

Deep convection lifecycle characteristics: a database from GoAmazon experiment

Camila da Cunha Lopes¹, Rachel Ifanger Albrecht¹, Douglas Messias Uba², Thiago Souza Biscaro², and Ivan Saraiva³

¹Departamento de Ciências Atmosféricas, Instituto de Astronomia, Geofísica e Ciências Atmosféricas, Universidade de São Paulo, Rua do Matão, 1226, 05508-090, São Paulo, SP, Brasil

²Divisão de Satélites e Sensores Meteorológicos, Coordenação-Geral de Ciências da Terra, Instituto Nacional de Pesquisas Espaciais, Rodovia Presidente Dutra, Km 40, SP-RJ, 12630-000, Cachoeira Paulista, SP, Brasil

³Centro Gestor e Operacional do Sistema de Proteção da Amazônia, Av. do Turismo, 1350, 69049-630, Manaus, AM, Brasil

Correspondence: Camila da Cunha Lopes (camilalopes.ccl@gmail.com)

Abstract. The Observations and Modeling of the Green Ocean Amazon (GoAmazon2014/5) Experiment provided a comprehensive suite of cloud-aerosol-precipitation observations with both *in situ* and remote sensing instruments. In this study, we apply a tracking methodology to volumetric radar data, creating a refined database focused on deep convective systems with full lifecycle, incorporating lightning data. This refined deep convection database is shown to be a robust sample of the complete dataset in terms of convective systems morphology. The analysis reveals significant seasonal and diurnal variations in convective morphology and intensity, with most intense systems occurring during the dry-to-wet season transition. The filtered dataset offers a robust sample for future studies on Amazonian convection.

1 Introduction

The Amazon tropical rainforest serves as a natural test bed for several studies on cloud-aerosol-precipitation and land-atmosphere interactions due to its large territorial extent which includes pristine forest, agricultural expansion and a large urban zone with an industrial center. This complex ecosystem is one of the main centers of convection regulating the climate (Nobre et al., 2009; Artaxo et al., 2022) and the South American Monsoon System (SAMS) (Zhou and Lau, 1998; Jones and Carvalho, 2002). Several convection patterns are present in the region, mainly modulated by the Hadley circulation and the corresponding position of the Intertropical Convergence Zone (ITCZ) which determines the wet (austral summer) and dry (austral winter) seasons.

Several field experiments were conducted in the region in order to study different aspects of the cloud-aerosol-precipitation interactions: the Amazon Boundary Layer Experiment ABLE 2A (Harriss et al., 1988) and 2B (Harriss et al., 1990)) focused on the che

[1] (0,76) : The sentence syntax in this section is awkward and not correct. It is a mix of a list with compound sentences and separate sentences. I suggest to get rid of the colon and semi-colons and make separate and shorter sentences.

Atmosphere Experiment in Amazonia LBA Program (Silva Dias et al., 2002) were responsible for numerous field experiments during late 1990s and early 2000s, including the first major mesoscale atmospheric campaign as part of the Tropical Rainfall Measuring Mission (TRMM) validation campaigns and the CHUVA Project (Cloud Processes of the Main Precipitation Systems in Brazil: A Contribution to Cloud-Resolving Modeling and to the Global Precipitation Measurement (GPM) (Machado

Deep convection lifecycle characteristics: a database from GoAmazon experiment

Camila da Cunha Lopes¹, Rachel Ifanger Albrecht¹, Douglas Messias Uba², Thiago Souza Biscaro², and Ivan Saraiva³

¹Departamento de Ciências Atmosféricas, Instituto de Astronomia, Geofísica e Ciências Atmosféricas, Universidade de São Paulo, Rua do Matão, 1226, 05508-090, São Paulo, SP, Brasil

²Divisão de Satélites e Sensores Meteorológicos, Coordenação-Geral de Ciências da Terra, Instituto Nacional de Pesquisas Espaciais, Rodovia Presidente Dutra, Km 40, SP-RJ, 12630-000, Cachoeira Paulista, SP, Brasil

³Centro Gestor e Operacional do Sistema de Proteção da Amazônia, Av. do Turismo, 1350, 69049-630, Manaus, AM, Brasil

Correspondence: Camila da Cunha Lopes (camilalopes.ccl@gmail.com)

Abstract. The Observations and Modeling of the Green Ocean Amazon (GoAmazon2014/5) Experiment provided a comprehensive suite of cloud-aerosol-precipitation observations with both *in situ* and remote sensing instruments. In this study, we apply a tracking methodology to volumetric radar data, creating a refined database focused on deep convective systems with full lifecycle, incorporating lightning data. This refined deep convection database is shown to be a robust sample of the complete dataset in terms of convective systems morphology. The analysis reveals significant seasonal and diurnal variations in convective morphology and intensity, with most intense systems occurring during the dry-to-wet season transition. The filtered dataset offers a robust sample for future studies on Amazonian convection.

