs account for about 40 % of total precipitation in the central Amazon (Greco et al., 1990).[1]

[1] Paragraph taken from Vieira Fischer (2025) with some adaptations.

Precipitation events during the night and early morning in the central basin are linked to coastal SLs propagating deep into the Amazon basin (L. A. T. Machado et al., 2002; Rickenbach, 2004). However, these SLs are often not continuously detectable throughout the time period spanning from initiation at the coast until causing a nocturnal rain event. SLs have been observed to undergo severe minima in convective activity, which I will categorise into two types for the purpose of this study:

**Type 1** regeneration events show a nocturnal minimum in activity and overall signs of dissipation with no clearly identifiable SL structure. Some convective system however remains and subsequently regains strength and re-forms into a SL. This kind of regeneration has been reported by Cohen (1989) and Rickenbach (2004).

**Type 2** regeneration events have the coastal SL completely dissipate, with no remaining convective activity. Later a new SL forms on the path of the coastal SL earlier. Even though the squall line dissipated and another initiated, they show clear signs of continuity. This kind of regeneration has been reported by Kousky (1980) and Alcântara et al. (2011).

The first type could be linked to the low level jet which could provide additional wind shear for nocturnal SL (Anselmo et al., 2020).

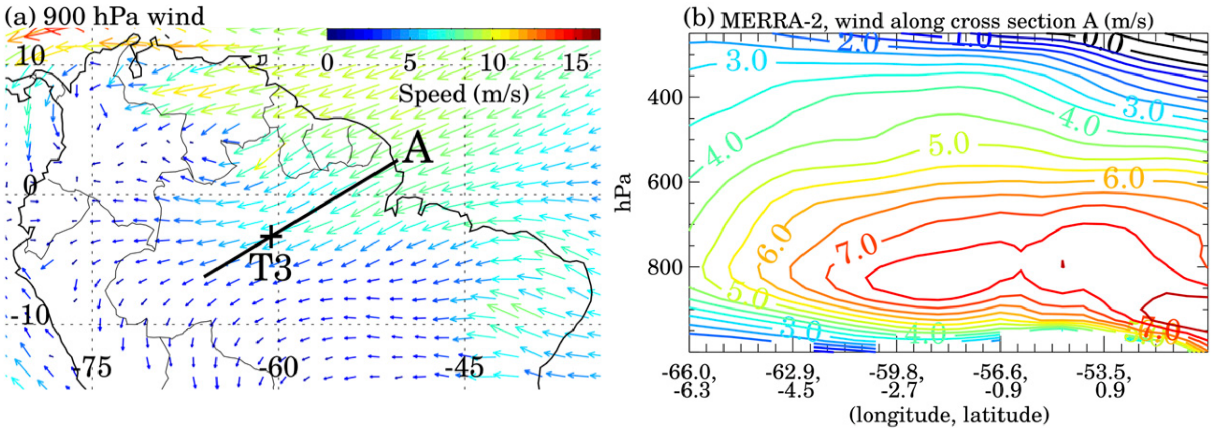


Figure 3.4: March, April, and May (MAM) horizontal winds ($\mathrm{m\,s^{-1}}$) from the Modern-Era Retrospective Analysis for Research and Applications, Version 2 (MERRA-2) reanalysis data over the Amazon basin and along the Amazon River showing the amazonian low level jet (Figure and Caption adapted from Figure 1 and 3 in Anselmo et al., 2020).

## 3.3 Identification and Tracking

Cloud tracking is one of the common approaches to study the temporal changes of clouds or cloud decks. Even before automatic detection by algorithms was feasible, manual tracking of SLs was performed on time series of satellite images (Desbois et al., 1988). Later semi-automated procedures were incorporated for tracking MCSs, utilising computers to preprocess the data and manual identification and tracking by the researches (L. A. T. Machado et al., 1998). Similar techniques have still been used in recent studies about Amazonian coastal SL initiation, due to the difficulties differentiating between mesoscale convective clusters and SLs (Oliveira and Oyama, 2019; Sousa et al., 2021).[1]

For more generalised tracking of deep convection there is currently a large number of trackers available. Feng et al. (2025) recently compared 10 MCS tracking algorithms using model output data from different DYAMOND (dynamics of the atmospheric general circulation modelled on non-hydrostatic domains) simulations. These trackers differ vastly both in their algorithmic approach and tracking performance. Especially the number of detected systems can differ by several hundred percent between different trackers. System lifetime and cloud shield size also differs significantly between trackers. The trackers also differ in the required input data, with some being highly dependant on specific input fields (usually brightness temperature and/or rain rates), while other trackers will take any input data. Each of the following sections describe a tracker, which I chose to get a general overview over different tracking approaches.[1]

### 3.3.1 ForTraCC

The Forecast and Tracking the Evolution of Cloud Clusters (ForTraCC) algorithm uses the widely adopted method of identifying clouds by applying a minimum cloud brightness temperature threshold. Contiguous areas are then considered to be a single cloud. After the identification step, tracking is performed by checking the overlap of cloudy areas in two adjacent images. If the overlap exceeds a minimum of 25%, either forward or backward in time, both areas are considered to be the same cloud at different times. If more than one cloudy area overlaps with another in the earlier or later time step, it is considered to be a split or merge. The area with the largest overlap contributes to the continuation of the trajectory, the split or merged parts are handled as beginning or ending trajectories. Based upon the minimum overlap and the cluster

[1] Paragraph taken from Vieira Fischer (2025) with some adaptations.

movement speed, the algorithm has a minimum cluster size for tracking, which filters out the early formation and late dissipation stages. ForTraCC further includes a nowcasting module (Mathon and Laurent, 2001; Vila et al., 2008).[1]

A new python based ForTraCC (pyForTraCC) was developed recently. The new pyForTraCC algorithm has some changes compared to the old ForTraCC:

- Two inner thresholds were added.
- The clustering by contiguous pixels was replaced with Density Based Spatial Clustering of Applications with Noise (DBSCAN).
- Minimum overlap for tracking was reduced to 10%.
- The overlap tracking now uses polygons which approximate the outer boundaries of the cloud areas.

According to the authors, the latter was done to smooth the movement of the centroid. Since merges and splits can still lead to unwanted jumps in the location of the centroid, the movement vectors are corrected for splits, merges, and based on the cores identified by the inner thresholds, which is important for the nowcasting module of the algorithm (Leal et al., 2022).[1]

### 3.3.2 tobac

The Tracking and Object-Based Analysis of Clouds (tobac) algorithm also does identification and tracking in two distinct steps. The identification of cloudy areas is divided into two parts and uses multiple thresholds. Features are identified as contiguous areas starting with the outermost threshold and a centroid is calculated for each. If an area encapsulates more than one area of an inner threshold, the centroid will be replaced by the centroids of the contiguous areas defined by the inner threshold. The data for the first part can be preprocessed and smoothed by tobac as needed. The second part of the identification step is a watershed segmentation, which attributes an area to each centroid based on the data. The first and second part of the identification step can use different datasets and the second part may use a three dimensional dataset. A minimum cloud size filter can be applied and the segmentation is gap allowing in order to connect these small cloudy areas to bigger areas. The tracking step is done by linking the centroids in time and does not take into account any splits or merges. The tracking allows both for random and

[1] Paragraph taken from Vieira Fischer (2025) with some adaptations.

for predictive linking, with the latter taking the previous movement of the centroid into account (Heikenfeld et al., 2019). A recent update of the tobac algorithm included the addition of 3D centroid detection and tracking as well as a post-processing procedure, that handles splits and merges (Sokolowsky et al., 2024).[1]

### 3.3.3 TOOCAN

The Tracking Of Organised Convection Algorithm through a 3-D segmantatioN (TOOCAN) differs significantly from the two described previously, as detection and tracking is accomplished in a single step. The time domain is treated as if it is a spacial domain, resulting in a three dimensional, spatio-temporal dataset. Within this space multiple thresholds are used, starting with the innermost and then iterating outwards. At each iteration step "seeds" are distributed in the spatio-temporal domain within not yet seeded contiguous volumes confined by the threshold. The seeds are then be expanded by watershed segmentation before moving to the next threshold. This should result in the tracks of individual convective cells and, according to the authors, no handling of splits and merges is necessary (Fiolleau and Roca, 2013; Fiolleau and Roca, 2024).[1]

[1] Paragraph taken from Vieira Fischer (2025) with some adaptations.

Part 4

# Data and Methodology

In this study I will generate a dataset of deep convective systems over the Amazon basin with the help of automated algorithmic cloud tracking. The individual system tracks will be classified into SLs, MCSs, and other systems. This dataset will then be evaluated statistically to produce a climatology and searched for SL regeneration events.[1]

## 4.1 Region and Dataset

For the purpose of studying Amazonian SLs, I chose the region of interest from 5 to −10 °N and from 45 to 70 °W (Figure 4.1 and 4.2). This covers the Amazon basin, with the western bound excluding the Andes Mountain Range and the eastern bound including the Amazon River delta, parts of the Brazilian coast and a bit of the Atlantic ocean. In between the central Amazon basin and the coast there is a region with higher elevation, through which the Amazon River cuts through to reach its delta. Both the north-west and south-east contain mountainous regions.

Capturing the evolution of SLs over their lifetimes requires data with a high temporal resolution, since individual convective updrafts have mean lifetimes of about 70 min Gupta et al., 2024. Time resolutions of 30 min are at the upper limit, with tracking benefitting from better time resolution. The horizontal resolution is preferably high enough to capture the cold overshooting tops of individual updrafts, with the updrafts having mean sizes in the range of 20–70 $km^2$.

The dataset produced by Geostationary Operational Environmental Satellite 16 (GOES-16) during its time as operational GOES-East satellite at 0 °N and 75,2 °W fulfills these prerequisites. GOES-16 was operational and positioned at 75,2 °W from 18.12.2017 to 18.03.2025, when it was nudged to 75,5 °W, and later on 07.04.2025 replaced by GOES-19 arriving at 75,2 °W (GOES-R, 2018; GOES-R, 2025). The data is available in 15 min time intervals and 10 min respectively

[1] Paragraph taken from Vieira Fischer (2025) with some adaptations.

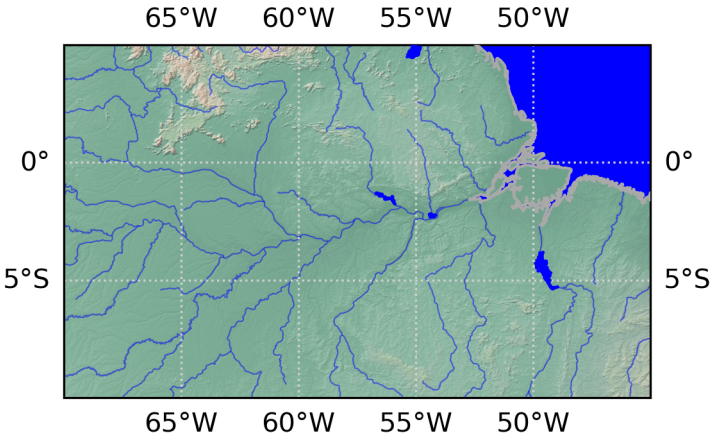


Figure 4.1: The topographic features of the domain displayed as cross-blended Hypsometric tint with shaded relief with rivers and ocean in blue (Kelso and Patterson, 2026).

starting April 2019 (see Fig. 4.3). The spatial resolution in the Amazon region is about 2 km by 2 km (GOES-R, 2020; GOES-R, 2021).

In order not to bias towards any season, I limit the data to the complete years 2018–2024 during the above mentioned timeperiod of GOES-16. During these 7 years, there are a total of 71 gaps of 1 h or longer occurred. The longest gap measures 22,5 h from the end of one to the beginning of the next scan (see Fig. 4.3). These gaps amount to a total of 10 d, 19:56:29 and are distributed reasonably well over all 7 years and all four seasons (see Fig. 4.4). The summer months June, July and August have the least gaps and only about 2 days of data missing, compared to about 2,5 to 3 days during the other seasons. Therefore the season with the least missing data coincides with the season of least convective activity in the region.

The GOES-16 satellite provides data from 16 spectral channels and a multitude of secondary retirevals, which are products of one or more chanels. For the detection of convection I chose the spectral channel with a wavelength of 10,33 µm, which covers the “clean” infrared long-wave window of the atmosphere (GOES-R and NOAA, 2017). The data is provided as top of the atmosphere (TOA) brightness temperature, which shows the observed temperatures of clouds over the full height of the atmosphere, down to the surface of the Earth, which is detected as well.

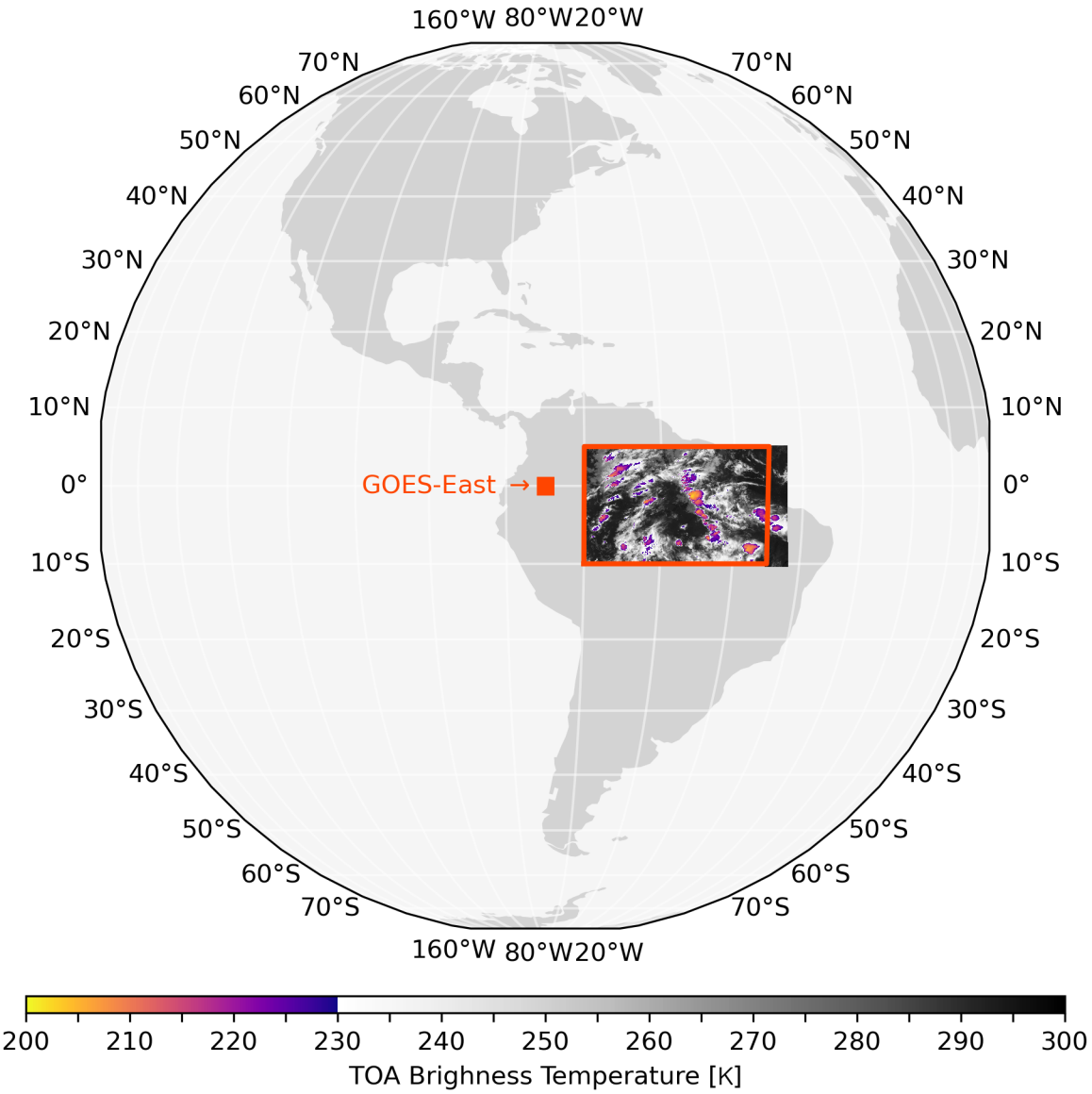


Figure 4.2: Schematic view of the GOES-East full disc scan area, with the GOES-East position at 75,2 °W, the region of interest from 5 to −10 °N and from 45 to 70 °W, and example data from 16.01.2020 02:15 UTC containing a SL.

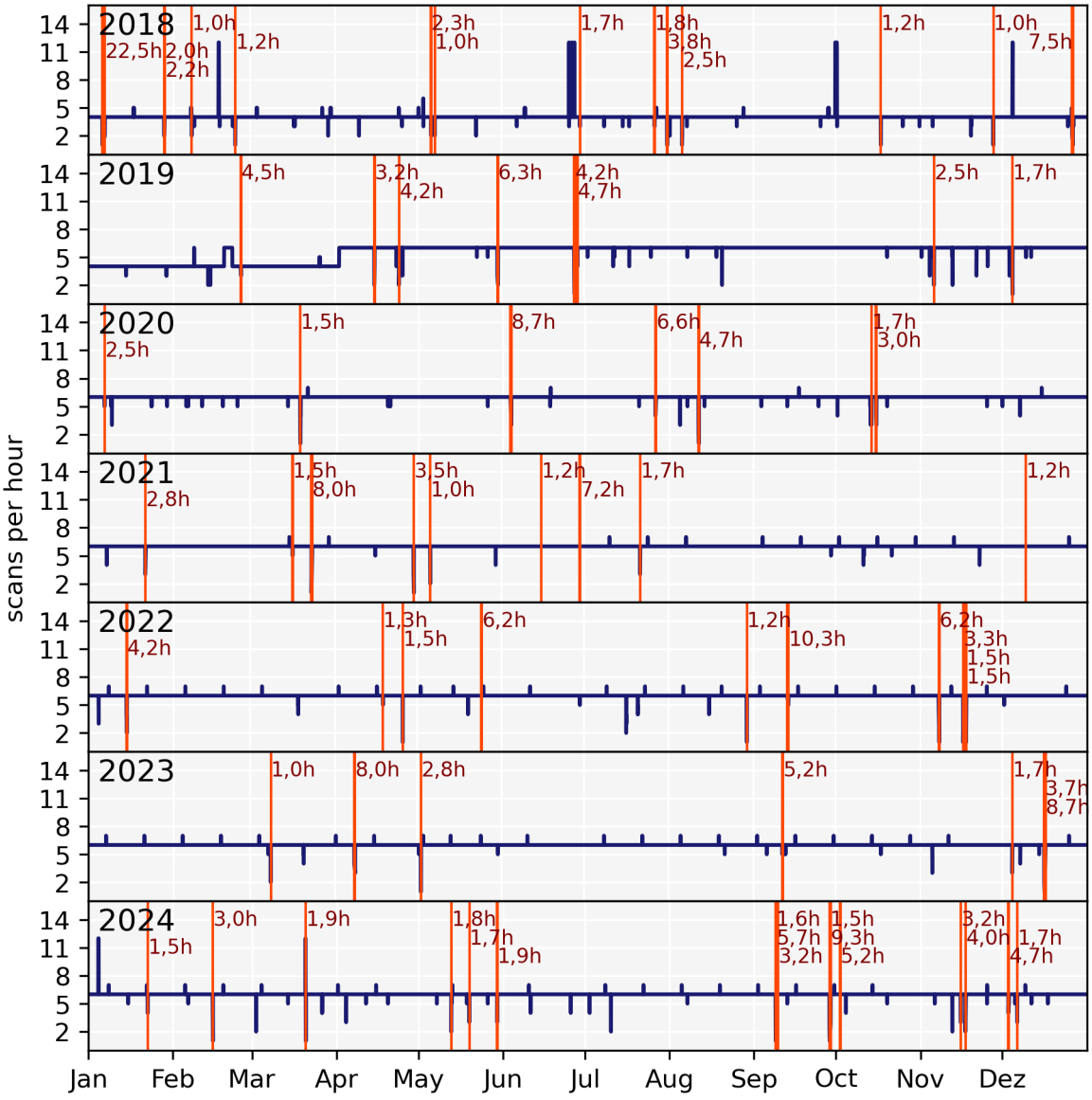


Figure 4.3: Hourly GOES-16 scan frequency for the entire time period from 2018 to 2024, with gaps with one hour or longer marked in red together with their length in hours. Gaps shorter than an hour can be seen as dips in the scan frequency.

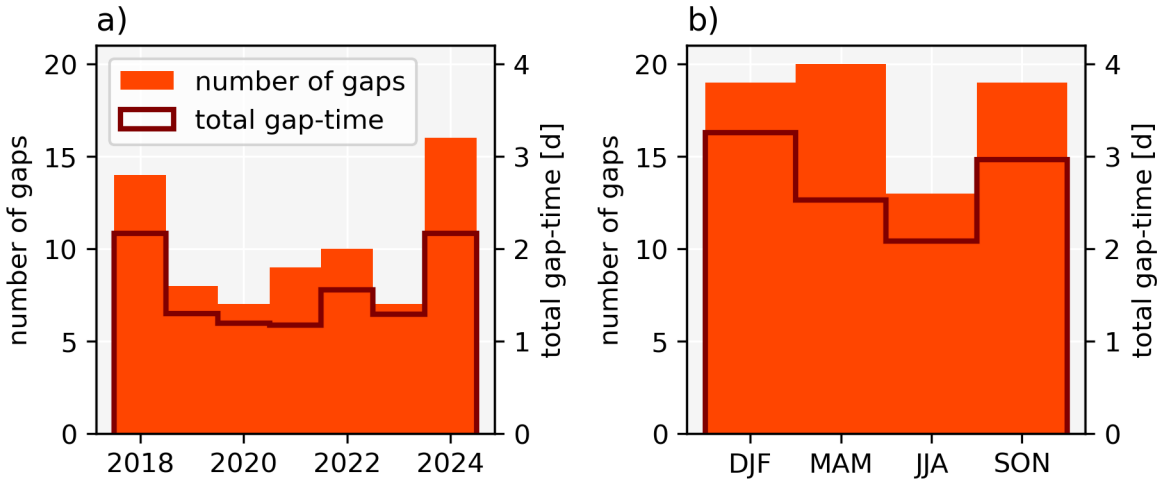


Figure 4.4: Number of gaps in the data with 1 hour or longer and their total accumulated time for *a)* each year and *b)* each season.

For optically thick clouds in the troposphere the detected temperature correlates with their height in the atmosphere, with colder being higher. The overshooting tops of the strong convection of a SL will therefore show very cold temperatures below 230–235 K, down to 200 K and below, which can be seen in Figure 4.2. Cirrus clouds have similarly cold cloud temperatures, but in the brightness temperature data they appear warmer because of their lower optical depth. I chose this dataset because deep convection can be clearly identified and also because of its high spatial and temporal resolution and coverage.

## 4.2 Identification and Tracking Algorithm

As I discussed in Section 3.3, there are different approaches to tracking leading to different results. The goal of this new algorithm is to detect large coherent features and track them through time for as long as possible.

For the algorithm I chose the classical two-step approach, where in the first step each individual frame is analysed in order to extract the features, and in a second step features from adjacent timesteps are linked, gradually forming continuous tracks.

### 4.2.1 Frame Analysis

During this first tracking step each frame is analysed without temporal context. It takes any data on a regular grid, extracts features from the data, and calculates a series of properties for each feature. This is done in two steps:

1. threshold based segmentation

2. analysis of individual features

The *threshold based segmentation* extracts deep convection from a TOA brightness temperature dataset by masking out all areas warmer than a chosen threshold. For the purpose of detecting organised deep convection in the tropics I chose a threshold of 230 K, which is 5–11 K colder than what is typically used for the detection of deep convective cells (Vila et al., 2008; Fiolleau and Roca, 2013; Feng et al., 2025). Due to the high tropopause in the tropics and the strong convection exhibited by SLs, I expect colder cloud-top temperatures than for typical cells. This may also help to reduce the size of the outflow region and focus more on the SL part containing the linearly organized convective cores. Figure 4.2 shows example data with highlighted temperatures below the threshold.

The areas in the mask, resulting from the threshold application, now have to be segmented into individual coherent features. This step is notoriously difficult for tropical SLs, because of their discontiguous nature. When segmenting based on contiguous areas, the large structure may be broken into smaller parts, loosing the broader structure of the coherent feature. The same holds true for watershed segmentation based on individual convective cores, because SLs are composed of several cores or a large elongated core with stronger and weaker parts.

Thus, more unconventional approaches are necessary which can be found in studies about SL detection algorithms for radar-based severe weather detection. Kamani et al. (2018) used skeleton matching, which skeletonizes the mask an tries to match the resulting “cloud skeletons” with a database containing “cloud skeleton types”. Wang et al. (2021) extracted the central axes of the clouds in the mask with a similar approach and then used a Houghs transformation to find the most probable SLs. Xie et al. (2023) used a deep learning method to detect SLs, which required a previously labelled dataset. Shi et al. (2024) applied a density based clustering method for segmentation, which is similar to the Density Based Spatial Clustering of Applications with Noise (DBSCAN) algorithm. As DBSCAN uses density based clustering, it is inherently gap permitting, and can distinguish between core, edge, and noise points (Ester et al., 1996;

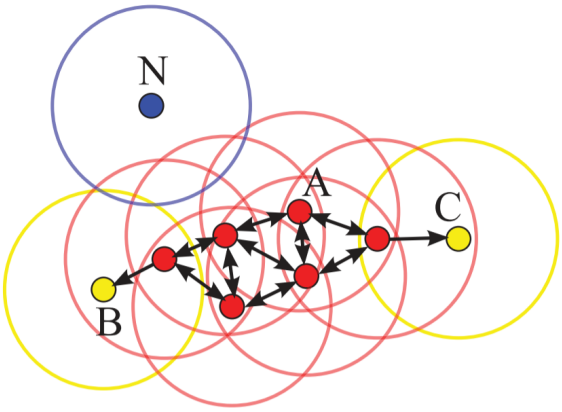


Figure 4.5: Illustration of the DBSCAN cluster model, depicting core points (red), edge points (yellow) and noise points (blue), each with their eps-range. (Figure and Caption adapted from Figure 1 in Schubert et al., 2017)

Schubert et al., 2017). Based on a radius "eps" the algorithm searches for a minimum number of points in the vicinity of each point and labels points that reach this criterion as "core point". The other points are either labelled as an "edge point" if they can reach a core point or else as noise, which is illustrated in Figure 4.5.[2]

The algorithm for this study will segment the convection mask into coherent features with DBSCAN. Only the core points produced by DBSCAN are used as parts of the cloud feature, discarding edge points, which should reduce accidentally clustered clutter and increase the focus on the main system. The eps has to be chosen large enough to jump gaps in SLs, and small enough not to cluster features together which should be kept apart. The number of minimum points can be calculated from a fraction which represents a density of masked to unmasked area in the eps-range. I use an eps (radius) of 45 km and a density of 30 %. This leaves a minimum detectable cloud area of about 2000 $km^2$ corresponding to about 25 km radius, which is smaller than the 2400 $km^2$ size threshold for MCSs employed by Vila et al. (2008) and Anselmo et al. (2021). This should result in well detected MCSs and compared to worldwide data in Figure 4.6, it only misses some small and short lived systems (Fiolleau and Roca, 2024). The maximally jumpable gap has a width of about $eps - 2r_{\text{pixel}} \approx 43\,\text{km}$, covering a high range of nearest neighbour distances (NNDs) of individual deep convective cells (Gupta et al., 2024), which can be seen in Figure 4.7.[2]

[2]Paragraph partially taken from Vieira Fischer (2025) with adaptations.

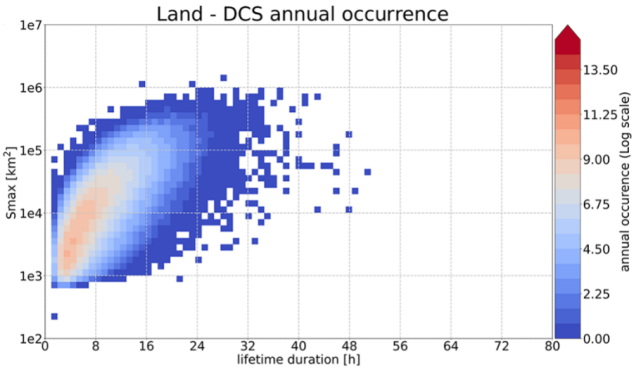


Figure 4.6: Annual occurrences of deep convective systems (DCS) over all continents binned by lifetime and maximum system size (Smax). (Figure from Fiolleau and Roca, 2024)

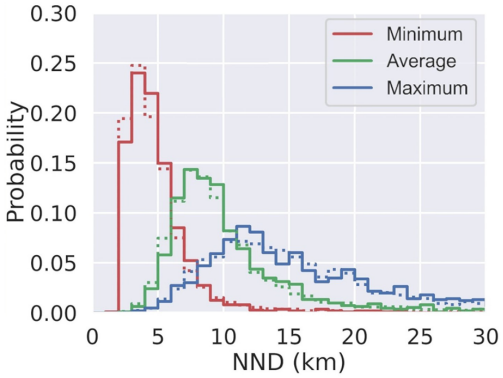


Figure 4.7: Track based probability distributions of minimum, average, and maximum NND between tracks of individual deep convective cells in the Amazon basin. (Figure from Gupta et al., 2024)

The next step in my algorithm is the *analysis of the individual features*. Based on the $x$ and $y$ coordinates of the features pixels, their values and the longitudinal and latitudinal grid resolutions $dx$ and $dy$, the properties listed in Table 4.1 are computed for each detected feature. Some of these properties rely on the central line and the convex hull of the feature. The convex hull is a mathematical construct which is the smallest convex shape, that contains the entire object of which it is the convex hull (Barber et al., 1996). In my case I have two dimensional cloud features, and their convex hull can be imagined like the shape of rubber band, when it gets tied around the cloud area. It will wrap tightly around the outermost edges and span the gaps, where the cloud has a concave feature or hole (see Figure 4.8).

The central line is derived by performing a weighted least squares fit with a linear function to the pixels of the cloud feature. The weights are computed the same way the tobac algorithm computes weights, by taking the absolute difference between the threshold and the individual cloud pixel value (Heikenfeld et al., 2019). The weights shift the location of the central line towards the colder convective cores of the SL, reducing the influence of the warmer outflow region. To confine the central line to the extend of the cloud feature, the intersection of the linear function from the fit and the convex hull is taken (see Figure 4.8). This method unfortunately biases the central line towards the axis of the independent variable. To counteract this bias, the calculation of the central axis is performed both for $y(x)$ and $x(y)$. From these two the one yielding the greater length is used as the central line, the other one is discarded.

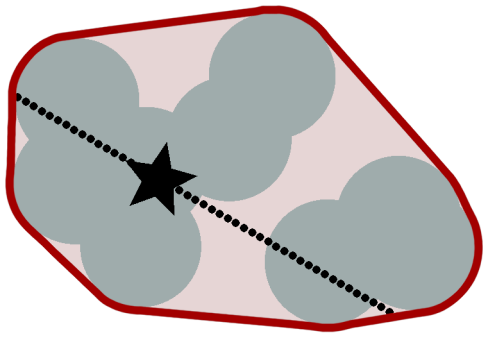

Figure 4.8: Schematic representation of a cloud (teal), its convex hull (dark red), its central line (black, dotted) and centroid (black star).

The position of the cloud feature is calculated as single point. This centroid is the weighted mean of all pixel positions, which is then snapped to the closest point on the central line. The latter is done for consistency reasons, so that the centroid and the endpoints of the central line are on the same line. For any other geometric property of the cloud feature (see Table 4.1), its shape is approximated as an ellipse, which is common praxis, since the irregular shape of a cloud is best represented by an ellipse, when keeping to simple geometries (e.g. Mathon and Laurent, 2001; Vila et al., 2008; Raut et al., 2021; Fiolleau and Roca, 2024; Shi et al., 2024).

All of the previously described cloud feature properties, are also calculated for regions within each system based on the lower brightness temperature thresholds 220 K, 210 K, and 200 K. This results in additional information about the area of stronger convection within each individual cloud feature.

The application of DBSCAN with a high *eps* in the Amazon basin can lead to very massive objects or clusters of many small convective cells produced by the afternoon convection in this region. Both are undesirable and thus two filters are set in place. First, any system with an area greater than 500 000 $km^2$, which is the upper limit of expected convective system size (Anselmo et al., 2021), will be subdivided. Secondly, any system with a CH_density ($area/CH\_area$) below 50 % and an aspect ratio above 1/3 was subdivided which are loosely clustered individual cells with a lot of 'clear sky' ($T_B < 230$ K) in between. Elongated convective systems (aspect $< 1/3$) were exempt from this, because they tend to have lower CH_densitys due to longer contours, which are usually not smooth and lead to more inclusion of 'clear sky' into the convex hull.

Table 4.1: This Table contains all scalar properties which are calculated for each feature during the analysis step, alongside with a description, the corresponding formula and an alternative value (else) if applicable.

| property | description | formula | else |
|---|---|---|---|
| peak | The lowest or highest value within the area encapsulated by the threshold. | $\min(values)$ or $\max(values)$ | - |
| mean | The average value within the area encapsulated by the threshold. | $\langle values \rangle$ | - |
| area | The pixel area within the threshold. | $i \cdot dx \cdot dy$ with $i$ as the number of pixels in the feature | - |
| x, y | Centroid position of the feature. Weighted or unweighted mean of all pixel positions. Can be snapped to the closest point on the central line. | $(x,y) = \frac{\sum_i w_i(x_i,y_i)}{\sum_i w_i}$ | - |
| x1, y1, x2, y2 | Endpoints of the central line derived as the intersection points between a (weighted) linear regression and the convex hull of the feature. | $central\ line \quad \cap$ $convex\ hull\ contour$ | $NaN$ |
| length | Distance between the endpoints. | $\sqrt{(x1-x2)^2+(y1-y2)^2}$ | $NaN$ |
| boundary_ contact_ length_ | The length of the features contour that touches the domin boundary. | $\lvert\{X_0,X_n\}\cap\{x_1,...,x_i\}\rvert \cdot dy$ $+\lvert\{Y_0,Y_n\}\cap\{y_1,...,y_i\}\rvert \cdot dx$ with $X,Y$ domain coordinates and $x,y$ feature coordinates | - |
| width | Width of the feature calculated as the width of an ellipse with the same area and length as major axis. Defaults to 0 if $length = 0$. | $\frac{4 \cdot area}{\pi \cdot length}$ | 0 or $NaN$ |
| aspect | Aspect ratio of the feature. Same as the eccentricity of its ellipse. Defaults to 1 if $length = 0$. | $\frac{width}{length}$ | 1 |
| angle | Angle between the x-axis and the central line of the feature. | $\arctan\left(\frac{y1-y2}{x2-x1} \cdot \frac{dy}{dx}\right)$ | $NaN$ or $\infty$ |

Table 4.1: (continuation)

| property | description | formula | else |
|---|---|---|---|
| CH_area | Convex hull area of the feature. Since the convex hull uses the pixel centers, the area ($A$) has to be corrected by the convex hull contour length ($C$) times halve the pixel width plus one pixel. | $(A+0,5\cdot C+1)\cdot dx\cdot dy$ | $NaN$ |
| CH_density | Density of the feature within its convex hull, calculated as the ratio between the pixel area and the convex hull area. | $\frac{area}{CH\_area}$ | $NaN$ |
| CH_width | Similar to *width* but calculated based on the convex hull of the feature. Defaults to 0 if $length=0$. | $\frac{4\cdot CH\_area}{\pi\cdot length}$ | 0 or $NaN$ |
| CH_aspect | Similar to *aspect* but calculated based on the convex hull of the feature. Defaults to 1 if $length=0$. | $\frac{CH\_width}{length}$ | 1 or $NaN$ |
| weight | Total weight of the feature, calculated as the sum of the individual weights $w_i$ of the features pixels. | $\sum_i w_i$ with $w_i=\lvert value_i-threshold\rvert$ or $w_i=1$ | - |

The subdividing is accomplished by first re-applying the *threshold based segmentation* with a new lower threshold and clustering with DBSCAN to the data of the feature to be subdivided. The new cloud features are then extended to the original threshold by applying watershed segmentation. After that their properties are recomputed. The watershed segmentation, which is also used by tobac and TOOCAN (Heikenfeld et al., 2019; Fiolleau and Roca, 2013), takes the brightness temperature data as a basin topography and the features detected with a higher threshold as source areas. From these the basins are slowly filled until the area is segmented into the basins which belong to each source area. This results in several new features which completely replace the original feature. The subdividing will first be tried with a 225 K threshold and then iterate down in 5 K steps until either all cloud features meet the criteria above or the final 200 K threshold is reached.

At this point all detected cloud features could be filtered according to criteria like Anselmo et al. (2021), Sousa et al. (2021) and Shi et al. (2024) implemented in their studies. I found, that this leads to a worse tracking performance in the next step. The classification is therefore performed in post-processing after the tracking.

### 4.2.2 Tracking

The tracking is performed similar to what is described by Vila et al. (2008), although I implemented some improvements. The basic idea behind tracking is that cloud features that overlap in timesteps with small timespans in between belong to the same track. For this study I chose a maximum timespan in between two frames of 1 h, larger gaps will break the tracking and can be seen in Figure 4.3. Usually there are 5 frames (3 before April 2019) within the 1 h maximum tracking timespan, so even minor outages or broken data in individual frames shouldn't break the tracking. This maximum tracking timespan was also implemented by Anselmo et al. (2021) and is within the 70 min average lifetime of individual deep convective cores in the Amazon region found by Gupta et al. (2024).

To ensure the overlap between two features, the cloud features have to overlap by at least 10 % compared to the area of the smaller one. More restrictive or complex tracking criteria can be used but they wouldn't improve the tracking significantly (L. A. T. Machado et al., 1998). During the development of the tracking algorithm I tried differential filtering, predicted minimum overlap based on ellipses, and predicted movement based on prior system speed. All of these methods either did perform similar or significantly worse to the minimum overlap criterium. As L. A. T. Machado et al. (1998) described, there is usually only one candidate overlapping with the current cloud feature in two related time steps, and even for the cases where more than one candidate overlaps, this can be attributed to merging or splitting of tracks. More restrictive tracking criteria therefore can break apart tracks, counteracting the stated goal for this algorithm of producing long and continuous tracks.

This leads to handling different tracking situations, which can be seen in Figure 4.9. The two most simple situations are firstly a cloud feature that does not overlap with another cloud feature in a prior time step, which will be marked as initial "I" time step. And secondly, a cloud feature that overlaps with exactly one feature in a prior time step, which will be marked as a continuity "C" situation. The other three tracking situations are more complex and involve three or more cloud features distributed over two time steps. When more than one feature in the

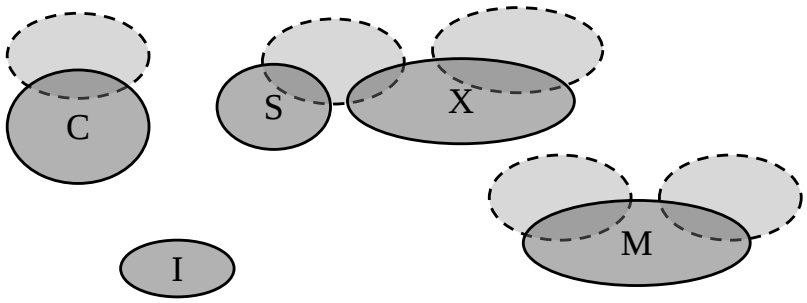


Figure 4.9: Schematic representation of possible tracking situations and their respective tracking flags. The dashed objects belong to a prior timestep, the continuous objects belong to the currently tracked timestep. The track flags are “C” for continuity, “S” for splits, “M” for merges, “X” if both splitting and merging occurs, and “I” for initiation.

prior time step overlaps with a feature in the current time step, they are considered to merge “M”, and the opposite way around they are considered to split “S”. In these cases the larger absolute overlap continues the track, while the other tracks terminate in the merge situation, or new tracks are started in the split situation. The continued track is saved as “main track” for all currently tracked cloud features, which are part of such a merge or split. It is possible that both merge and split situations “X” apply to a cloud feature, which are handled sequentially first as a merge and then as a split.

During the tracking phase not only the linking of features into tracks is done, but further properties with temporal dependencies are added to the cloud features. A list of these properties can be found in Table 4.2. Most of these properties are calculated based on the previous cloud feature of the main track of the current feature, which is the track which is continued during a merge and/or split situation. For initial and continuity situations the track and main track are the same. The centroid speed and geometric speed however differ if a merge and/or split situation occurs. For a merge situation all cloud features from the prior time step that overlap with the current cloud feature are taken into account according to their weights (centroid speed) or area (geometric speed). A split situation however works a bit different, because each feature in the currently tracked time step, that split from the feature in the previous time step, can have a different speed. Therefore the speed has to be corrected, which is achieved by interpolating between the combined centroid/geometric center of all cloud features belonging to the splitting

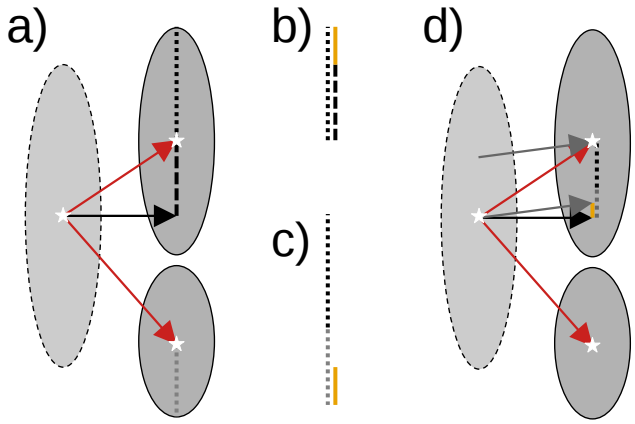


Figure 4.10: Schematic representation of the split speed correction. The ellipses represent cloud features like in Figure 4.9, with white stars as centroid/geometric center. The arrows represent speeds, with red being the center based speeds, black the speed with combined centroids, and gray the interpolated speed. The dotted lines represent 1/2 cloud feature lengths and the dashed line the offset of the features center from the combined center. The yellow line in *b)* shows the difference between the offset and feature length, and c) compares the this length with 1/2 the combined feature lengths.

and the specific cloud feature for which the speed is calculated. The interpolation is based on the offset of the cloud feature center from the combined center and the length $l$ of the cloud features:

$$f_i = \frac{\frac{l_i}{2} - \sqrt{(x_i - \langle x \rangle)^2 + (y_i - \langle y \rangle)^2}}{\sum_j \frac{l_j}{2}}$$

The interpolation defaults to $f_i = 0$ if $f_i$ would be negative due to a greater center offset than feature length. The grater $f_i$ the more influence the center of the individual feature has on the speed. This interpolation is visualized in Figure 4.10.

Table 4.2: This Table contains all scalar properties which are calculated for each feature during the tracking step, alongside with a description, the corresponding formula and an alternative value (else) if applicable.

| **property** | **description** | **formula** | **else** |
|---|---|---|---|
| track | Unique ID of the track. | | - |
| track_ overlap | Overlap with the feature in the previous time step $t-1$. Defaults to 0 if not linked to a feature in a prior time step. | $feature_t \cap feature_{t-1}$ | 0 |
| track_ overlap_ frac | Maximum overlap fraction relative to the feature area at either $t$ or $t-1$. Defaults to 0 if not linked to a feature in a prior time step. | $\max\left(\left\{\frac{track\_overlap}{area_t}, \frac{track\_overlap}{area_{t-1}}\right\}\right)$ | 0 |
| track_flag | Flag indicating if merges and/or splits occurred between the time steps $t-1$ and $t$. Defaults do I if not linked to a feature in a prior time step. | C: continuity, M: merge, S: split, X: split&merge, I: initial | I |
| speed_x, speed_y | Speed components along $x$ and $y$ of the centroid. Defaults to 0 if not linked to a feature in a prior time step. | $\frac{i_t - i_{t-1}}{track\_delay} \ \forall \ i \in \{x, y\}$ | 0 |
| speed | Absolute speed of the centroid. Defaults to 0 if not linked to a feature in a prior time step. | $\sqrt{speed\_x^2 + speed\_y^2}$ | 0 |
| speed_x_ geo, speed_y_ geo | Speed components along $x$ and $y$ of the geometric center. Defaults to 0 if not linked to a feature in a prior time step. | $\frac{i_t - i_{t-1}}{track\_delay} \ \forall \ i \in \{xgeo, ygeo\}$ | 0 |
| speed_geo | Absolute speed of the geometric center. Defaults to 0 if not linked to a feature in a prior time step. | $\sqrt{speed\_x\_geo^2 + speed\_y\_geo^2}$ | 0 |
| track_main | Unique ID of the main track indicating the continued track if the track flag is M, S or X. | | $track$ |
| xgeo, ygeo | Geometric center of the feature. | $\langle\{x_1, ..., x_n\}\rangle$, $\langle\{y_1, ..., y_n\}\rangle$ | - |

Table 4.2: (continuation)

| property | description | formula | else |
|---|---|---|---|
| track_delay | Time between the linked feature in a prior time step and the current feature. Defaults to 0 if not linked to a feature in a prior time step. | $time_t - time_{t-1}$ | 0 |
| merge_ count | Number of tracks that merged into the main track at this time step ($t$). Defaults to 0 if no merge occurred. | $N_{elements} - 1$ in $\{tarcks_{t-1} \bowtie track_t\}$ | 0 |
| split_count | Number of tracks that the feature in the last time step ($t-1$) split into. Defaults to 0 if no split occurred. | $N_{elements}$ in $\{tarcks_t \bowtie track_{t-1}\}$ | 0 |
| growth_ factor | Relative growth (or decay) of the feature compared to the feature in the previous time step. Defaults to $\infty$ if not linked to a feature in a prior time step. | $\frac{area_t}{area_{t-1}}$ | $\infty$ |

## 4.3 Post Processing and Classification

During post processing a few variables were corrected and tracks were classified into SLs, MCSs, and other. The variables to be corrected were all aspect ratios and the growth factor.

While most aspect ratios range from 0 to 1, there were some significantly higher. In itself that is not a problem, since it means that the width of the system is higher than its length. This however complicates the filtering and analysis, while not adding much information. I therefore added a corrected version of each aspect ratio, which is calculated by dividing 1 through the aspect ratio if the aspect is greater than 1.

The growth factor was calculated as the area of the previous cloud feature divided by the area of the current cloud feature in a track. The time differences between previous and current cloud feature can range from around 5 min to 1 h, limited by the satellites scan frequency and the maximum timespan of my tracking algorithm. This leads to variation in the growth factor, both from the actual system size change, but also from the time difference. The simple solution is to calculate a growth rate from the growth factor and time difference $\Delta T$:

$$growth\_rate = \sqrt[\Delta t]{growth\_factor} - 1 \tag{4.1}$$

The growth rate is not simply the growth factor divided by the time, but calculated analogue to compounding probabilities or compounding interest. While the growth factor was centred around 1, the growth rate is now centred around 0.

### 4.3.1 Classification of Mesoscale Convective Systems

The classification criteria for MCSs were taken from Anselmo et al. (2021), in order to be able to compare my results to that study. Their criteria were a minimum lifetime of 2,5 h and a minimum size of 2400 $km^2$. Any track in my dataset that fits these criteria and isn't classified as a SL (see below) was marked as a MCS.

### 4.3.2 Classification of Squall Lines

For the classification of tracks into SLs I used a list of visually identified SLs during the year of 2024, based on my dataset derived from the GOES-16 data. This list contains both the track identifier of 76 SLs, and the respective time periods during which they underwent clearly identifiable SL stages. With this list I created two subsets from the data Gaussian kernel density

estimations. One derived from the clearly identified SL steps (hereafter "SL"), and another derived from the rest of the tracks with no visually identified SL stages (hereafter "R").

A Gaussian kernel density estimation acts like a probability density function which can be derived from a dataset using Gaussian smoothing. The goal is to classify SLs based on the ratio between the probability densities derived from Gaussian kernel density estimations which will be set up with the data subsets "SL" and "R". I chose six variables from my analysis which span a six dimensional latent space for the kernel density estimation. These six variables I chose are listed below:

- *peak*: The coldest brightness temperature of each cloud feature should be a very cold minimum for SLs, because they exhibit strong deep convection.
- *length*: The length of each cloud feature is generally higher for SLs, than for the average system.
- *CH_aspect_c*: The (corrected) aspect ratio of the convex hull is lower for SLs, due to their linear shape.
- *CH_density*: The convex hull density should help to avoid cloud features which are loose clusters of individual cells.
- *length_210*: The length of the region with strong deep convection ($T_\mathrm{b} < 210\,\mathrm{K}$) should be similarly long as the cloud feature length.
- *CH_aspect_210_c*: The convex hull aspect ration (corrected) of the region with strong deep convection ($T_\mathrm{b} < 210\,\mathrm{K}$) should reflect the line shape of a SL as well.

Cloud features that do not reach brightness temperatures below 210 K have missing values for the latter two variables. Since they therefore can't contribute to the Gaussian kernel density estimation, they had to be removed from the data subsets. All variables were normalized to values between 0 and 1 as preparation for the Gaussian kernel density function (see Figure 4.11):

$$\bar{x}_i = \frac{x_i - x_\mathrm{min}}{x_\mathrm{max} - x_\mathrm{min}} \tag{4.2}$$

The Gaussian kernel density estimation also needs a smoothing bandwidth for the Gaussian smoothing. Too low smoothing can lead to over fitting to the setup data, too high smoothing will deteriorate the quality of the resulting probability density. I therefore set up six pairs of Gaussian kernel density estimations from the aforementioned SL and R subset with the

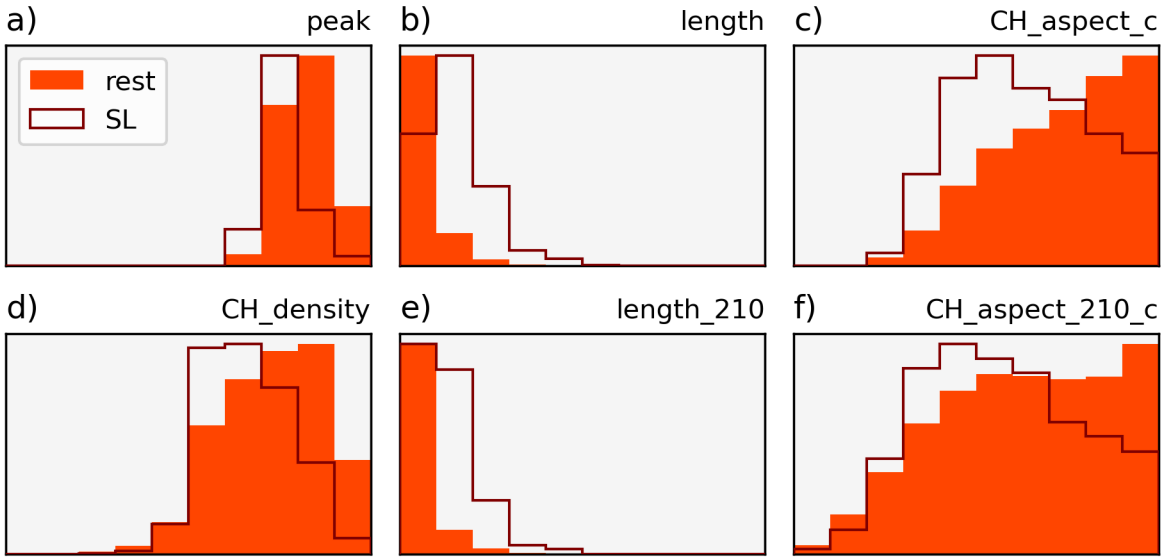


Figure 4.11: Normalised distributions of the normalised variables for creating the Gaussian kernel density estimation. The "SL" graphs contain the data from the visually identified SL stages during 2024, the "rest" graphs contain the data from all tracks from 2024 that do not contain any SL stage.

smoothing bandwidths 0,1, 0,15, 0,2, 0,25, 0,3, and 0,5. After applying these kernel density estimations I get a SL probability density $p_{\mathrm{SL}}$ and a "rest" probability density $p_{\mathrm{R}}$ for each cloud feature and each smoothing bandwidth. The difference $p_{\mathrm{R}} - p_{\mathrm{SL}}$ can be seen in Figure 4.12 with the fraction that each year from my dataset contributes to each bin. The change from over fitting to over smoothing can be seen in the change in contribution of the year 2024 combined with the change in distribution near the lower end. Systems that had missing data in one or more of the six variables were assumed not to be SLs and got the lowest $p_{\mathrm{SL}}$ and the highest $p_{\mathrm{R}}$, leading to the persisting peak at the right of the distribution. I made this assumption for the reason that most missing data was due to the cloud features not having an area below 210 K, which would be necessary for a SL.

Figure 4.12 shows a relatively uniform contribution for positive values of $p_{\mathrm{R}} - p_{\mathrm{SL}}$, which have a low probability of being a SL. SL candidates contribute to the negative values which seem to over proportionately represent the training year 2024. However, this impression is created by the logarithmic y-axis. Figure 4.13 shows that the contribution of each year for $p_{\mathrm{R}} - p_{\mathrm{SL}} < 0$ is much more uniform, with 2024 actually contributing more with higher smoothing bandwidths. The figure also shows, that stronger smoothing increases the number of potential SL candidates.

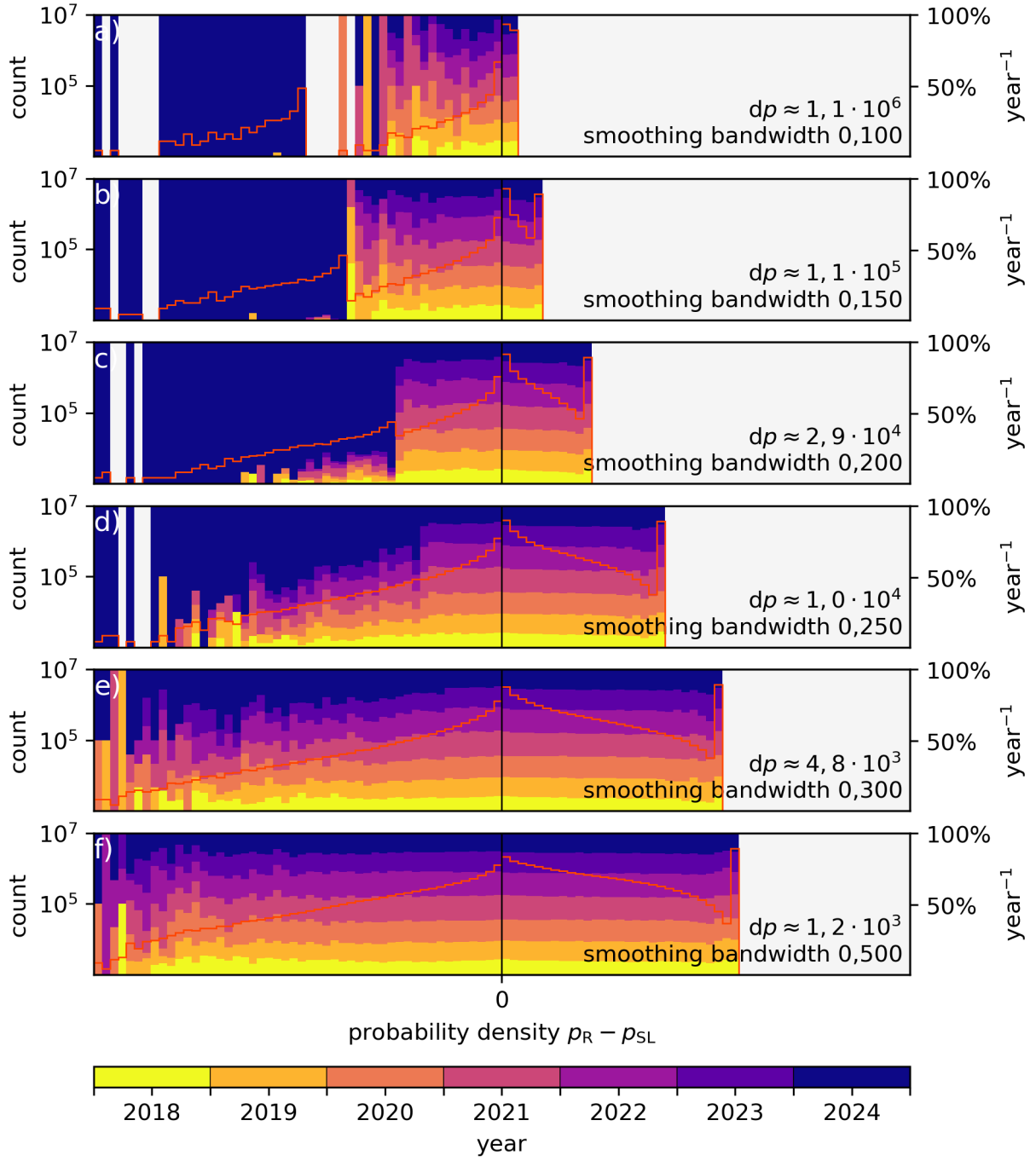


Figure 4.12: Probability density distribution for the difference between the "rest" and SL probability density $p_R - p_{SL}$ in red. The shading in the background indicates the fraction each year from the dataset contributes to each bin. The different subplots contain the probability density differences derived from Gaussian kernel density estimations with different smoothing bandwidths.

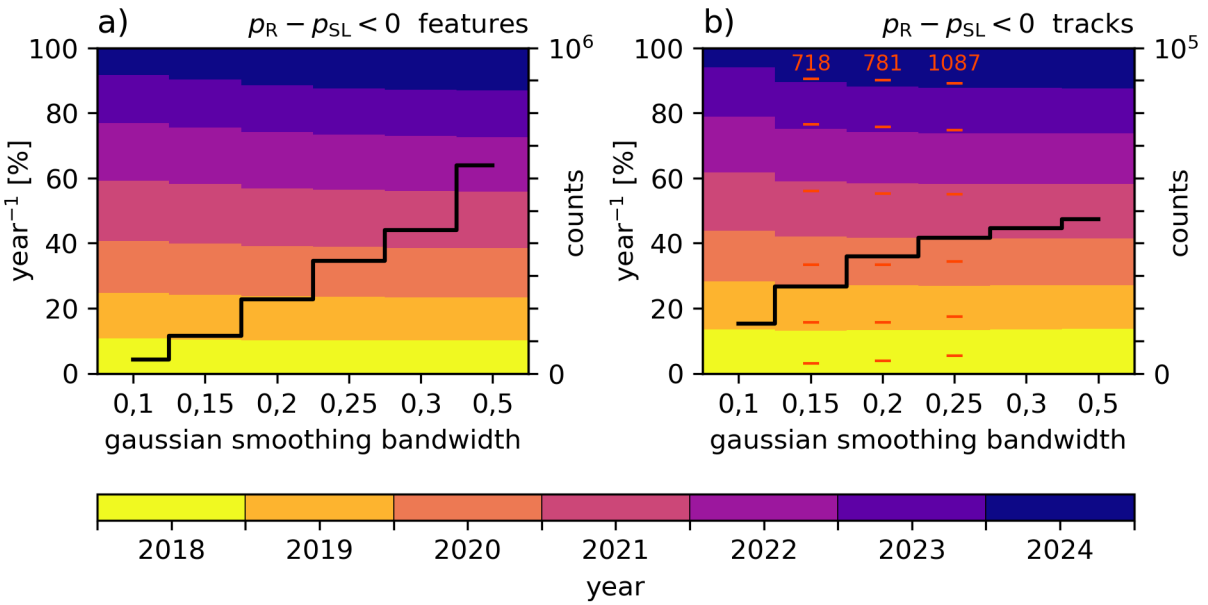


Figure 4.13: Number of potential SL candidates $p_R - p_{SL} < 0$ depending on the smoothing bandwidth used in the Gaussian kernel density estimation. The number of individual cloud features is shown in a) and b) shows the number unique tracks. The shading in the background indicates the relative contribution of each year to the total count. The red numbers indicate counts and the red bars indicate the edges of the relative contribution of each year for three chosen sets of smoothing bandwidth, factor and threshold.

For the final filter I need frames which have a much higher SL probability density compared to the "rest" probability density. I achieve this by artificially increasing the "rest" probability density $p_R$ by a certainty factor $f$:

$$f \cdot p_R \overset{!}{<} p_{SL} \tag{4.3}$$

Furthermore it is not sufficient for a single cloud feature in a track to fulfil Equation 4.3. Rather a track should have several of such cloud features. My filter therefore requires, that from all cloud features $i$ of a track, the number of SL cloud features has to be grater or equal to a threshold $t$:

$$t \leq \sum_i \begin{cases} 1 & p_{i,SL} > f \cdot p_{i,R} \\ 0 & p_{i,SL} \leq f \cdot p_{i,R} \end{cases} \tag{4.4}$$

Figure 4.14 shows the ration of falsely identified SLs (false positives) to falsely misidentified SLs (false negatives) in 2024 for the range of factors $f$ and thresholds $t$ for each smoothing bandwidth. The higher smoothing bandwidths 0,3 and 0,5 perform rather bad, requiring high factors and

Table 4.3: Three sets of filter parameters, each consisting of a smoothing bandwidth, a certainty factor, and a minimum SL cloud feature threshold. The rates of false negatives and false positives are derived from the training data from the year 2024. The last column contains the number of all detected SLs in the entire dataset from 2018 to 2024.

| bandwidth | factor $f$ | threshold $t$ | false negatives | false positives | SL 2024 | SL total |
|---|---|---|---|---|---|---|
| 0,15 | 1 | 20 | 19,7 % | 11,6 % | 69 | 718 |
| 0,20 | 5 | 20 | 19,7 % | 21,7 % | 78 | 781 |
| 0,25 | 10 | 16 | 10,5 % | 42,9 % | 119 | 1087 |

still having both high false negatives and high positives indicating a low skill. Contrary to that, the low smoothing bandwidths 0,1 and 0,15 perform quite well even for the low factors 1 and 2. Both the previous two figures and this indicate that the smoothing bandwidth of 0,1 might be severely over fitting the training data.

For the remaining three bandwidths I chose two sets of factor-threshold-pairs with a low rate of false positives on the cost of higher false negatives for the bandwidths 0,15 (factor 1 and threshold 20) and 0,2 (factor 5 and threshold 20). For the bandwidth 0,25 I chose a factor-threshold-pair with a lower rate of false negatives, producing more false positives (factor 10 and threshold 16). These three sets should provide some variety, whilst being reasonably conservative choices. Their properties can both be seen in Figure 4.13 and Table 4.3.

All three of these filter sets seem to be a reasonable choice, which is further supported by the similar patterns of monthly SL counts for these three sets, and can be seen in Figure B.1 in the Appendix. The first does a good job in reducing the false positives, producing a high certainty for identified SL tracks. The smoothing bandwidth however might be over fitting the training dataset. The second set performs similarly to the first, but Figures 4.12 and 4.13 show less over fitting and more consistent performance for 2018 (see Figure 4.12 b) and c)). Furthermore its number of 78 detected SLs in 2024 is very close to the number of 76 visually identified SLs in 2024, making it probably a good predictor of SL count, even though the false positives are higher. The third filter set has a better rate of false negatives than the other two but performs much worse in respect of false positives. This however would be expected if one would assume that the list of visually identified SLs is incomplete. However, this also reduces the certainty of identified SL tracks.

For this study I decided to continue with the the second set of filter parameters, because it shows less over fitting, has a strong certainty factor of $f = 5$, and predicts the total number of SLs better than the other two.

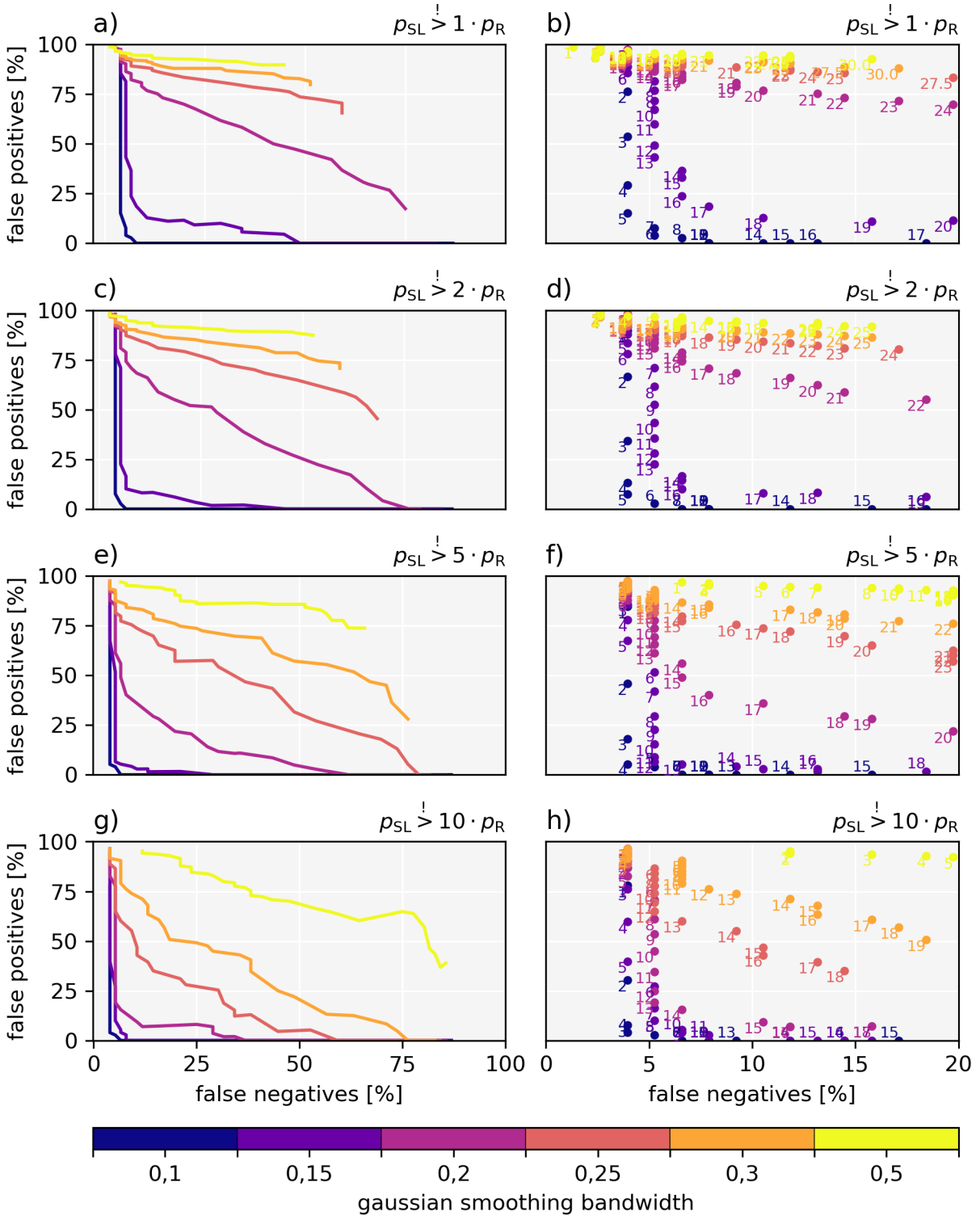


Figure 4.14: Ratios of falsely identified SLs (false positives) to falsely misidentified SLs (false negatives) in the training data from 2024 for different sets of smoothing bandwidth (color), certainty factor (rows), and minimum SL cloud feature threshold (independent variable). The left column a) – g) contains the entire domain, the right column b) – h) is limited to false negatives below 20 % and the numbers show the thresholds belonging to each point.

Part 5

# Results

The algorithm analysed 345135 images and identified a total of 4944355 cloud features. These features were linked by the tracking into 418812 tracks. The classification then identified 781 SL tracks (0,2 %), 90138 MCS tracks (21,5 %), and 327884 other tracks (78,3 %).

All tracks can either form by themselves and therefore start with an initial "I" cloud feature, or split "S" from another track. From all tracks, a total of 217037 (51,8 %) tracks start from an initial cloud feature, for SLs that is 62,2 % and 69,2 % for MCSs. Swaths were calculated for all tracks, with a "swath" being the entire area that the system covered during its lifetime. Figure 5.1 shows a tracking example, including the swath of a MCS and a single cloud feature thereof as comparison.

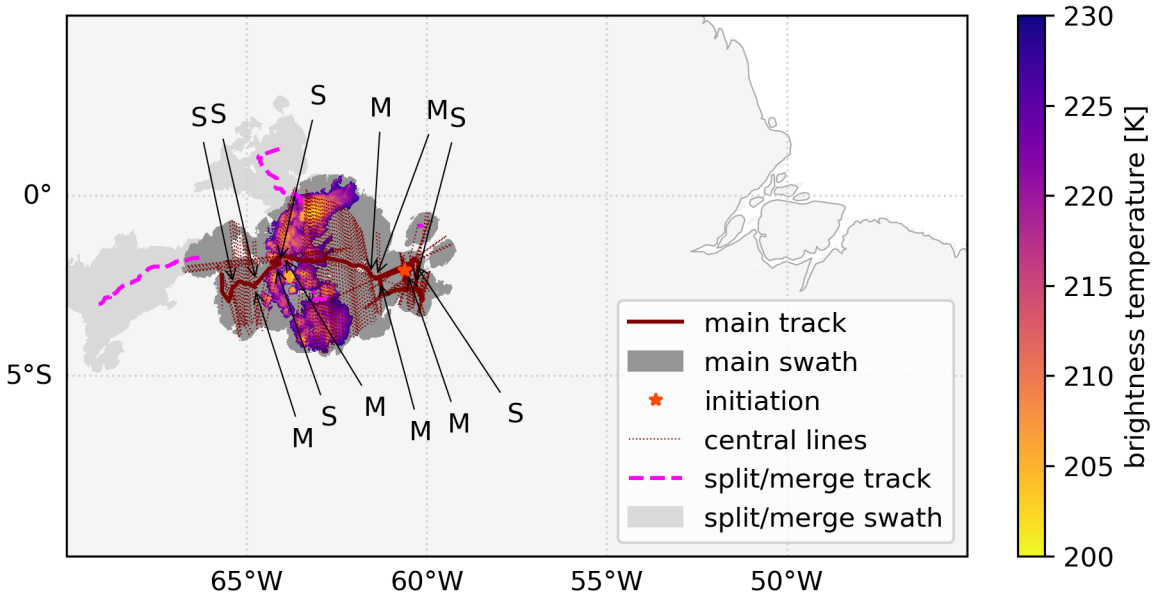


Figure 5.1: Tracking example of a single MCS from 19.10.2022 with example brightness temperature data from 10:15 UTC. Aside from the centroid track, central lines, and swath of the main system, the figure also displays merges (M), splits (S), and their respective tracks and swaths.

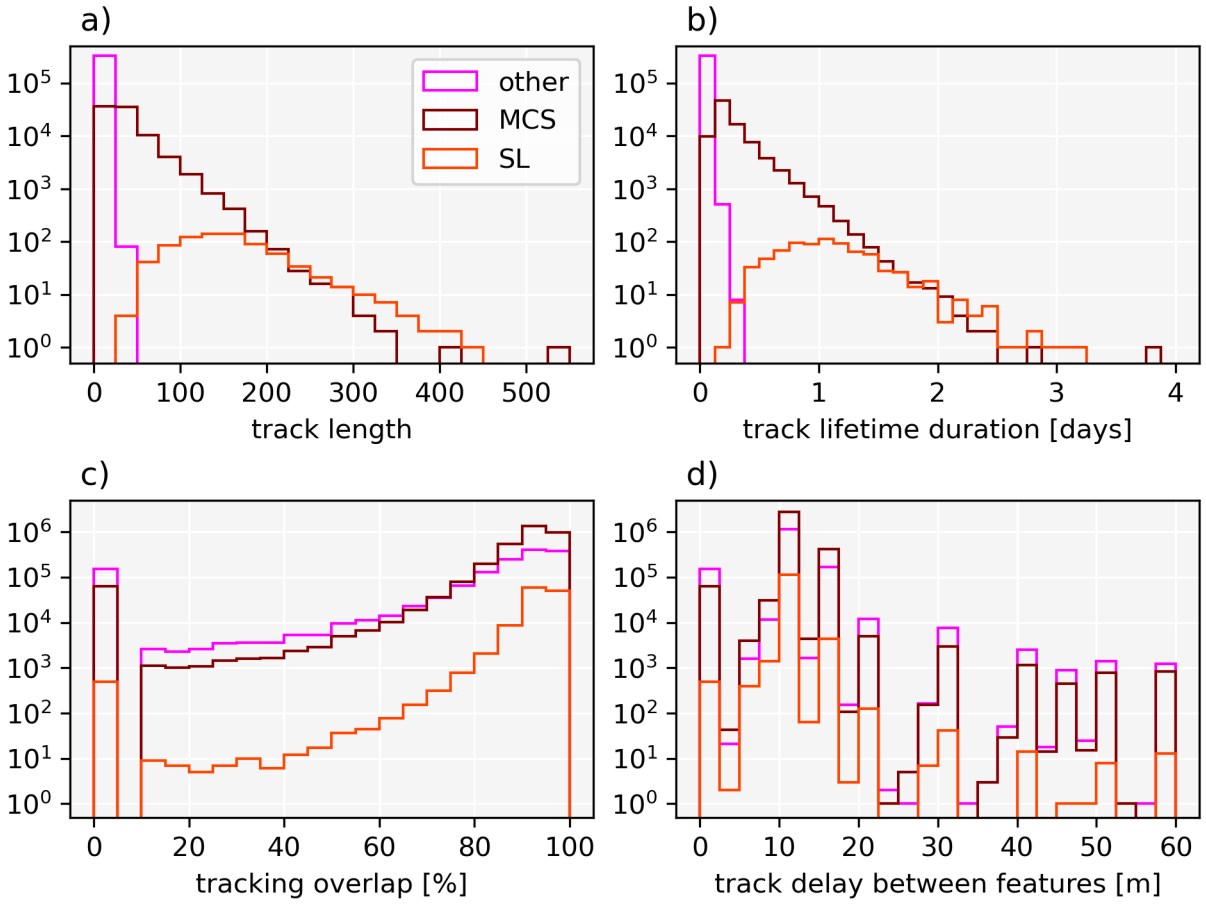


Figure 5.2: Histograms of some general tracking properties by system type. Histogram a) shows the numbers of cloud features linked together to a track, b) shows the lifetimes of the tracks, c) shows the relative overlaps of each feature their prior feature, and d) shows the time between each feature and their prior feature.

The SL tracks have a mean size of $83 \cdot 10^3\,\mathrm{km}^2$ and mean lifetime of 1 day 02:26 hours, and MCSs $23 \cdot 10^3\,\mathrm{km}^2$ and 06:24 hours. Both system types have tracks exceeding lifetimes of 2 days, which can be seen in Figure 5.2 b). The unclassified "other" convective cells or systems have lifetimes of less than 9 hours, with most of them existing for less than 3 hours. This classification also includes 103078 tracks from which only a single time step could be detected. These could be tracks, that barely reach the detection minima, but reasons for short tracks also include larger system dissipating and splitting into smaller dissolving features, cells initiating in the vicinity of larger systems subsequently merging with them, and parts of larger systems splitting and re-merging (see Figure 5.1).

The distribution of lifetimes also reflects in the track length, which means the number of cloud features linked into a track (Figure 5.2 a)). Both distributions have a strong bias

towards short tracks for MCSs and other systems. The distribution of the overlap between two adjacent cloud features has its maximum close to 100 %, falling off several orders of magnitude in occurrence towards the minimum of 10 % (Figure 5.2 c)). This behaviour is strongest for SL tracks, and weakest for other, correlating with their size distributions (Figure B.2 and B.3 in the Appendix). Smaller overlaps are expected for smaller features (e.g. early time steps during a systems lifetime), because their diameter is smaller, resulting in less overlap for the same travel distance. The time between two adjacent cloud features from a track is typically in the range of 5 to 25 minutes, which corresponds to 0 or 1 missed time steps. This distribution again trails off towards the 1 hour mark after which tracking stops (Figure 5.2 d)). Both latter distributions show data at 0 which is the expected behaviour caused by initial steps, which do not have a prior step.

In the Appendix Figures B.2 and B.3 show histograms of a wide variety of system properties. The usual distribution has the large, long lived, and elongated SLs towards one end of the distribution, and the small, short lived, and more symmetrical than other classified systems towards the other end of the distribution. The MCSs are located in between, with their expected properties also covering the range between SLs and other. The angles of the central axis however show a weird behaviour, by preferring angles close to 0° or ±90°. This was probably caused by the way the algorithm determines the central line. The peaks of the mean track speed distributions are in the range of expected values, with their maxima around $45\,\mathrm{km\,h^{-1}}$ for SLs, $35\,\mathrm{km\,h^{-1}}$ for MCSs, and $25\,\mathrm{km\,h^{-1}}$ for other systems. The maximum speeds exceed $100\,\mathrm{km\,h^{-1}}$ which can be attributed to splitting and merging of systems, where both weighted and geometric centroid can exhibit large jumps.

## 5.1 Diurnal, Seasonal, and Annual Variations

For the analysis of the diurnal and seasonal cycles, the region was split into its four quadrants. This leaves us with two coastal regions in the north-east (NE) and south-east (SE), and two "basin" regions in the north-west (NW) and south-west (SW). Figure 5.3 shows the diurnal cycles for all four regions and all three system types. Figure 5.4 shows the respective seasonal cycles. The fractions for each system type show the number of tracks occurring in the specific time bin divided by the total number of systems. The fractions add up to more than a 100 %, because tracks may persist for more than one time bin and pass through more than one region.

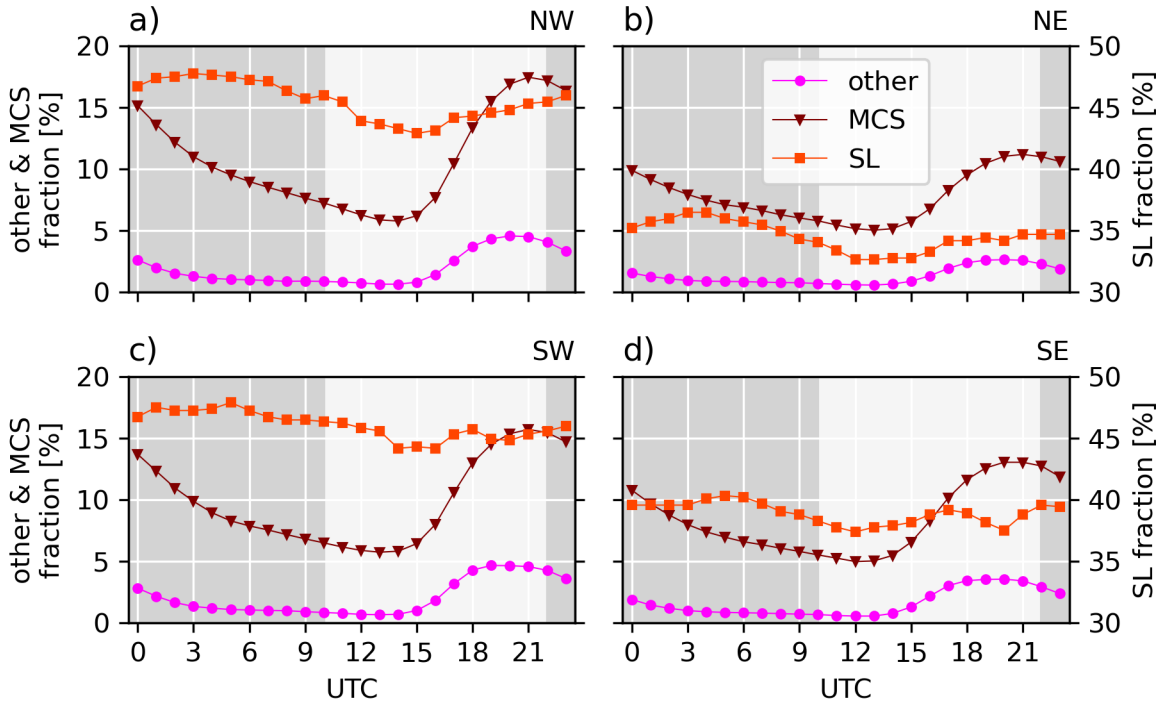


Figure 5.3: Diurnal cycle of all three system types by quadrant of the region of interest, showing the fraction of tracks per hour of day and region. The time between 18:00 and 06:00 UTC–4 has a darker background shading to imply nighttime.

The MCSs and other systems follow the typical diurnal cycle of local convection in the Amazon region, starting from a minimum before local noon, and a rapid increase after the local noon, with a maximum shortly before sunset, and subsequently followed by a decrease. The decrease differs somewhat, for it drops off much more gradually throughout the night for MCSs and faster for the "other" category. This is the expected behaviour when comparing large complex convective clusters to individual deep convective cells. The pattern is the same in all four regions, the amplitude however differs. It is highest in the basin regions, followed by the south-eastern, and then by the north-eastern region. Considering that the north-eastern region includes a bit of ocean and the south-eastern region is also rather close to the coast, the amplitude varies with the coastal and Atlantic influence.

Similar to the other two categories, the SLs in the two basin regions have their minimum before or around local noon. The increase thereafter differs and is comparatively more gradual with its maximum around local midnight. The coastal regions have the same midnight peak, but drop faster towards the minimum in the morning hours around 12:00 UTC or around 08:00 local

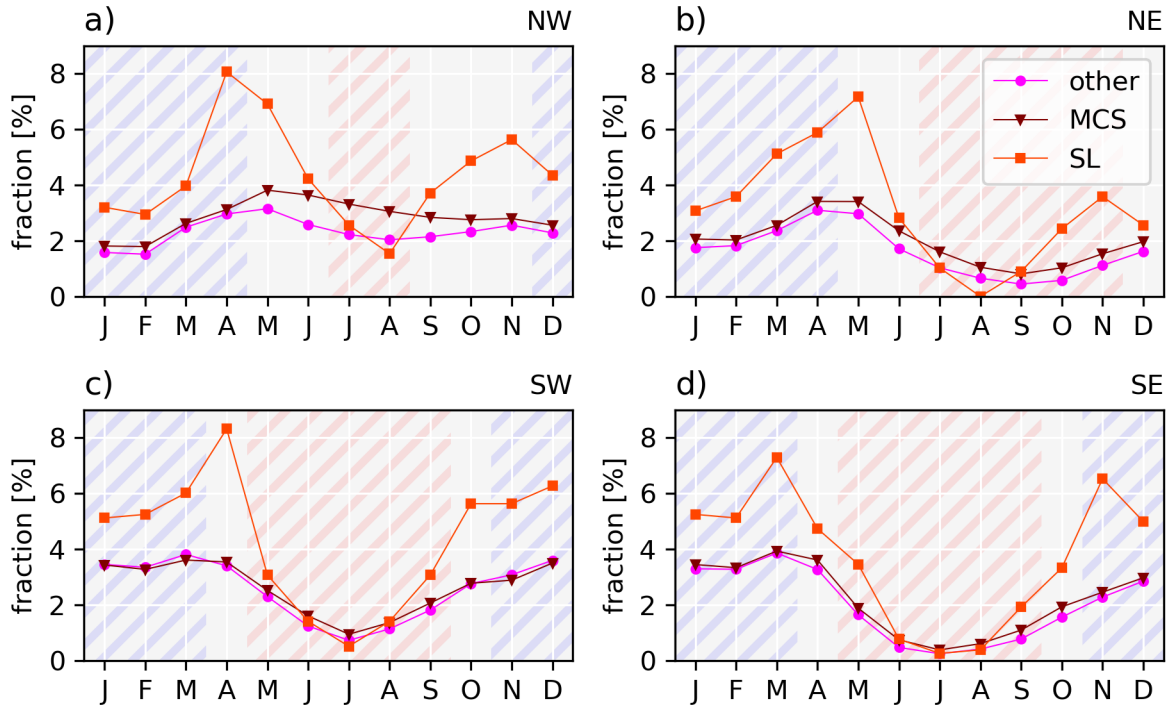


Figure 5.4: Seasonal cycle of all three system types by quadrant of the region of interest, showing the fraction of tracks per month and region. The hatching indicates dry (red) and wet season (blue) for the cities a) Manaos, b) Belém, c) Vilhena, and d) Brasilia (seasonal data from L. A. Machado et al., 2004).

time. All regions have a slight plateau or dip around the time of afternoon convection, which could be either due to the algorithm having detection problems in the noisier environment, or to less distinction from the surrounding convection (leading to merges, splits, or dissipation), or both. The overall amplitude of the diurnal cycle is lower than for the other system types, but the baseline is higher due to long lifetimes and far travel distances.

For the seasonal cycles I added blue and red hatching to Figure 5.4 to indicate the wet and dry seasons of the cities a) Manaos, b) Belém, c) Vilhena, and d) Brasilia (L. A. Machado et al., 2004). The latter two are further south than the regions, outside of the Amazonian biome, and the first two cities near the edge of the northern and southern regions. The occurrences of the MCSs and other systems fit to the indicated wet and dry seasons quite well (especially obvious for the basin regions a) and c)) and the northern two regions cover the equator, whereas the southern two do not, which are the two reasons why I chose to include the seasons from the two southern cities.

For all regions, except the north-western, the MCSs and other systems have their minimum during dry season, increasing at the end of the dry and throughout the wet season, and with a peak at the end of the wet season. Both have almost identical seasonal cycles for these three regions. The north-western region still shows the maximum of MCSs and other systems at the end of wet season, but the MCSs drop of more gradually throughout the dry season and towards a minimum in the middle of wet season. The other systems have a minimum in during dry season, a maximum at the transition into the wet season, and a minimum in the middle of dry season similar to the MCSs.

The Seasonal cycle for SLs is somewhat different. Additional to the peak at the end of the wet season they also exhibit a peak when transitioning from the dry to the wet season (see Figure 5.4 a), c), and d)) or near the end of the dry season (see Figure 5.4 b), and d) if the dry season from b) is applied). The dry season minimum of SL activity is very pronounced for all four regions and accompanied by a second much weaker local minimum at the beginning or during wet season. Both the second maximum and the second minimum are strongest in the north-western region and weakest in the south western region where the activity almost plateaus during the wet season, only peaking slightly at the beginning and end. The south-western region shows more a slightly ascending plateau followed by a small dip and a large peak during the wet season. This behaviour with clearer seasons in the southern regions and less clear seasons with a dip during wet season correlates with the amount of land north of the equator.

Figure 5.5 shows the occurrences for all three system types. All show a maximum for 2021 and drop of continuously in both directions. The number of SLs is lowest in 2018 and second lowest in 2024, for other systems this is the opposite way around. MCSs show only very little change in occurrence throughout the seven years from 2018 to 2024. The maximum occurrences of all system types coincides with La Niña, although the the lower occurrences coincide both with El Niño and neutral years. The behaviour of MCSs is not consistent with the findings of Sousa et al. (2021), who found maxima for “mesoscale convective cloud clusters” during strong La Niña years. They also found both low and high occurrences of Amazonian coastal SLs for strong La Niña years and no clear correlation between SL and MCS frequencies.

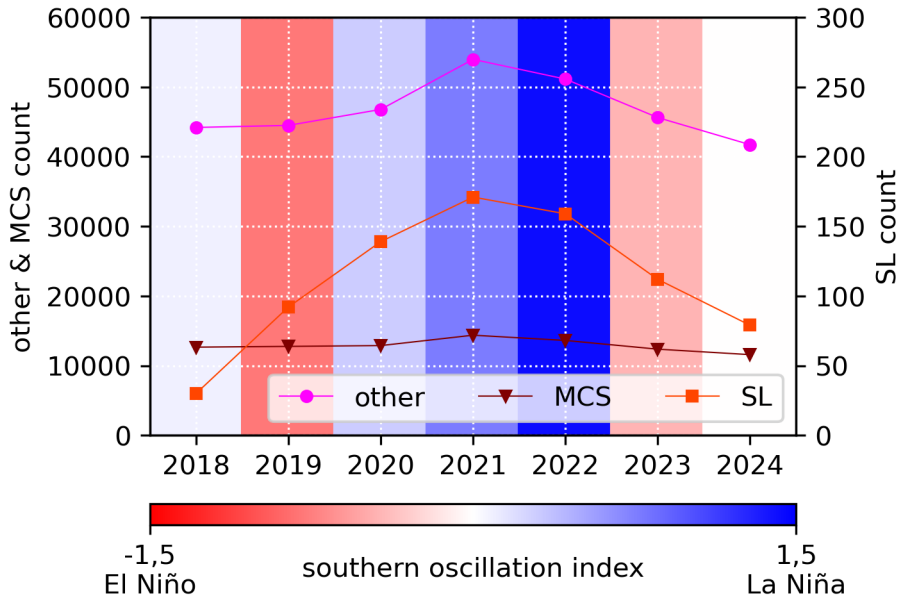


Figure 5.5: Yearly counts of all three system types and yearly mean southern oscillation index indicating El Niño, La Niña, and neutral years (southern oscillation index data according to Ropelewski and Jones, 1987).

## 5.2 Spacial and Seasonal Distributions

The spacial distribution of the different systems can be inferred from a composites of their swaths. For these the number of swaths covering each individual pixel is summed up into a total count of swath coverage per pixel, which can be seen in Figure 5.6. The count therefore directly reflects the number of systems per type detected in each area over the course of seven years.

The distribution for other convective systems shows the highest amount of fine details, followed by the MCSs and the SLs. The Amazon River and its affluents can be clearly seen as minima in the spacial distribution of the smaller, short lived convective systems in Figure 5.6 a) and a bit less clear in b) which displays the spacial distribution of larger and more long lived MCSs. The spacial distribution of SLs in Figure 5.6 c) shows a lot less fine details and more noise, due to their lower count, larger size, and longer lifetime. Still, the Amazon River is identifiable as a minimum, even though much weaker. All three distributions show a large maximum within the central Amazon basin and a smaller maximum at the coast. Between these maxima and to the north of the central basin there is a region with a higher elevation that shows less convective activity. The coastal occurrence maximum of the systems categorised as other is strongly localised at the Amazon delta and less localised, while extending south-east along

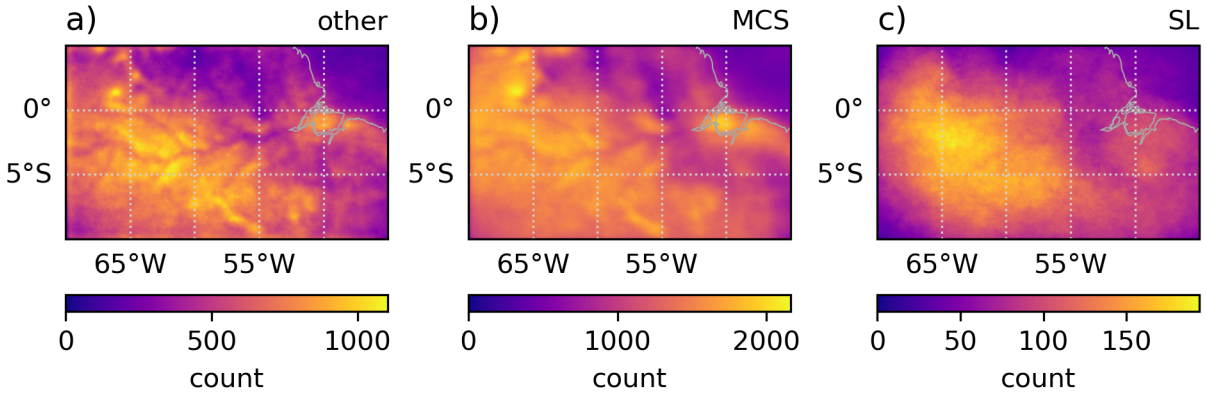


Figure 5.6: Spacial distribution of occurrences for a) other convective systems, b) MCSs, and c) SLs, calculated based on GOES-16 pixels covered by the swaths of the different convective systems.

the coast for MCSs and SLs. The maximum for SLs is less strong, but even further elongated away from the coast towards the south-west. The MCSs and the other systems show strong activity near mountain ridges in the north-west and south. SLs show less activity towards the boundaries of the studied region, where both algorithmic detection is worse and less SLs are expected. The ocean in the north-east shows little activity for all three system types.

All three system types show strong seasonality following mostly the Inter-Tropical Convergence Zone (ITCZ) over its yearly cycle, which can be seen in Figure 5.7. During December, January, and February (DJF) most convective activity is located south of the Equator, whereas June, July, and August (JJA) show most activity north of 5 °S. This explains the dip during wet season discussed in Section 5.1, which is caused by the distinct lack of convection in the northern highlands during DJF. March, April, and May (MAM) show a lot of activity in the entire region, even some activity over the ocean. The coastal activity is highest during this season, with up to 80 % of the yearly activity for SLs (see Figure B.10 d) – f) in the Appendix), followed by DJF. September, October, and November (SON) show a strong difference between the north-eastern and south-western part of the region, with low activity in the north-east and high activity in the south-west. For the purpose of further analysis the seasons will be divided into the seasons from December to May with plenty of coastal convection, and June to November with little coastal convection. Since they mostly coincide with the wet and dry seasons, they will be referred as such hereafter.

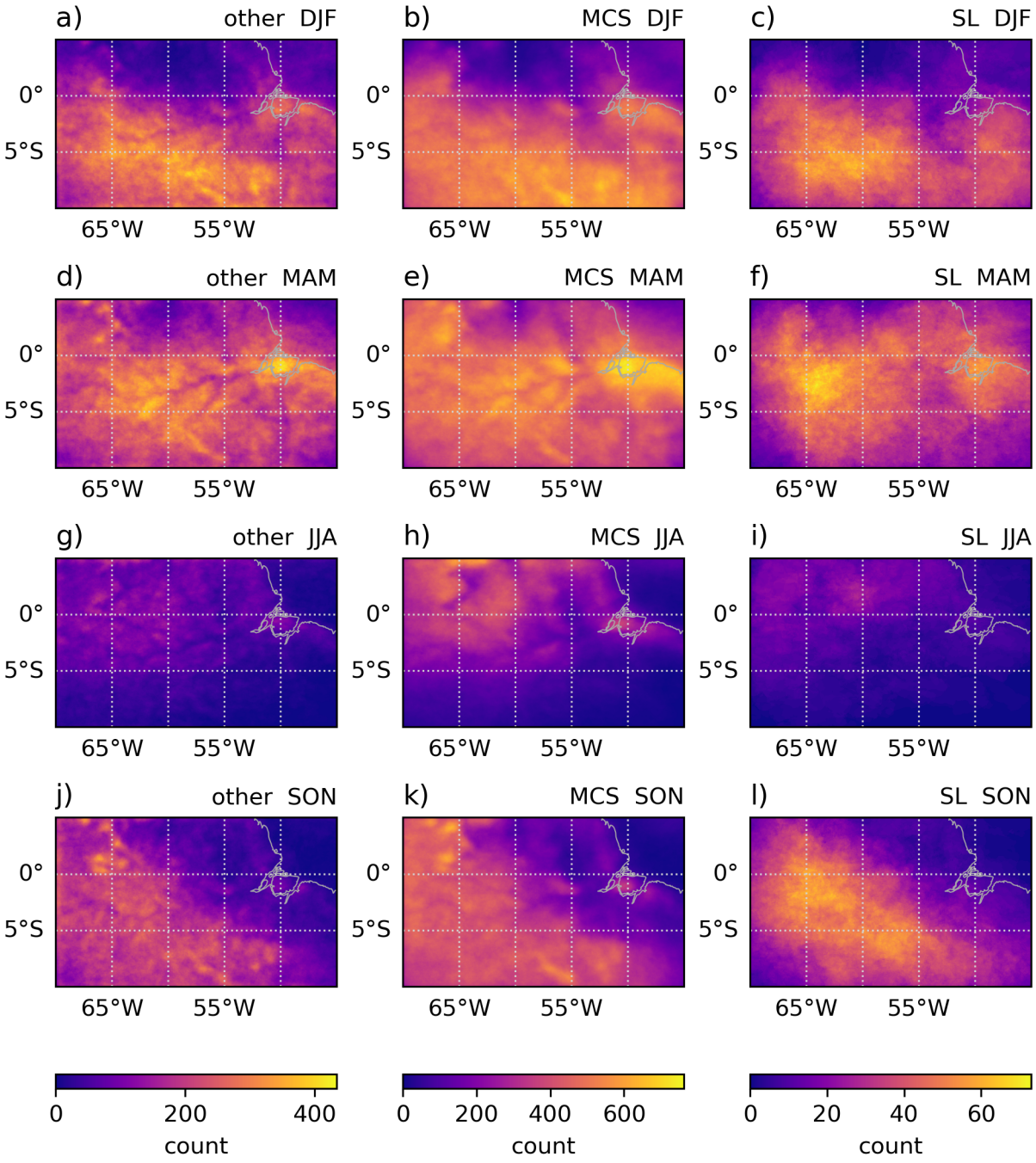


Figure 5.7: Seasonal spacial distribution of convective system occurrences. Same as Figure 5.6, but filtered by seasons. Columns left to right are a)–j) other convective systems, b)–k) MCSs, and c)–l) SLs, and rows top to bottom are a)–c) December, January, and February (DJF); d)–f) MAM; g)–i) June, July, and August (JJA); and j)–l) September, October, and November (SON).

## 5.3 Squall Line Lifecycle

Amazonian SLs have two major formation regions, one being the coast at and surrounding the Amazon river delta, and the other being the eastern and central Amazon basin, as can be seen in Figure 5.8 a). In the region of higher elevation there is an elongated minimum along the solid pink line. This distribution is similar to the occurrences in Figure 5.6 c).

SL dissipation follows a gradient from least dissipation in the east and most dissipation at the western domain boundary, as can be seen in Figure 5.8 b). The latter suggests, that many SLs don't dissipate within the domain, but rather leave the domain through its western border to propagate further into the western Amazon basin and towards the Andes Mountain Range. The solid pink line, which was previously centred on the initiation minimum is at the edge between almost no dissipation in the east and medium dissipation occurrences towards the west.

Considering north-easterly winds, this onset of dissipation during a minimum in initiation at the pink line and subsequently followed by initiation could hint towards type 2 regeneration, where a SL fully dissipates followed by initiation of a new SL along the path of the dissipated SL. In this case the initiation and re-initiation would be about 5° or more apart, which is the distance covered by a SL in 12 hours in the studied domain, with $45\,\mathrm{km\,h^{-1}}$ being close to the expected upper speed limit of a SL (L. A. T. Machado et al., 1998).

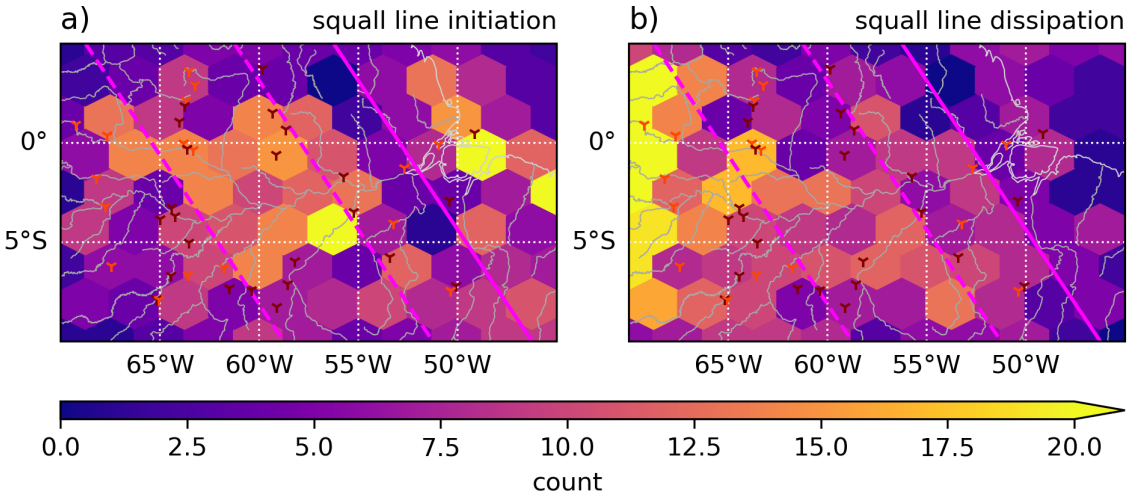


Figure 5.8: Binned spacial occurrences of SL initiation and dissipation, derived from the first and last centroid location of a track. The markers indicate the minima from Figure 5.11 and pink lines are purely visual aids. Gray lines indicate the coastline in the north-east and rivers.

### 5.3.1 Diurnal Lifecycle

As discussed in Section 5.2, there are seasonal differences between the dry and the wet season. For this reason the diurnal cycle is analysed for each season separately. To observe the lifecycle, the individual frames of each SL track were categorised into frames exhibiting clear growth $> 5\,\%/10\text{min}$ compared to the previous, clear decay $< -5\,\%/10\text{min}$, and propagation. The growth rate was limited to values between $-50\,\%/10\text{min}$ and $50\,\%/10\text{min}$, in order to avoid very large rates from splits and merges. The thresholds $\pm 5\,\%/10\text{min}$ are the closest rounded percent value of the 1/3 and 2/3 quantiles of the entire Dataset. Both are rounded by about $0{,}3\,\%/10\text{min}$ benefitting the propagation bin. Similar to Section 5.2 occurrence composites were made, which can be seen in Figures 5.9 and 5.10. These are composites from the individual frames ant not the swaths, therefore the shading correlates with accumulated time covered by SLs and not with occurrence. The four six hour time bins cover evening and midnight (0–6 UTC, 20–2 UTC–4), night and early morning (6–12 UTC, 2–8 UTC+4), late morning and noon (12–18 UTC, 8–14 UTC–4), and mid to late afternoon (18–24 UTC, 14–20 UTC–4). These cover the diurnal cycle of convection, consisting of the peak of activity in the afternoon, followed by reduction and decay, nocturnal boundary layer stabilisation, and a minimum with onset of afternoon convection. During both seasons SL growth happens predominately during the second half of the day, although wet season shows most activity at the coast and some in the basin, whereas dry season shows most growth in the basin. Minor differences are SL growth in mountainous regions during dry season and nocturnal SL growth over the ocean. These seasonally differing growth patterns result in different propagation and decay patterns over the diurnal cycles.

During the wet season the nocturnal and morning/noon SL propagation off the coast of Brazil is directly adjacent to the coastal SL growth during morning/noon and afternoon. SL propagation and decay are located in the same coastal region during the afternoon, with higher values at the coast south of the Amazon delta. During evening/midnight (0–6 UTC) the northern coastal SL propagation strengthens and moves several degrees to the west, while the southern propagation also detaches from the coast to move inland. During the night and early morning (6–12 UTC) both propagation maxima are again located further away from the coast than during the prior time bin. During morning and noon (12–18 UTC) the signal of the coastal SLs has weakened and merged with the maximum in the central basin, with the southern maximum potentially moving south. The basin maximum regains strength during the afternoon

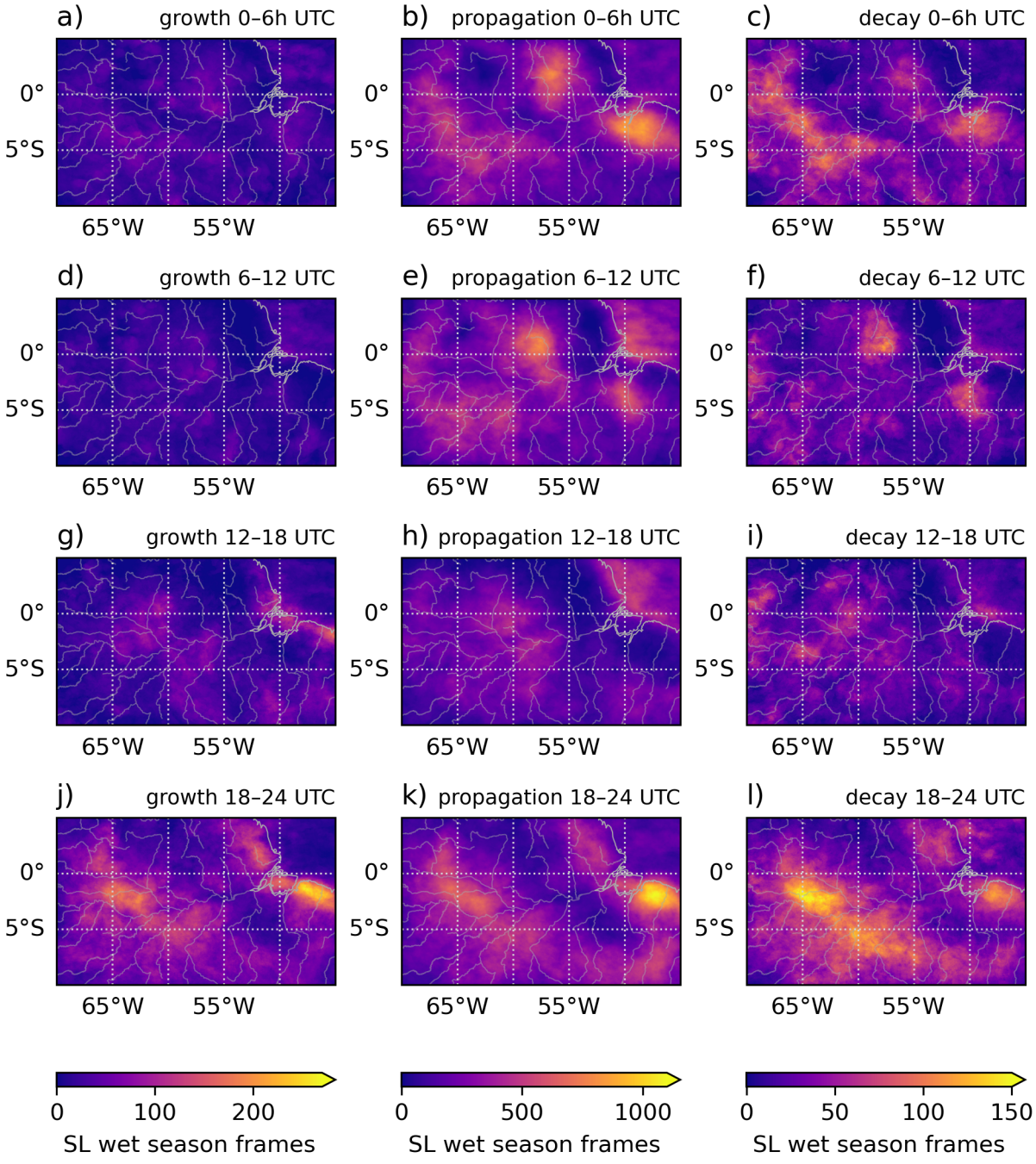


Figure 5.9: SL occurrences during wet season (December to May), by time of day in rows and change in size in columns. Growth is any size increase between 5 and 50 %/10min and decay any reduction in the same range. Propagation refers to little change in size of −5 to 5 %/10min.

(18–24 UTC) and propagates westwards to meet the western domain boundary during the night or early morning (6–12 UTC), while widening and becoming more diffuse. This creates a two day cycle of SLs initiating at the coast and travelling westwards through the domain, following the the predominant weather direction (Anselmo et al., 2021). The decay maxima of the SLs throughout the whole day and night are co-located with the propagation and are especially strong during the afternoon. This was discussed previously in Section 5.1 and can be attributed to problems in the detection caused by the noisy afternoon convection, or the decay of the SLs, which could also be due to the widespread convection in the vicinity.

During dry season most signals are more stationary. South of the Amazon delta is still a signal propagating inland, but much weaker and diffuse. During the evening, night and early morning a maximum in both growth and propagation of SLs can be seen in the north west, which is the strongest nocturnal growth throughout the year. It is probably related to the adjacent mountain ridge and too close to the domain boundaries for any clear correlation with other features. Again the decay patterns match the propagation patterns, although the night and early morning (6–12 UTC) are an exception. During this time there are propagation maxima both in the center and close to the western domain boundary, but the central maximum isn't matched by a similar decay maximum. This might hint towards nocturnal stabilisation of SLs in the during dry season.

Compared to SL, the MCSs show strong afternoon growth, propagation, and decay throughout the whole domain only avoiding the region of higher elevation and the ocean (Figure B.7 in the Appendix). The coast and mountain ridges show strong growth, but MCS propagation doesn't reach as far into the basin and decays faster. The ration between coastal to basin MCS propagation is much higher than for squall lines, indicating a high amount of very localised coastal convection leading to MCSs. The decay however is lower and could hint towards merges with SLs, which would agree with previous findings that four "convective cloud clusters" in this region typically merge into a SL (Sousa et al., 2021).

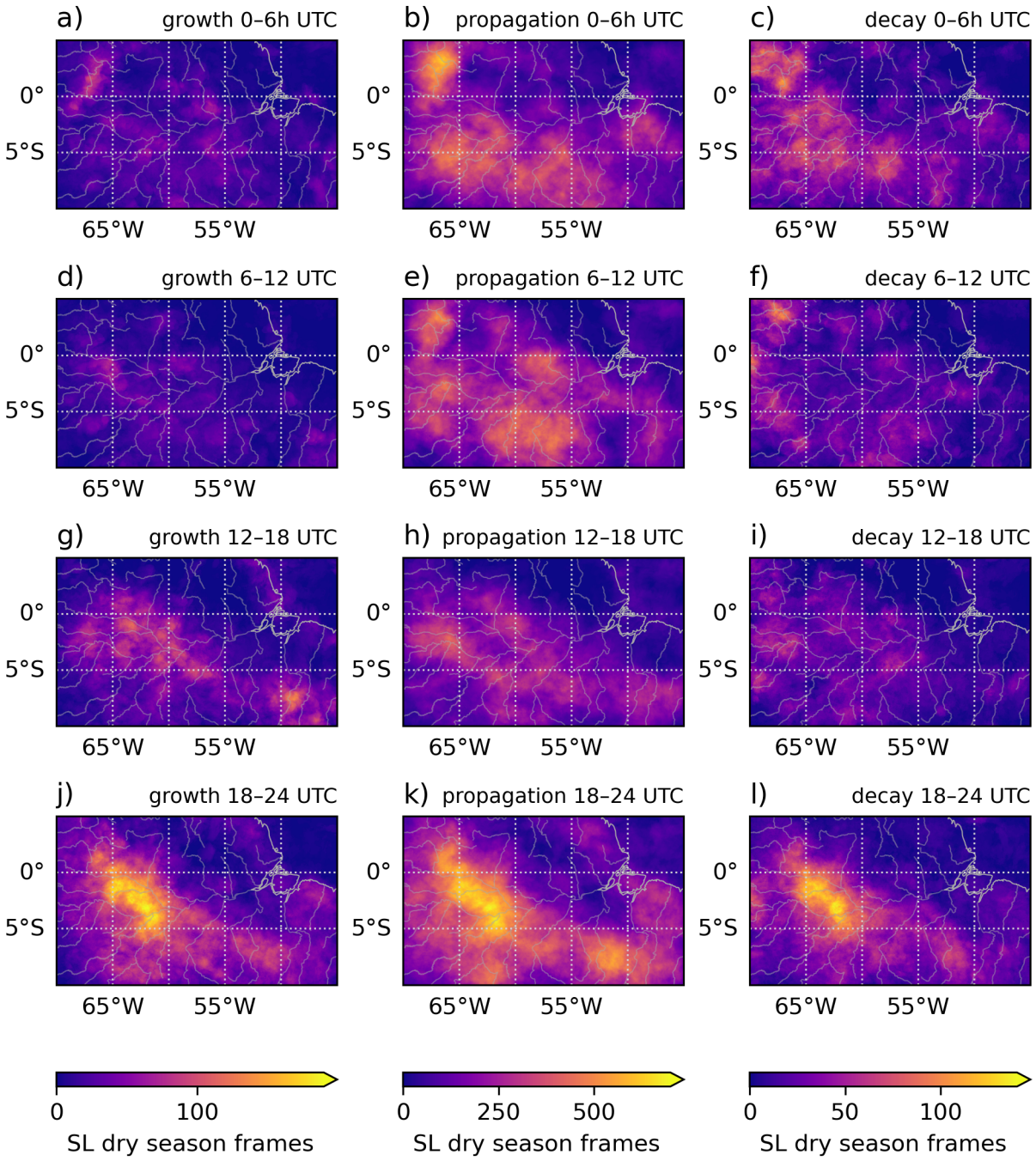


Figure 5.10: SL occurrences during dry season (June to November), by time of day in rows and change in size in columns. Refer to Figure 5.9 for further information.

### 5.3.2 Type 1 Regeneration

The type 1 SL regeneration consists of a period with weaker convection before the squall line regains strength. For the detection of these types of regeneration, each individual SL track was analysed by a peak finding algorithm (Virtanen et al., 2020). It was configured to look for maxima in brightness temperature with a prominence of 20 K, which should indicate a significant weakening in convection. I further assured, that SL stages were detected both before and after the peak. Figure B.3 a) in the Appendix shows, that the minimum cloud brightness temperature for squall lines is usually in the range 180–190 K for SLs, 190–200 K for MCSs, and 200–210 K for other convection. Thus an increase on brightness temperature by 20 K should indicate a significant loss in organisation. The peaks had to be at least 6 time steps wide and apart from each other, to avoid detecting noise as signals.

A total of 41 type 1 regeneration events were detected in 37 SL tracks, which is about 5 % of all SL tracks. These tracks exhibited mostly negative median zonal speeds (westwards propagation) and a distribution of median meridional speeds centred around 0. Figure 5.11 shows the locations and time of all detected minima, color coded dark for night and morning (3–15 UTC, 23–11 UTC–4) with little and bright for noon until evening with lots of convection

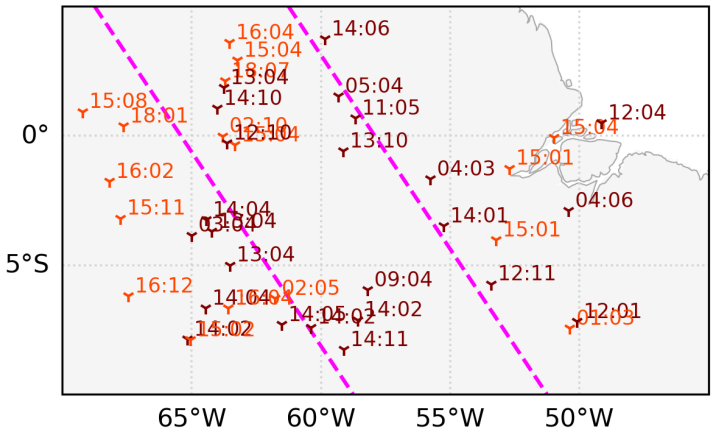


Figure 5.11: Location and time in UTC of prominent maxima in the minimum system brightness temperature of SL tracks. Darker marker color indicates times between 03:00 and 15:00 UTC to imply night and early morning and coinciding with the strongest time period of the Amazonian low level jet.

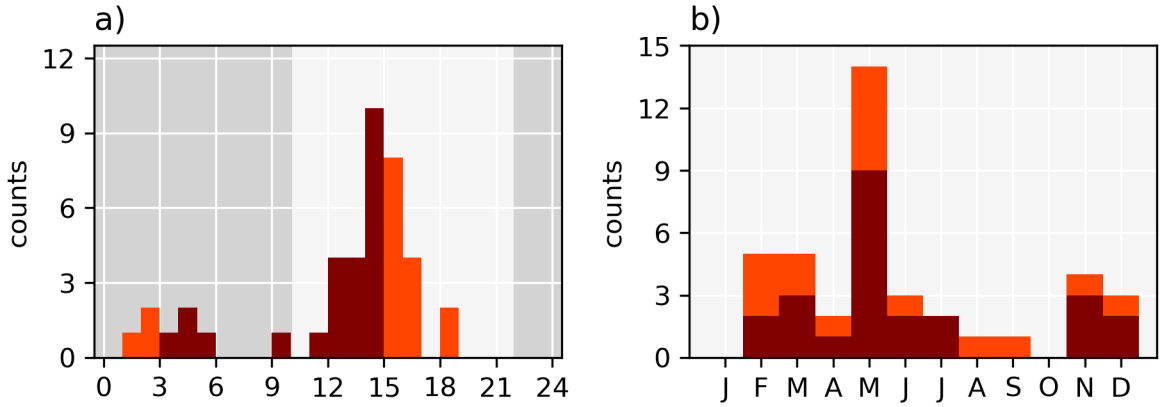


Figure 5.12: Diurnal and seasonal occurrence of detected regeneration events. The time between 18:00 and 06:00 UTC–4 has a darker background shading to imply nighttime, the colors of the histogram are the same as the colors in Figure 5.11.

during the diurnal cycle (see Section 5.1). The regeneration events during night and morning are located in an almost circular shape around the center of the domain, the type 1 regeneration from noon until evening are located mostly to the west and some to the east of the circle formation. The dashed pink lines are semi parallel to the coast and through a majority of the nocturnal and morning regeneration events. The same lines can be found in Figure 5.8, where they are located mostly at the edges of the SL initiation maximum in the central basin area.

The diurnal cycle of the detected type 1 regeneration events favours the time around local noon, but already starting during the morning hours, coinciding both with the increase of available energy from the sun light and the strongest phase of the Amazonian low level jet (Anselmo et al., 2020). The minimum of regeneration events coincides with the maximum of convective activity during the mid to late afternoon and evening. Reasons for this can be a lack of regeneratable SL, which would have already regained strength earlier during the day, but also include the weak phase of the Amazonian low level jet. A smaller group of 10 type 1 regeneration events was detected in the middle of the night, which could be related to the strengthening of the low level jet in the central basin region after local midnight.

The seasonal cycle of type 1 regenerations follows in general the seasonal cycle of SLs, a major exception being the month of April. This month shows a lot of SL activity (see Figure 5.4, but almost no type 1 regeneration events. Furthermore, the dry season minimum of type 1 regeneration is shifted about a month back, compared to the SL activity.

Type 1 regeneration events occur over the timespan of several hours, starting with an increase in minimum and mean brightness temperature 6 or more hours before regeneration, indicating a weakening of the convection (see Figure 5.13). This is followed by a reduction in size and length, both of the entire squall line and the region of strong convection with a brightness temperature below 210 K within it. Once the reduction in size begins about 4–5 h before the minimum, the area and length of the cold parts drop faster and reaching their minimum at about the same time as the brightness temperatures peaks. The minimum of area and length of the entire SL lag behind by about 1 h, which is the time the outflow of the convective cores need to influence the edges of the SL. The density only dips for a short time, implying signs of dissipation, and slowly regaining density about 3 h after the regeneration. The convex hull aspect ratio shows an initial gradual increase and later decrease with no clear structure like a dip or peak. The systems exhibiting more extreme regeneration events with a length below 100 km during their regeneration probably appear as a small convective system, whereas the larger systems with a length above 400 km may still resemble a SL.

On average the area of the SL drops by about 74 % over seven hour and its length by about 44 %. The area of the colder parts of the SL reduces slightly more in size by 79 % over six hours, the length reduces by 69 %, which is significantly more. Most shape variables, but especially the length show a stronger increase in size than the initial decrease. The length of the SL is nearly 15 % larger 6 h after the regeneration compared to 6 h than before. The area shows no increase, being 7 % smaller than initially, but still showing a growth trend. The colder regions with strong convection show over 18 % length and only 2 % in area. This goes in line with the time of day most type 1 regenerations happen, where they slowly weaken towards their minimum in occurrence before local noon. Subsequent strengthening through increased convection during noon, afternoon, and into the night happens faster.

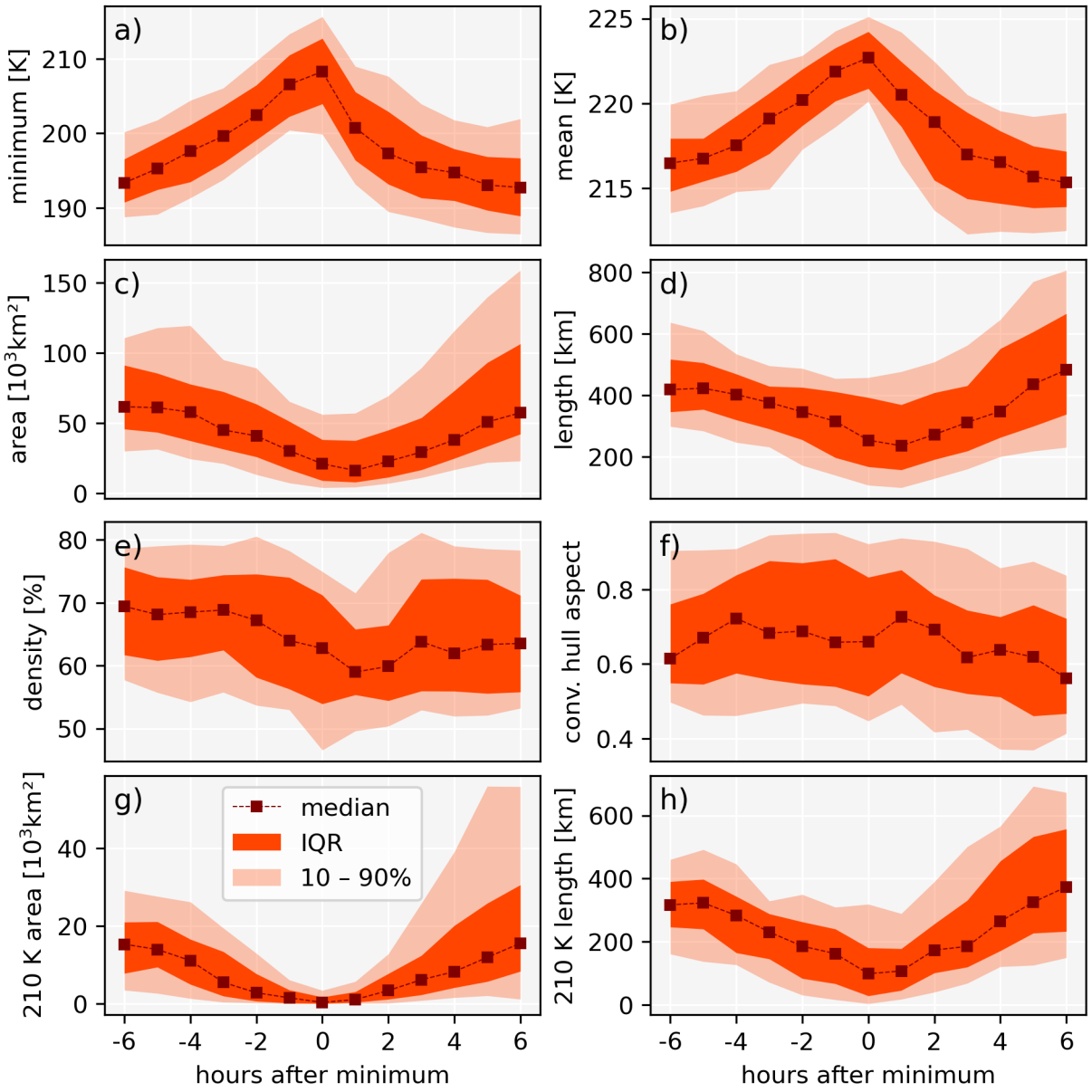


Figure 5.13: SL regeneration properties from 41 regeneration events binned hourly for 13 h centred around the regeneration event. The subplots show a) minimum brightness temperature, b) mean brightness temperature, c) area and d) length of the SL, e) the convex hull density, f) the convex hull aspect ratio, and g) area and h) length of the region with stronger convection with a brightness temperature lower than 210 K within the SL.

# Part 6
# Discussion

The development of an algorithm to automatically detect and track coherent features such as SL was overall successful, although there are some caveats. Like all algorithms it relies on some fixed thresholds and by detecting larger coherent features it may also cluster together what doesn't belong. This is dependent on the parameters applied to DBSCAN and even though my choice is reasonably based on previous studies, they sometimes produce very large coherent features. Whilst I implemented the subdividing method to mitigate this problem, it may introduce bias into the data based on the triggering thresholds. Although, this is partially mitigated, by the subdividing method not forcing systems apart, if they show strong coherency within their convective structure. The density distribution (Figure B.2 g)) shows some influence from the subdividing, but wasn't used in the further analysis and its impact on other variables should be small. The maximum system area was artificially limited by the same means as well, but since only a single track managed to exceed this limit after attempted subdividing, the threshold choice of $5 \cdot 10^5\,\mathrm{km}^2$ seems justified as upper limit.

A problem with the algorithm is the aforementioned bias of the central line towards 0° or ±90°. Fortunately, due to the nature of the Chi-squared optimisation of linear fits, this affects rounder systems with aspect ratios near 1 stronger than elongated systems with lower aspect ratio, which means that the stronger a systems is affected, the smaller the error. Still this should be fixed for future studies.

The tracking relies on less thresholds and the underlying methods have been historically proven to work for large systems. Overlap based tracking has been used by L. A. T. Machado et al. (1998), Mathon and Laurent, 2001, and more recent Anselmo et al. (2021), just to name a few. It is rather insensitive to the threshold used and can even track splits and merges. In the context of my algorithm the tracking relies on the data from the analysis step and consequently any uncertainty there is, may affect the tracking and result in increased numbers of splits and merges. This however should have little impact in the analysis, since the setup ensures that

larger systems keep continuity. Larger systems themselves may split and later re-merge, although it can be argued, that this is intended behaviour in such a case. This behaviour leads to noisier tracks, but since the further analysis is performed in a manner that individual frames have little weight, the impact should be negligible.

The length of tracks, the tracking overlap and the track delay (Figure 5.2) further reflect on the tracking performance, with the latter to showing nothing unexpected. The tracking did produce about $1 \cdot 10^5$ tracks with only a single cloud feature (length 1). But compared to the total number of close to $5 \cdot 10^6$ detected cloud features, this affects the statistics by only about 2 %. All of these short lived tracks were classified as other, which in itself is a category consisting of a mix of systems not classified as SL or MCS. Within this category they contribute about 31 % of the tracks and much less in terms of cloud features. Because the study is not aimed at smaller systems, such short lived behaviour is expected. The other category also includes short tracks of larger systems that split from or merged with a MCS or SL, making this category truly a mix of systems. The results however show, that this category mostly reflects the behaviour of smaller convective systems, and as long as its limitations are kept in mind, it can be argued as such.

The classification is tuned on a manually identified dataset. After tuning about 20 % false negatives and about 22 % false positives remain, leading to some missed SLs and some MCSs misidentified as SLs. Considering that the manually identified data may have missed a few SLs, and that the classification predicted the total SL number for the manually identified year quite well, my analysis is rather at the lower end of the expected range of detected SLs. It detected approximately 112 SLs per year or one every three days, which is close to the 124 per year reported by Sousa et al. (2021) and lower than one squall line per two days reported by Alcântara et al. (2011). Both of them focussed on the coastal region, and contrary to my analysis they didn't account for SLs initiated in the basin. This adds to the assumption that the total number of SLs may be higher, although there are no studies covering the same time period. Nevertheless, the total counts of tracks of all three classes should supply a solid base for statistical analysis.

Over all, the results show clear differences between all three system types which meet the expectations. The lifetimes and sizes of MCSs are larger than of the "other" category consisting of mostly individual cells. The mean size and lifetime values are mostly in the same range as previous studies, although the lifetime is a bit larger (L. A. T. Machado et al., 2018; Anselmo

et al., 2021). Since maximum MCS lifetimes also agree with previous studies, the mean lifetime is within an acceptable range. For the system size doesn't show the same behaviour, this is probably not a result of tracking coherent features instead of contiguous features. The results show further a clear distinction between the diurnal cycles of MCSs and other, while their seasonal cycles match well.

The mean lifetime of SLs is also in agreement with prior studies in the Amazon region (Cohen et al., 1995), their length and area however are less easily compared. The size of SLs varies widely, ranging from hundreds to over a thousand kilometres (Cohen et al., 1995; Rickenbach, 2004). This range is matched well by my results, validating further my approach of tracking coherent features. The inter-annual results of Sousa et al. (2021) seem to somewhat contradict mine, since they found no strong correlation with La Niña but my results seem to support such a conclusion. However, I managed to cover only a single La Niña phase and did not control for other influences. With neutral years being the years with lowest SL occurrence, my results also suggest, that other factors play a role. The spacial distribution of SL initiation and growth at the coast is in agreement to Oliveira and Oyama, 2019 and Sousa et al., 2021, lending credence to my results of spacial distributions. The seasonal cycles further match wet and dry season patterns of cities in the vicinity, indicating strong regional homogeneity and supporting my results (L. A. Machado et al., 2004).

My results show, that the diurnal cycle of SLs differs significant from the typical diurnal cycle of convection in the Amazon basin. The low variance in occurrence throughout the day can be attributed to the long life times of squall lines. Simply put, a SL detected at any time during the day is likely to persist throughout an entire diurnal cycle. The nocturnal maximum in occurrence coincides with the timespan where two spacial maxima are visible during wet season, which is the season most SLs occur. The diurnal cycle of the spacial evolution of SLs further shows a strong link between basin occurring SLs and coastal SLs. They propagate away from the coast throughout the night to mostly dissipate in the late morning, although some traces could still be detected during that time. A few hours later, the maximum growth of SLs can be observed in continuation of the coastal SLs. Considering the strength of the occurrence maxima, I propose that most westwards propagating basin occurring SLs are a direct consequence of a coastal SL during the prior afternoon and night.

The detected type 1 regeneration events support this conclusion as well. They are distributed along the paths of the nocturnal SL propagation maxima and in the central basin, co-located with the respective growth maximum. Both diurnal and seasonal distribution of type 1 regeneration matches as well. Unfortunately I couldn't retrieve exact locations of type 2 regeneration events from the tracking dataset yet, but the locations of dissipation and initiation of squall lines put their locations in the same region. The properties of SLs during type 1 regenerations show massive reductions in size and activity, with the lower 10 % being even lower. If this is consistent for even further decay and subsequent re-initiation, type 1 and 2 regeneration events may be governed by the same mechanics and just different magnitudes of the same kind of event.

The statistics of type 1 regeneration are consistent with the expectation that it is governed by the strength of the convective cores. Consequently the larger 230 K cloud shield of the system follows the changes of the convective 210 K area of the system with a delay of about 1 h. With the density of the system (within its convex hull) as an indicator of spacial cohesion, a loss thereof during the time of lowest convective activity and reorganisation about two hours after regeneration was observed. This would further indicate, that type 2 regeneration events could be a consequence of a decaying SL further loosing spacial cohesion before regeneration.

## 6.1 Outlook

The algorithm which was developed within the scope of this study is in a working condition, but not yet finalised. Aside from fixing the aforementioned problem of the detection of the central axis of a system, it might further benefit from some additional features. And furthermore it has yet to be released to the public domain.

The sensitivity of the results to the choice of threshold values should be assessed. This can be done by varying individual thresholds and determining the sensitivity of the results to the threshold. Another method is a direct comparison with other Algorithms.

Extending the dataset by using older satellite data could yield further interesting results and improve the statistics. It would further allow to compare inter-annual results to older studies.

Finally another goal for the future is to detect individual type 2 regeneration events. Performing case studies by simulating regeneration events could be a further analysis based on this study.

## 6.2 Conclusion

Within the scope of this Thesis, I developed a new algorithm specific for the detection of SLs. The goal of detecting larger coherent features and link them together into long continuous tracks could be achieved. By applying this algorithm to TOA brightness temperature data, I derived a large dataset of tracks categorised into SL, MCS, and other tracks. Climatological findings derived from this dataset match other studies and the total counts suggest solid statistics. The following findings were made with a varying degree of certainty.

- The diurnal occurrences of SLs differ from MCSs and other convective systems. Even though all share a minimum before local noon, SLs peak significantly around local midnight.
- The results somewhat support a link between La Niña and SLs, but other factors may play a major role as well.
- Convective systems in the Amazonas region adheres to climatological convection phenomena around rivers, lakes, and mountains, although their influence decreases with increasing convective system size. Latter are probably not able to "resolve" smaller surface features.
- SL lifecycles differ significantly during wet and dry season. During wet season the spacial distribution exhibits a coastal and a basin maximum and clear propagation patterns. During dry season the activity is generally lower, and mostly focussed in the basin region with no clear propagation patterns.
- Basin occurring SLs have a strong link to coastal squall lines. They may be direct continuations after a minimum during the late morning.
- Several instances of type 1 SL regeneration have been observed. As expected, they are governed by the convective cores within the SL. Results further indicate a stronger growth by 15 % in length after the regeneration, compared to the initial decay.
- Finally the results suggest, that type 2 regeneration may be similar to type 1 regeneration. Their main difference may consist in how far the decay is able to progress before convection is re-initiated.

Due to the difficulties in detecting SLs, the generation of a large dataset of SLs with plenty of diagnostic data is one of the achievements of this study. Further analysis of this dataset may yield interesting insights into type 2 SL regeneration.

# Appendix A
# Abbreviations

| | |
|---|---|
| **DBSCAN** | Density Based Spatial Clustering of Applications with Noise |
| **DJF** | December, January, and February |
| **ForTraCC** | Forecast and Tracking the Evolution of Cloud Clusters |
| **GOES-16** | Geostationary Operational Environmental Satellite 16 |
| **ITCZ** | Inter-Tropical Convergence Zone |
| **JJA** | June, July, and August |
| **MAM** | March, April, and May |
| **MCS** | mesoscale convective system |
| **MERRA-2** | Modern-Era Retrospective Analysis for Research and Applications, Version 2 |
| **NND** | nearest neighbour distance |
| **pyForTraCC** | python based ForTraCC (see ForTraCC) |
| **SL** | squall line |
| **SON** | September, October, and November |
| **TOA** | top of the atmosphere |
| **tobac** | Tracking and Object-Based Analysis of Clouds |
| **TOOCAN** | Tracking Of Organised Convection Algorithm through a 3-D segmantatioN |

Appendix B

# Additional Figures

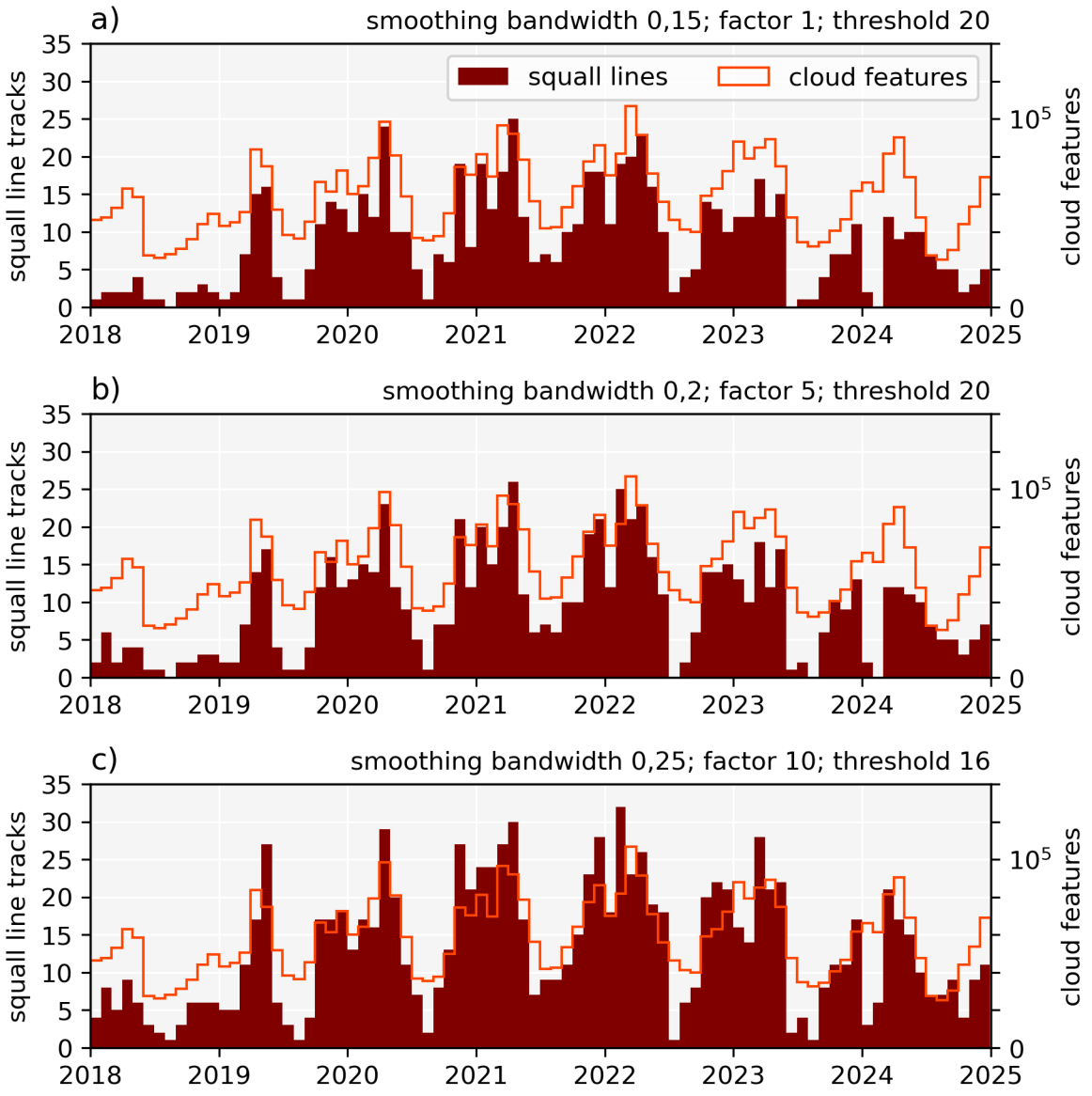


Figure B.1: Monthly number of SL tracks derived from the three filters in Table 4.3 and monthly totals of individual detected cloud features.

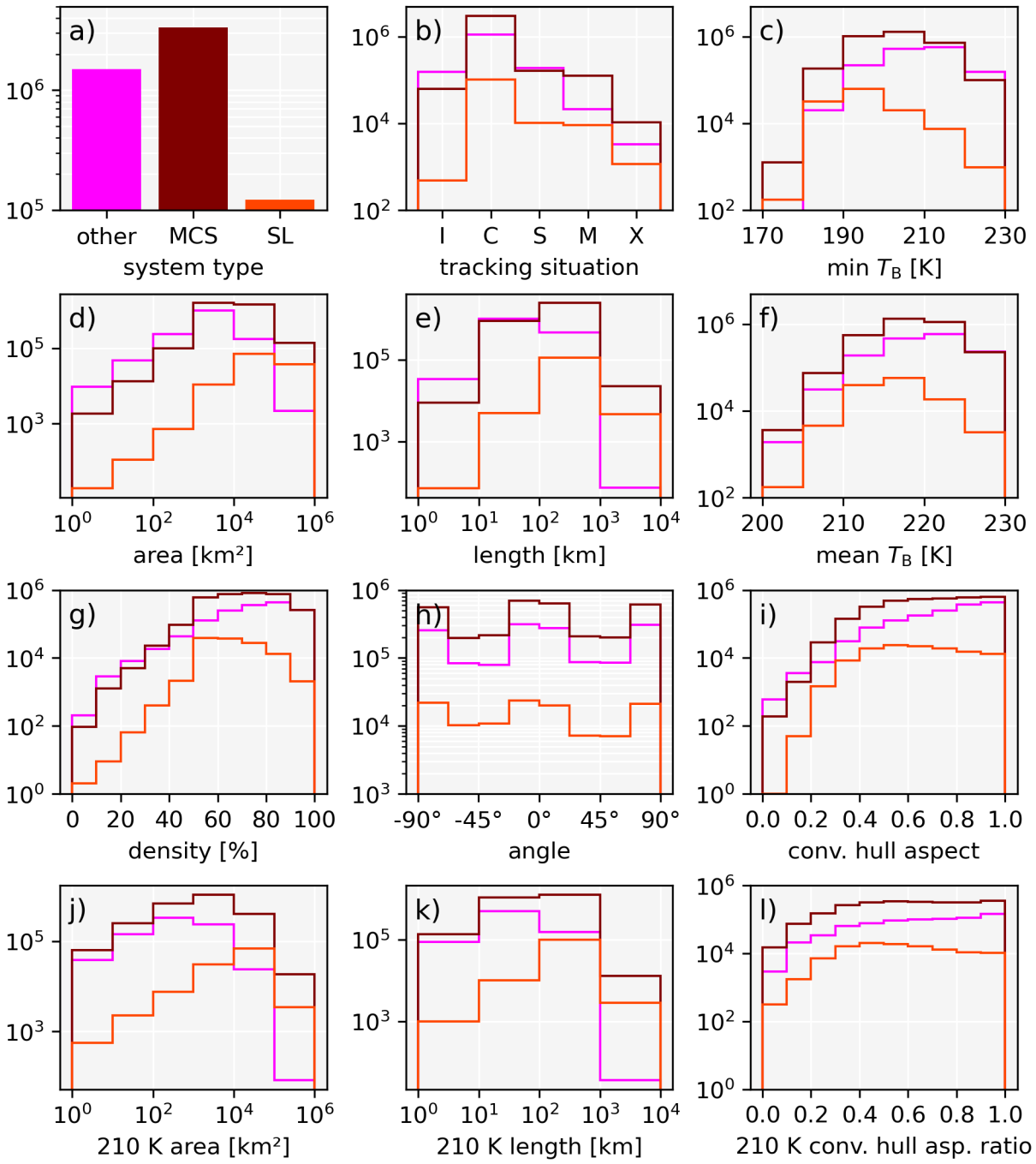


Figure B.2: Histograms of different variables based on the individual cloud features of all three system categories.

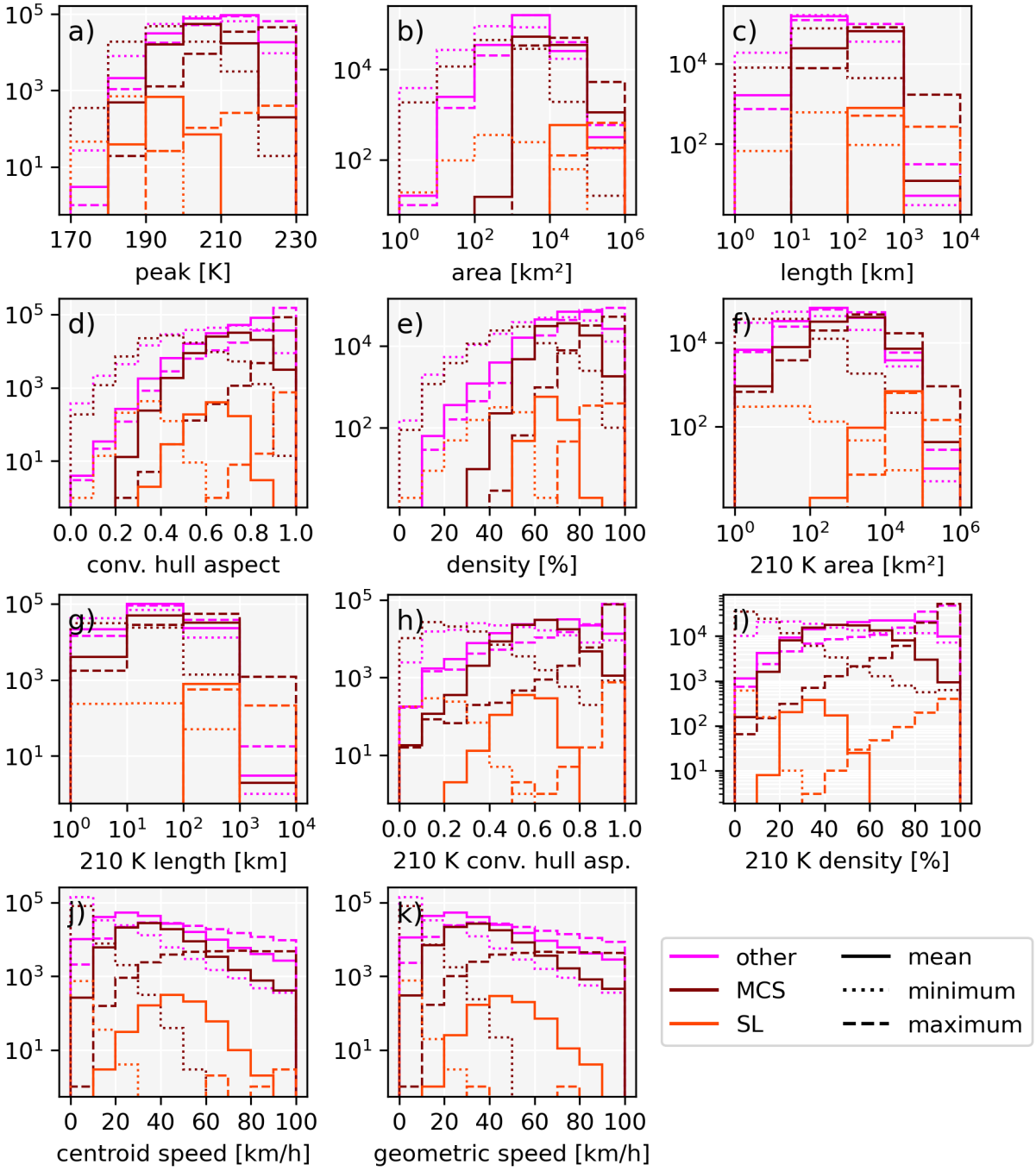


Figure B.3: Histograms of track based minima, maxima and means of different variables for all three system categories.

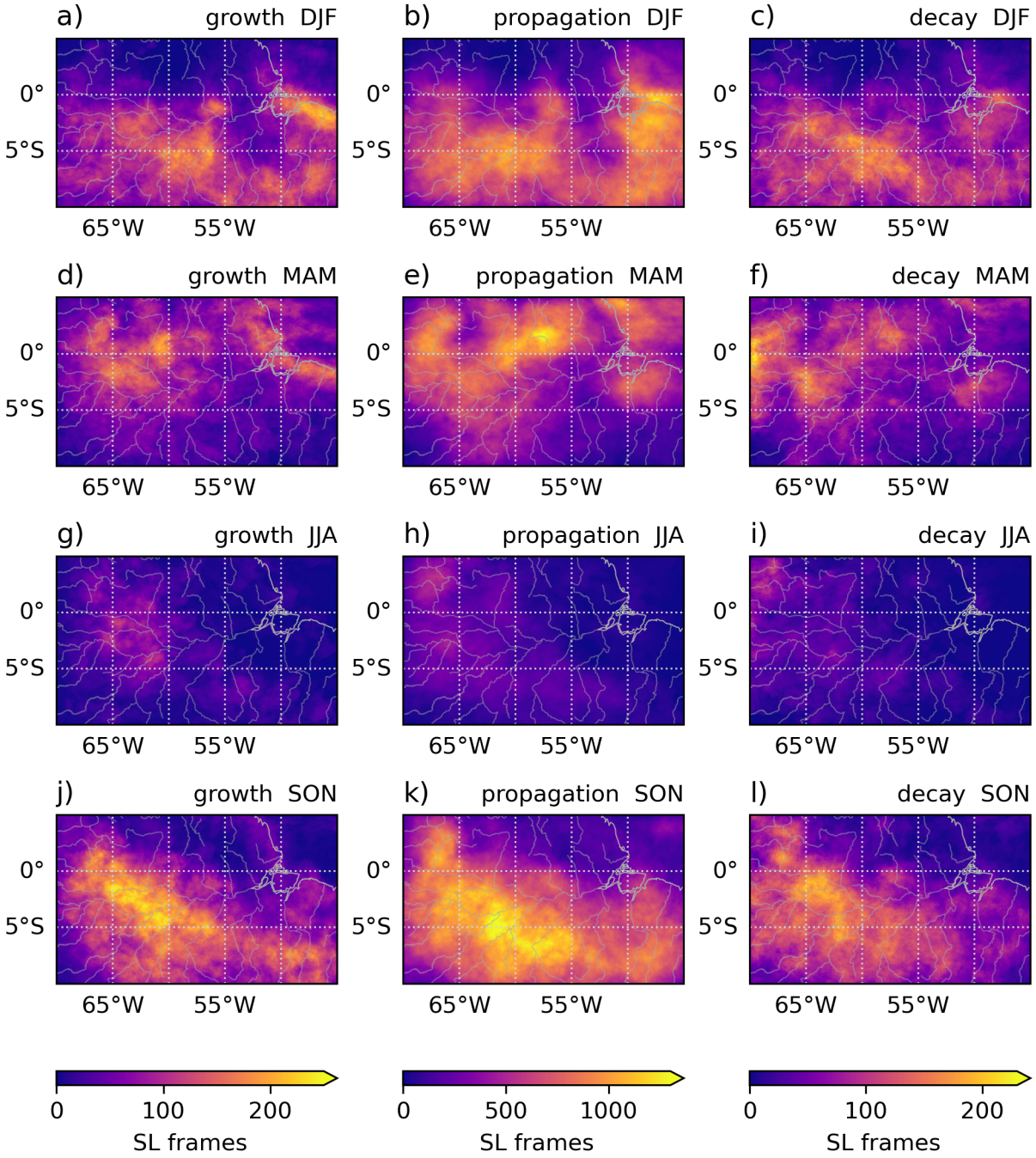


Figure B.4: Seasonal spacial distribution of SL lifecycle.

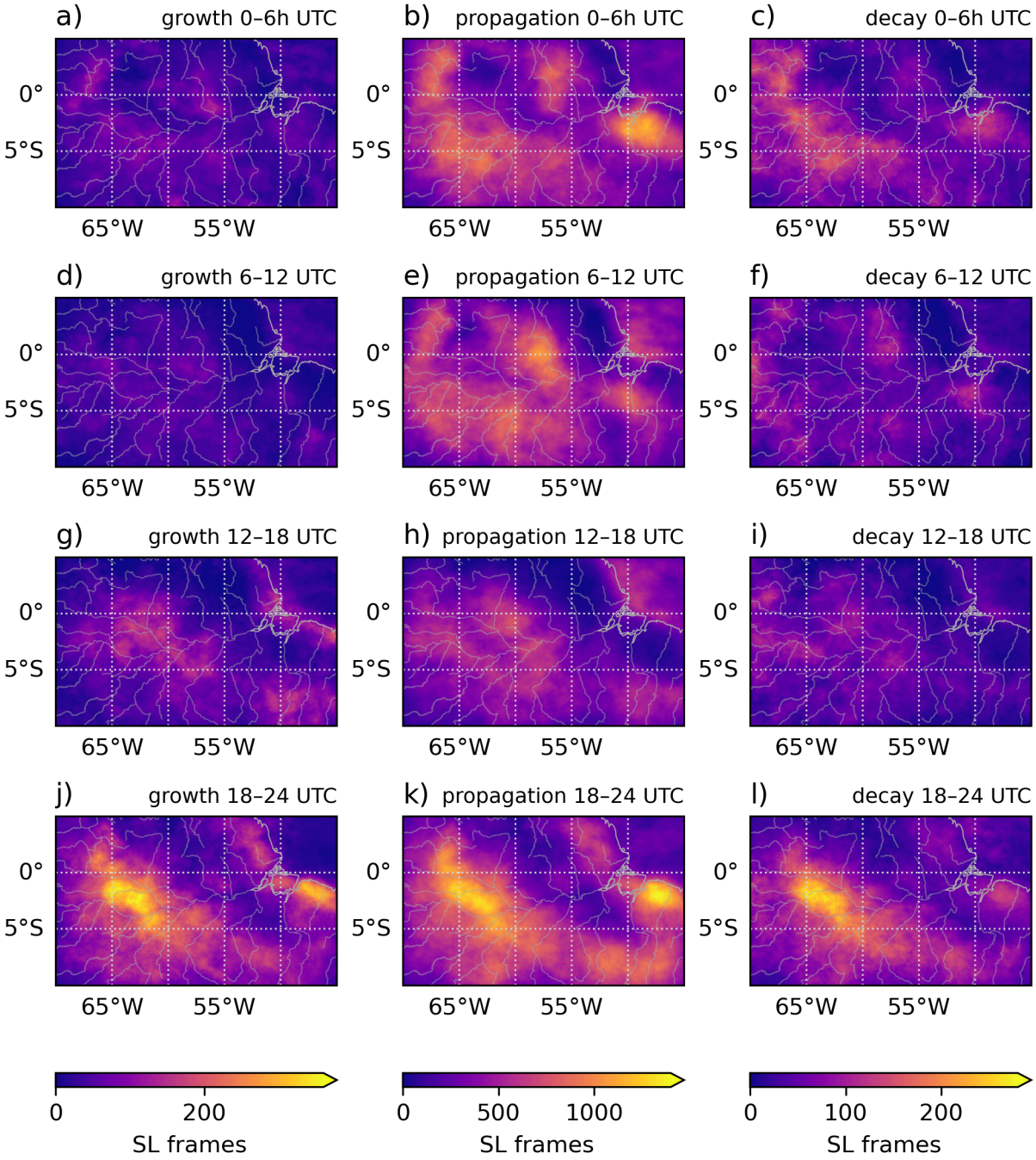


Figure B.5: Diurnal spacial distribution of SL lifecycle.

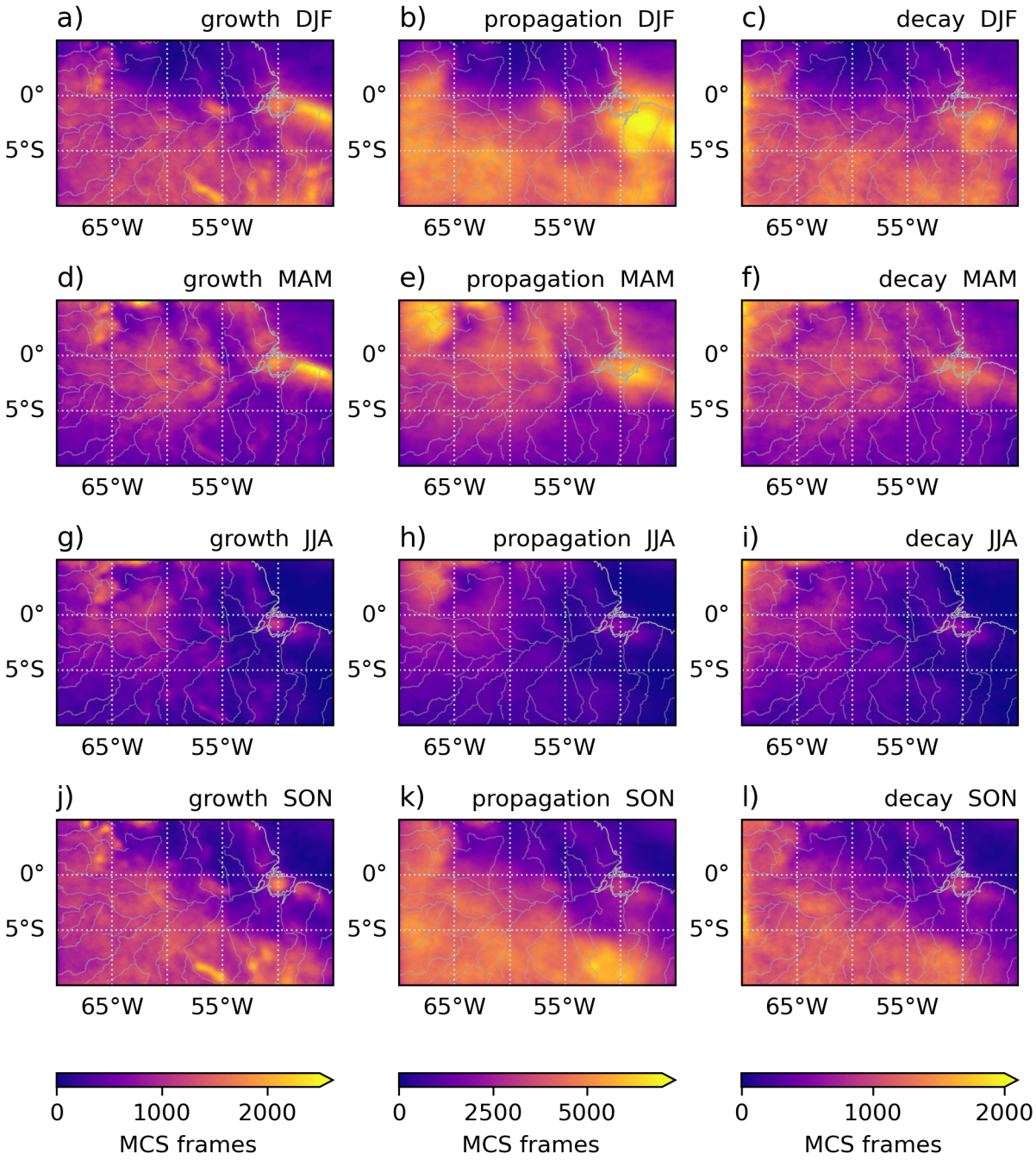


Figure B.6: Seasonal spacial distribution of MCS lifecycle.

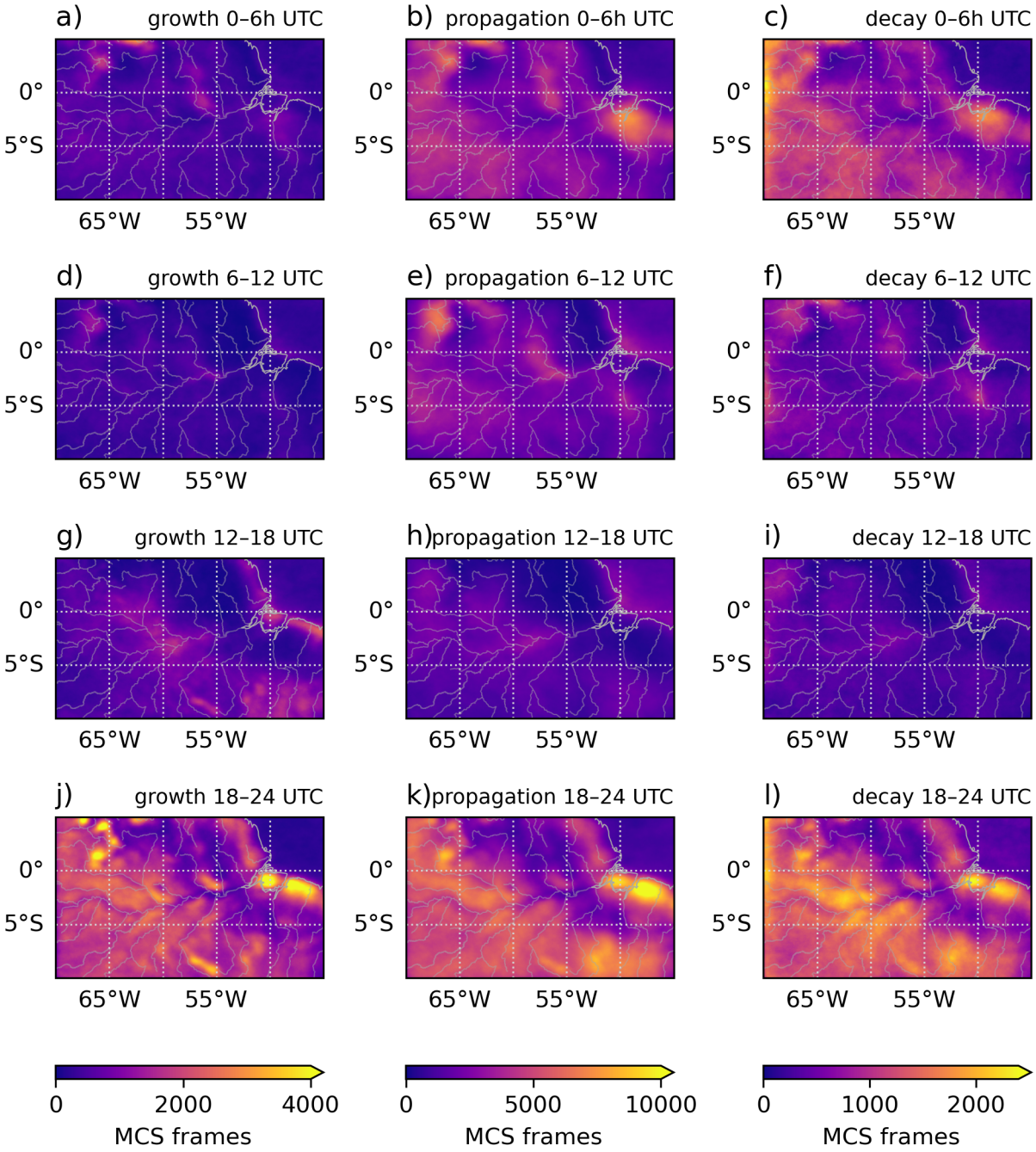


Figure B.7: Diurnal spacial distribution of MCS lifecycle.

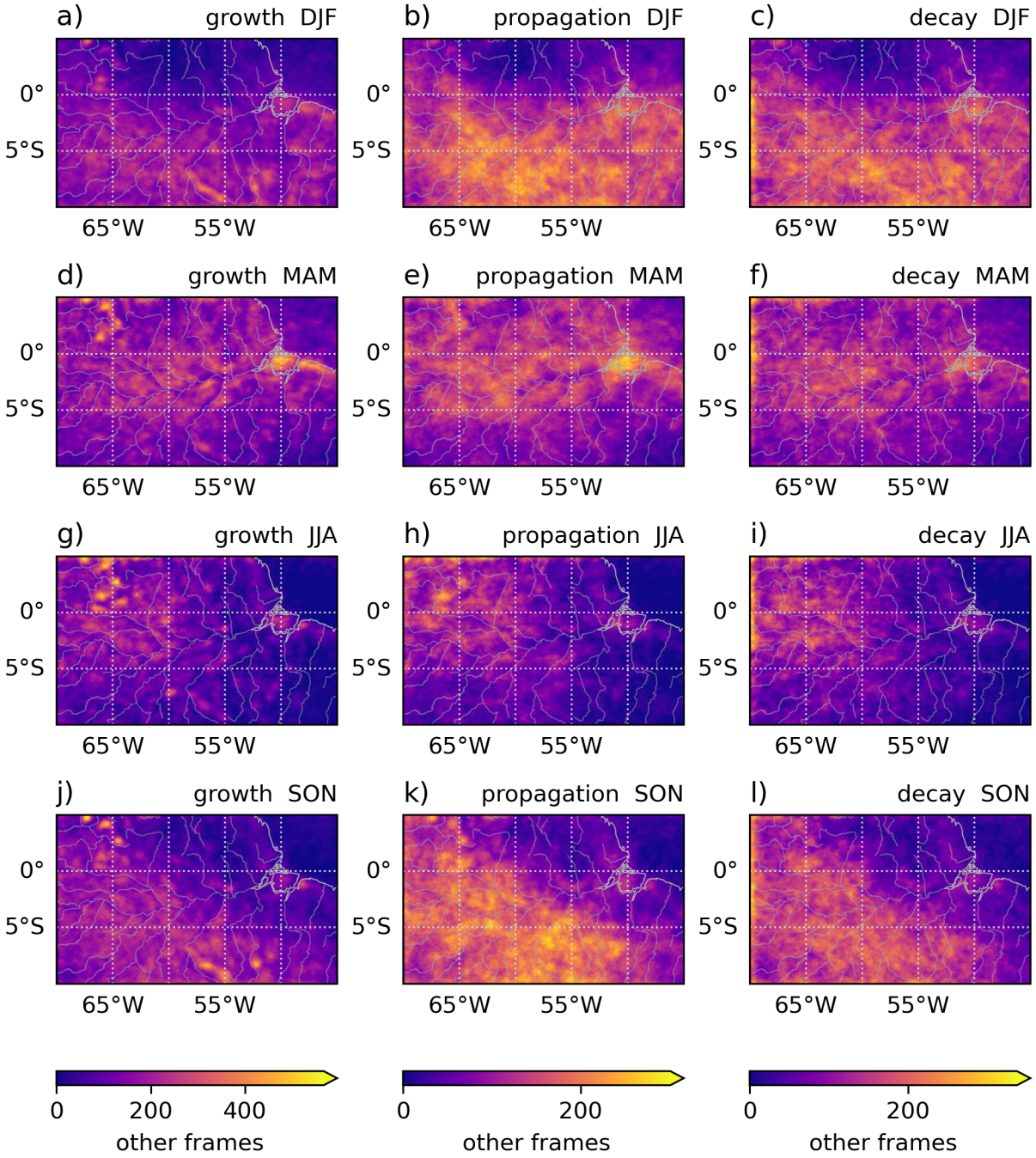


Figure B.8: Seasonal spacial distribution of the lifecycle of other systems.

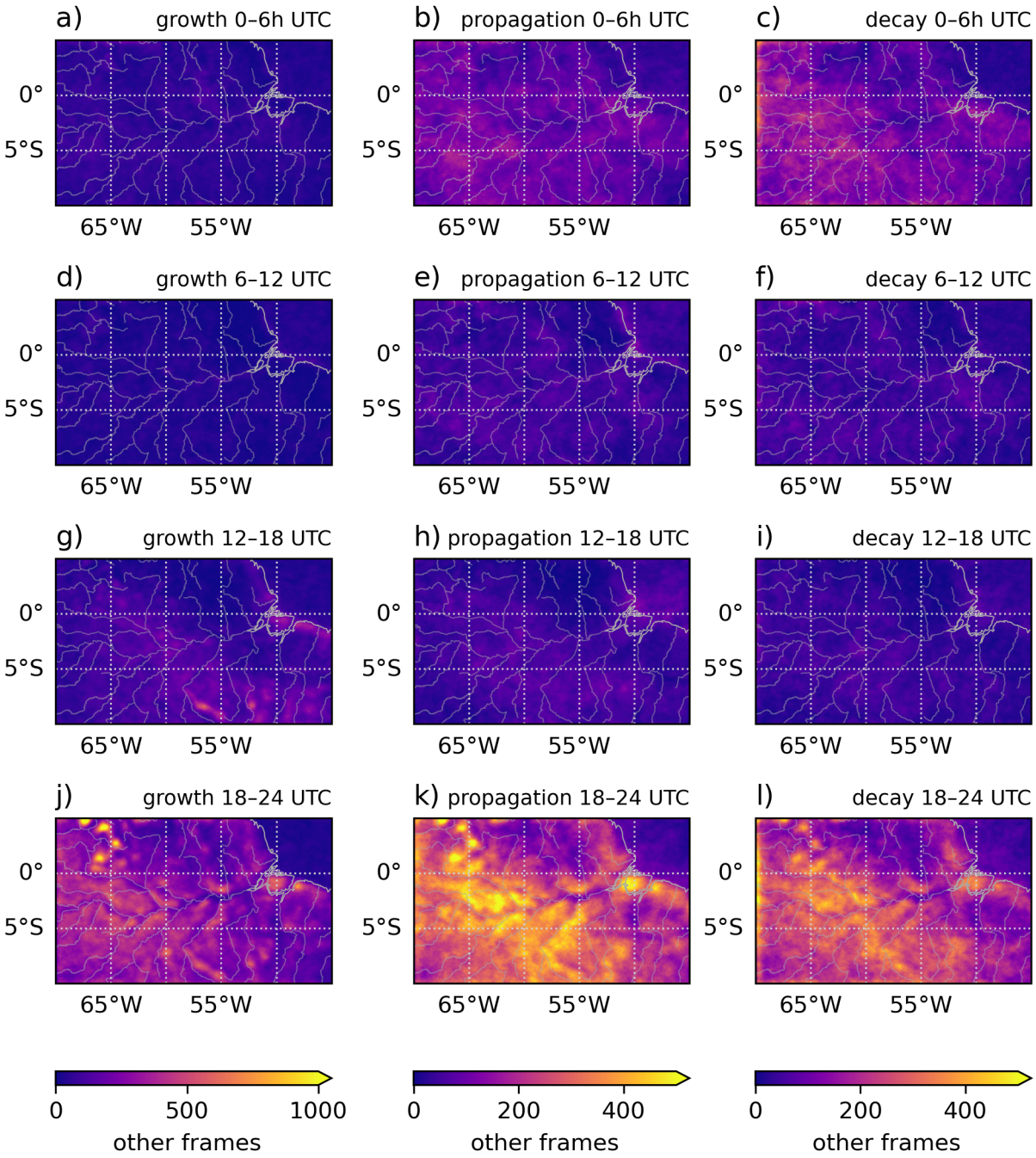


Figure B.9: Diurnal spacial distribution of the lifecycle of other systems.

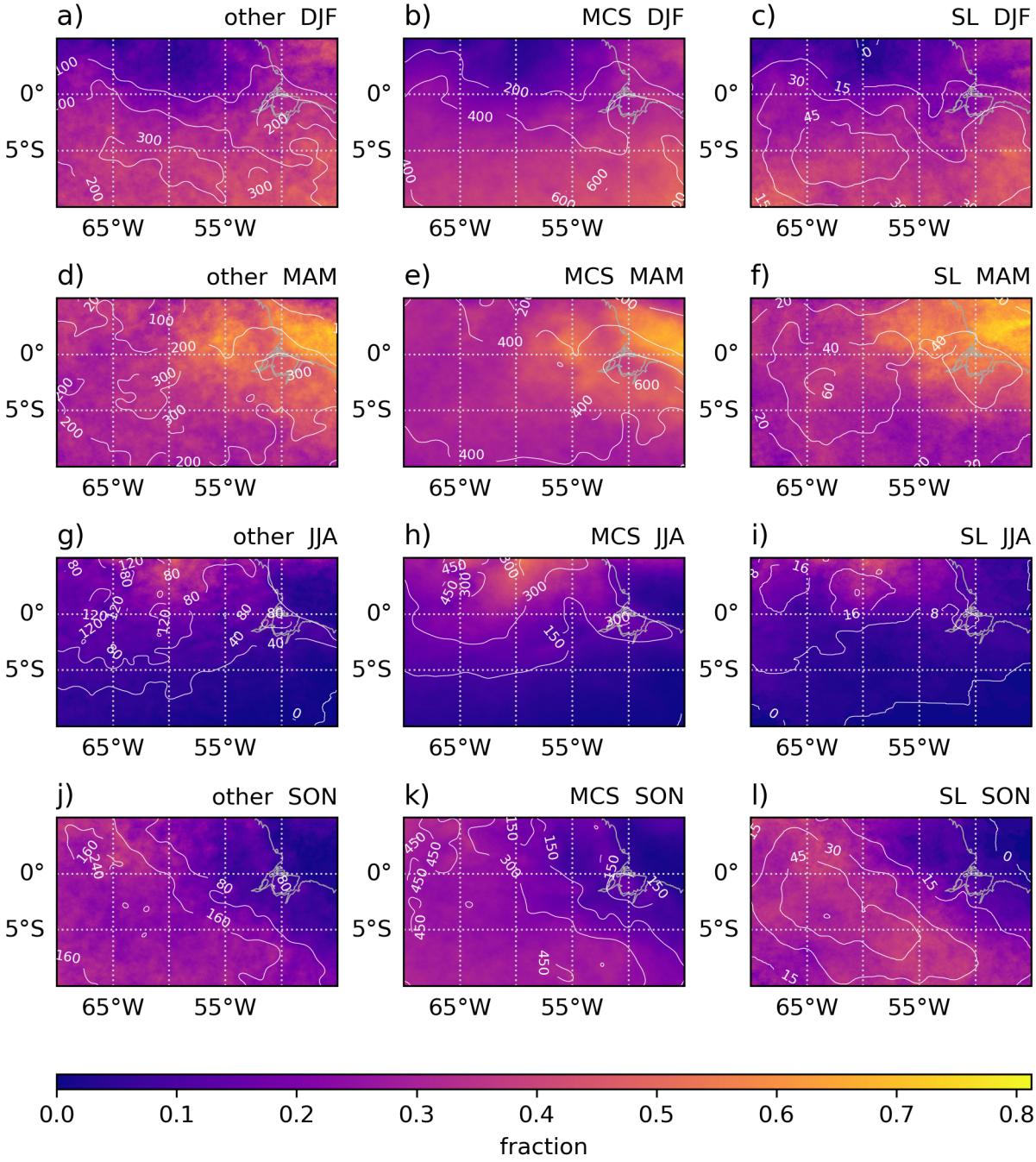


Figure B.10: Relative seasonal spacial distribution of convective system occurrences. Same as Figure 5.7, but relative to the counts in Figure 5.6, with contour lines indicating the absolute values.

## Appendix C

# References


Alcântara, Clênia R., Maria A.F. Silva Dias, Enio P. Souza and Julia C.P. Cohen (2011). 'Verification of the role of the low level jets in Amazon squall lines'. In: *Atmospheric Research* 100 (1), pp. 36–44. ISSN: 01698095. DOI: 10.1016/j.atmosres.2010.12.023.

Anselmo, Evandro M., Luiz A.T. Machado, Courtney Schumacher and George N. Kiladis (2021). 'Amazonian mesoscale convective systems: Life cycle and propagation characteristics'. In: *International Journal of Climatology* 41 (7), pp. 3968–3981. ISSN: 10970088. DOI: 10.1002/joc.7053.

Anselmo, Evandro M., Courtney Schumacher and Luiz A.T. Machado (2020). 'The amazonian low-level jet and its connection to convective cloud propagation and evolution'. In: *Monthly Weather Review* 148 (10), pp. 4083–4099. ISSN: 15200493. DOI: 10.1175/MWR-D-19-0414.1.

Barber, C. Bradford, David P. Dobkin and Hannu Huhdanpaa (1996). 'The quickhull algorithm for convex hulls'. In: *ACM Trans. Math. Softw.* 22.4, pp. 469–483. ISSN: 0098-3500. DOI: 10.1145/235815.235821.

Bluestein, Howard B. and Michael H. Jain (1985). 'Formation of Mesoscale Lines of Precipitation: Severe Squall Lines in Oklahoma during the Spring'. In: *Journal of Atmospheric Sciences* 42.16, pp. 1711–1732. DOI: 10.1175/1520-0469(1985)042<1711:FOMLOP>2.0.CO;2.

Bryan, George H, Richard Rotunno and Morris L Weisman (2012). 'What is RKW theory?' In: *26th Conference on Severe Local Storms*. American Meteorological Society, 4B.6. URL: https://ams.confex.com/ams/26SLS/webprogram/Paper211731.html.

Cohen, Júlia Clarinda Paiva (1989). 'Um estudo observacional de linhas de instabilidade na Amazônia'. PhD thesis. Instituto Nacional de Pesquisas Espaciais.

Cohen, Júlia Clarinda Paiva, Maria A. F. Silva Dias and Carlos A. Nobre (1995). 'Environmental Conditions Associated with Amazonian Squall Lines: A Case Study'. In: *Monthly Weather Review* 123 (11), pp. 3163–3174. DOI: 10.1175/1520-0493(1995)123<3163:ECAWAS>2.0.CO;2.

Desbois, Michel, Théoneste Kayiranga, Brigitte Gnamien, Saad Guessous and Laurence Picon (1988). 'Characterization of Some Elements of the Sahelian Climate and Their Interannual Variations for July 1983, 1984 and 1985 from the Analysis of METEOSAT ISCCP Data'. In: *Journal of Climate* 1 (9), pp. 867–904.

Ester, Martin, Hans-Peter Kriegel, Jörg Sander and Xiaowei Xu (1996). 'A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise'. In: *Second International Conference on Knowledge Discovery and Data Mining. KDD-96*. AAAI (www.aaai.org), pp. 226–231. URL: https://aaai.org/papers/kdd96-037.

Feng, Zhe, Andreas F. Prein, Julia Kukulies, Thomas Fiolleau, William K. Jones, Ben Maybee, Zachary L. Moon, Kelly M. Núñez Ocasio, Wenhao Dong, Maria J. Molina, Mary Grace Albright, Manikandan Rajagopal, Vanessa Robledo, Jinyan Song, Fengfei Song, L. Ruby Leung, Adam C. Varble, Cornelia Klein, Remy Roca, Ran Feng and John F. Mejia (2025). 'Mesoscale Convective Systems Tracking Method Intercomparison (MCSMIP): Application to DYAMOND Global km-Scale Simulations'. In: *Journal of Geophysical Research: Atmospheres* 130 (8). ISSN: 21698996. DOI: 10.1029/2024JD042204.

Fiolleau, Thomas and Rémy Roca (2013). 'An algorithm for the detection and tracking of tropical mesoscale convective systems using infrared images from geostationary satellite'. In: *IEEE Transactions on Geoscience and Remote Sensing* 51 (7), pp. 4302–4315. ISSN: 01962892. DOI: 10.1109/TGRS.2012.2227762.

— (2024). 'A database of deep convective systems derived from the intercalibrated meteorological geostationary satellite fleet and the TOOCAN algorithm (2012–2020)'. In: *Earth System Science Data* 16 (9), pp. 4021–4050. ISSN: 1866-3516. DOI: 10.5194/essd-16-4021-2024. URL: https://essd.copernicus.org/articles/16/4021/2024/.

Fovell, Robert G. and Yoshi Ogura (1989). 'Effect of Vertical Wind Shear on Numerically Simulated Multicell Storm Structure'. In: *Journal of the Atmospheric Sciences* 46 (20), pp. 3144–3176. DOI: 10.1175/1520-0469(1989)046<3144:EOVWSO>2.0.CO;2.

Fovell, Robert G. and Pei-Hua Tan (1998). 'The Temporal Behavior of Numerically Simulated Multicell-Type Storms. Part II: The Convective Cell Life Cycle and Cell Regeneration'. In: *Monthly Weather Review* 126 (3), pp. 551–577. DOI: 10.1175/1520-0493(1998)126<0551:TTBONS>2.0.CO;2.

GOES-R (2018). *GOES-16 Drift and Transition to Operations*. NOAA & NASA. URL: https://goes-r.gov/users/transitionToOperations16.html.

— (2025). *GOES-19 Post-Launch Testing and Transition to Operations*. NOAA & NASA. URL: https://goes-r.gov/users/transitionToOperations19.html.

GOES-R and NOAA: GOES-R Algorithm Working Group; NOAA Geostationary Operation Environmental Satellite-R Series (2017). *NOAA Geostationary Operation Environmental Satellite-R Series (GOES-R Series) Advanced Baseline Imager (ABI) Level 2 Cloud and Moisture Imagery Products (CMIP). GOES-16 Full Disk channel 13 brightness temperature at the Top-Of-Atmosphere*. NOAA National Centers for Environmental Information. https://doi.org/10.7289/V5736P36. Accessed 25.07.2025.

GOES-R: Geostationary Operational Environmental Satellites-R Series Program Office (2020). *GOES-R Series Concept of Operations (CONOPS), Version 3.0*. URL: https://goes-r.gov/syseng/docs/CONOPS.pdf.

— (2021). *GOES-R Series Product Definition and Users' Guide (PUG). Volume 3: Level 1B Products Revision 2.3*.

Greco, Steven, Robert Swap, Michael Garstang, Stanley Ulanski, Mark Shipham, R. C. Harriss, R. Talbot, M. O. Andreae and P. Artaxo (1990). 'Rainfall and surface kinematic conditions over central Amazonia during ABLE 2B'. In: *Journal of Geophysical Research: Atmospheres* 95.D10, pp. 17001–17014. DOI: 10.1029/JD095iD10p17001.

Gupta, S., D. Wang, S. E. Giangrande, T. S. Biscaro and M. P. Jensen (2024). 'Lifecycle of updrafts and mass flux in isolated deep convection over the Amazon rainforest: insights from cell tracking'. In: *Atmospheric Chemistry and Physics* 24.7, pp. 4487–4510. DOI: 10.5194/acp-24-4487-2024. URL: https://acp.copernicus.org/articles/24/4487/2024/.

Heikenfeld, Max, Peter J. Marinescu, Matthew Christensen, Duncan Watson-Parris, Fabian Senf, Susan C. Van Den Heever and Philip Stier (2019). 'Tobac 1.2: Towards a flexible framework for tracking and analysis of clouds in diverse datasets'. In: *Geoscientific Model Development* 12 (11), pp. 4551–4570. ISSN: 19919603. DOI: 10.5194/gmd-12-4551-2019.

Kamani, Mohammad Mahdi, Farshid Farhat, Stephen Wistar and James Z. Wang (2018). 'Skeleton matching with applications in severe weather detection'. In: *Applied Soft Computing Journal* 70, pp. 1154–1166. ISSN: 15684946. DOI: 10.1016/j.asoc.2017.05.037.

Kelso, Nathaniel Vaughn and Tom Patterson (2026). *1:10m Cross-blended Hypsometric Tints. With Shaded Relief Only*. Natural Earth Data. https://naturalearthdata.com. Accessed 15.01.2026.

Kousky, Vernon E. (1980). 'Diurnal Rainfall Variation in Northeast Brazil'. In: *Monthly Weather Review* 108 (4), pp. 488–498. DOI: 10.1175/1520-0493(1980)108<0488:DRVINB>2.0.CO;2.

Leal, Helvecio B., Alan J.P. Calheiros, Henrique M.J. Barbosa, Adriano P. Almeida, Arturo Sanchez, Daniel A. Vila, Sâmia R. Garcia and Elbert E.N. Macau (2022). 'Impact of Multi-Thresholds and Vector Correction for Tracking Precipitating Systems over the Amazon Basin'. In: *Remote Sensing* 14 (21). ISSN: 20724292. DOI: 10.3390/rs14215408.

Machado, Luiz A. T., Alan J.P. Calheiros, Thiago Biscaro, Scott Giangrande, Maria A.F.Silva Dias, Micael A. Cecchini, Rachel Albrecht, Meinrat O. Andreae, Wagner F. Araujo, Paulo Artaxo, Stephan Borrmann, Ramon Braga, Casey Burleyson, Cristiano W. Eichholz, Jiwen Fan, Zhe Feng, Gilberto F. Fisch, Michael P. Jensen, Scot T. Martin, Ulrich Pöschl, Christopher Pöhlker, Mira L. Pöhlker, Jean François Ribaud, Daniel Rosenfeld, Jaci M.B. Saraiva, Courtney Schumacher, Ryan Thalman, David Walter and Manfred Wendisch (2018). 'Overview: Precipitation characteristics and sensitivities to environmental conditions during GoAmazon2014/5 and ACRIDICON-CHUVA'. In: *Atmospheric Chemistry and Physics* 18 (9), pp. 6461–6482. ISSN: 16807324. DOI: 10.5194/acp-18-6461-2018.

Machado, Luiz A. T., Henri Laurent and Alexandra A. Lima (2002). 'Diurnal march of the convection observed during TRMM-WETAMC/LBA'. In: *Journal of Geophysical Research: Atmospheres* 107 (20), LBA 31-1-LBA 31–15. ISSN: 01480227. DOI: 10.1029/2001JD000338.

Machado, Luiz A. T., W. B. Rossow, R. L. Guedes and A. W. Walker (1998). 'Life Cycle Variations of Mesoscale Convective Systems over the Americas'. In: *Monthly Weather Review* 126, pp. 1630–1654.

Machado, Luiz A.T., Henri Laurent, Nadine Dessay and Ieso Miranda (2004). 'Seasonal and diurnal variability of convection over the Amazonia: A comparison of different vegetation types and large scale forcing'. In: *Theoretical and Applied Climatology* 78 (1-3), pp. 61–77. ISSN: 0177798X. DOI: 10.1007/s00704-004-0044-9.

Markowski, Paul and Yvette Richardson (2010). 'Mesoscale Convective Systems'. In: *Mesoscale Meteorology in Midlatitudes*. John Wiley & Sons, Ltd. Chap. 9, pp. 245–272. ISBN: 9780470682104. DOI: 10.1002/9780470682104.ch9.

Mathon, Vincent and Henri Laurent (2001). 'Life cycle of Sahelian mesoscale convective cloud systems'. In: *Quarterly Journal of the Royal Meteorological Society* 127 (572), pp. 377–406. ISSN: 0035-9009. DOI: 10.1002/qj.49712757208.

Negrón-Juárez, Robinson I., Jeffrey Q. Chambers, Giuliano Guimaraes, Hongcheng Zeng, Carlos F.M. Raupp, Daniel M. Marra, Gabriel H.P.M. Ribeiro, Sassan S. Saatchi, Bruce W. Nelson and Niro Higuchi (2010). 'Widespread Amazon forest tree mortality from a single cross-basin squall line event'. In: *Geophysical Research Letters* 37 (16). ISSN: 00948276. DOI: 10.1029/2010GL043733.

Oliveira, Fernando P. and Marcos D. Oyama (2019). 'Squall-line initiation over the northern coast of Brazil in March: Observational features'. In: *Meteorological Applications* 27 (1). ISSN: 14698080. DOI: 10.1002/met.1799.

Parker, Matthew D. and Richard H. Johnson (2000). 'Organizational Modes of Midlatitude Mesoscale Convective Systems'. In: *Monthly Weather Review* 128.10, pp. 3413–3436. DOI: 10.1175/1520-0493(2001)129<3413:OMOMMC>2.0.CO;2.

— (2004). 'Simulated Convective Lines with Leading Precipitation. Part I: Governing Dynamics'. In: *Journal of Atmospheric Sciences* 61 (14), pp. 1637–1655. DOI: 10.1175/1520-0469(2004)061<1637:SCLWLP>2.0.CO;2.

Raut, Bhupendra A., Robert Jackson, Mark Picel, Scott M. Collis, Martin Bergemann and Christian Jakob (2021). 'An adaptive tracking algorithm for convection in simulated and remote sensing data'. In: *Journal of Applied Meteorology and Climatology* 60 (4), pp. 513–526. ISSN: 15588432. DOI: 10.1175/JAMC-D-20-0119.1.

Rickenbach, Thomas M (2004). 'Nocturnal Cloud Systems and the Diurnal Variation of Clouds and Rainfall in Southwestern Amazonia'. In: *Monthly Weather Review* 132 (5), pp. 1201–1219. DOI: 10.1175/1520-0493(1980)108<0488:DRVINB>2.0.CO;2.

Ropelewski, C. F. and P. D. Jones (1987). 'An Extension of the Tahiti–Darwin Southern Oscillation Index'. In: *Monthly Weather Review* 115.9, pp. 2161–2165. DOI: 10.1175/1520-0493(1987)115<2161:AEOTTS>2.0.CO;2.

Rotunno, Richard, Joseph B. Klemp and Morris L. Weisman (1988). 'A Theory for Strong, Long-Lived Squall Lines'. In: *Journal of the Atmospheric Sciences* 45 (3), pp. 463–485. DOI: 10.1175/1520-0469(1988)045<0463:ATFSLL>2.0.CO;2.

Schubert, Erich, Jörg Sander, Martin Ester, Hans Peter Kriegel and Xiaowei Xu (2017). 'DBSCAN revisited, revisited: Why and how you should (still) use DBSCAN'. In: *ACM Transactions on Database Systems* 42 (3). ISSN: 15574644. DOI: 10.1145/3068335.

Shi, Zhao, Yuxiang Wen and Jianxin He (2024). 'A clustering-based method for identifying and tracking squall lines'. In: *Atmospheric Measurement Techniques* 17 (13), pp. 4121–4135. ISSN: 18678548. DOI: 10.5194/amt-17-4121-2024.

Sokolowsky, G. Alexander, Sean W. Freeman, William K. Jones, Julia Kukulies, Fabian Senf, Peter J. Marinescu, Max Heikenfeld, Kelcy N. Brunner, Eric C. Bruning, Scott M. Collis, Robert C. Jackson, Gabrielle R. Leung, Nils Pfeifer, Bhupendra A. Raut, Stephen M. Saleeby, Philip Stier and Susan C. Van Den Heever (2024). 'tobac v1.5: introducing fast 3D tracking, splits and mergers, and other enhancements for identifying and analysing meteorological phenomena'. In: *Geoscientific Model Development* 17 (13), pp. 5309–5330. ISSN: 19919603. DOI: 10.5194/gmd-17-5309-2024.

Sousa, Aline Corrêa de, L. A. Candido and P. Satyamurty (2021). 'Convective Cloud Clusters and Squall Lines along the Coastal Amazon'. In: *Monthly Weather Review* 149 (11), pp. 3589–3608. ISSN: 15200493. DOI: 10.1175/MWR-D-21-0045.1.

Vieira Fischer, Frederik L. (2025). 'Climatology of Amazonian Squall Lines derived by Algorithmic Detection and Tracking of Coherent Features'. Project Report related to the Master Thesis.

Vila, Daniel Alejandro, Luiz Augusto Toledo Machado, Henri Laurent and Inés Velasco (2008). 'Forecast and tracking the evolution of cloud clusters (ForTraCC) using satellite infrared imagery: Methodology and validation'. In: *Weather and Forecasting* 23 (2), pp. 233–245. ISSN: 08828156. DOI: 10.1175/2007WAF2006121.1.

Virtanen, Pauli, Ralf Gommers, Travis E. Oliphant, Matt Haberland, Tyler Reddy, David Cournapeau, Evgeni Burovski, Pearu Peterson, Warren Weckesser, Jonathan Bright, Stéfan J. van der Walt, Matthew Brett, Joshua Wilson, K. Jarrod Millman, Nikolay Mayorov, Andrew R. J. Nelson, Eric Jones, Robert Kern, Eric Larson, C J Carey, İlhan Polat, Yu Feng, Eric W. Moore, Jake VanderPlas, Denis Laxalde, Josef Perktold, Robert Cimrman, Ian Henriksen, E. A. Quintero, Charles R. Harris, Anne M. Archibald, Antônio H. Ribeiro, Fabian Pedregosa, Paul van Mulbregt and SciPy 1.0 Contributors (2020). 'SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python'. In: *Nature Methods* 17, pp. 261–272. DOI: 10.1038/s41592-019-0686-2.

Wang, Xing, Hao-xuan Bian, Dai-li Qian, Chun-sheng Miao and Shao-wei Zhan (2021). 'An automatic identifying method of the squall line based on Hough transform'. In: *Multimedia*

*Tools and Applications* 80 (12), pp. 18993–19009. ISSN: 15737721. DOI: 10.1007/s11042-021-10689-3.

Weisman, Morris L. and Richard Rotunno (2004). '"A Theory for Strong Long-Lived Squall Lines" Revisited'. In: *Journal of the Atmospheric Sciences* 61 (4), pp. 361–382. DOI: 10.1175/1520-0469(2004)061<0361:ATFSLS>2.0.CO;2.

Xie, Peilong, Zhiqun Hu, Shujie Yuan, Jiafeng Zheng, Hanyuan Tian and Fen Xu (2023). 'RADAR Echo Recognition of Squall Line Based on Deep Learning'. In: *Remote Sensing* 15 (19). ISSN: 20724292. DOI: 10.3390/rs15194726.