1 Introduction

The Amazon tropical rainforest serves as a natural test bed for several studies on cloud-aerosol-precipitation and land-atmosphere interactions due to its large territorial extent which includes pristine forest, agricultural expansion and a large urban zone with an industrial center. This complex ecosystem is one of the main centers of convection regulating the climate (Nobre et al., 2009; Artaxo et al., 2022) and the South American Monsoon System (SAMS) (Zhou and Lau, 1998; Jones and Carvalho, 2002). Several convection patterns are present in the region, mainly modulated by the Hadley circulation and the corresponding position of the Intertropical Convergence Zone (ITCZ) which determines the wet (austral summer) and dry (austral winter) seasons.

Several field experiments were conducted in the region in order to study different aspects of the cloud-aerosol-precipitation interactions: the Amazon Boundary Layer Experiment ABLE 2A (Harriss et al., 1988) and 2B (Harriss et al., 1990)) focused on the che

[1] (0,76) : The sentence syntax in this section is awkward and not correct. It is a mix of a list with compound sentences and separate sentences. I suggest to get rid of the colon and semi-colons and make separate and shorter sentences.

Atmosphere Experiment in Amazonia LBA Program (Silva Dias et al., 2002) were responsible for numerous field experiments during late 1990s and early 2000s, including the first major mesoscale atmospheric campaign as part of the Tropical Rainfall Measuring Mission (TRMM) validation campaigns and the CHUVA Project (Cloud Processes of the Main Precipitation Systems in Brazil: A Contribution to Cloud-Resolving Modeling and to the Global Precipitation Measurement (GPM) (Machado

et al., 2014a). The Green Ocean Amazon Experiment (GoAmazon 2014/5) (Martin et al., 2017) was the first long-term experiment to analyze the effects of the Manaus pollution plume in different experimental sites around Manaus, and included two intensive operation periods (IOPs) in the wet and dry seasons. Unlike previous experiments, an operational weather radar was available during GoAmazon, operated by *Sistema de Proteção da Amazônia* (SIPAM, System for the Protection of the Amazon), which covers all experimental sites with cloud remote sensing data in great temporal and spatial resolutions.

A few studies provide insights about convection characteristics during GoAmazon, such as Giangrande et al. (2017), Machado et al. (2018), Giangrande et al. (2020), and Biscaro et al. (2021). Each of them uses different definitions of cloud features and different definitions of wet and dry seasons, which makes it difficult for other studies to follow their lead in measurements, specially the ones that analyze cloud-aerosol-precipitation interactions. For this reason, this study aims to create a comprehensive database of convective systems based on radar data that can be used in future studies regarding GoAmazon data.

[2] (1,22) : Do you mean that different remote sensing data are used and hence result in different definitions or use the same data with different definitions. Not clear.

[3] (1,29) : Following the previous comment, are you adding to the confusion by yet another different data set and definition?

2 Materials and Methods

2.1 Data

Data sources for this study are the field experiments GoAmazon (Martin et al., 2017) and CHUVA-Manaus (Machado et al., 2014b) that occurred between January 2014 and December 2015 around Manaus, Amazonas. The main goal of the GoAmazon experiment was to analyze cloud-aerosol-precipitation interactions between the forest and the atmosphere, specially the transformation of air plumes from the pristine forest to the Manaus pollution plumes and its eastern propagation. The goal of the CHUVA-Manaus experiment was to characterize the convection regimes in the region by remote sensing, with the installation of a X-Band radar during the experiment and partnerships with CENSIPAM (*Centro Gestor e Operacional do Sistema de Proteção da Amazônia*, Manager and Operational Center of the System for the Protection of the Amazon) for surface radar data and NASA-JAXA (National Aeronautics and Space Administration — Japan Aerospace Exploration Agency) for satellite radar data from TRMM (Tropical Rainfall Measuring Mission) and GPM (Global Precipitation Measurement Mission).

[4] (1,46) : It seems that this sentence should be next to the first sentence on this page.

The main site of these experiments, named T3, was located in Manacapuru, Amazonas, Brazil (3.213°S, 60.598°W), about 70 km west from Manaus. A wide range of cloud, precipitation, aerosols and atmospheric instruments were installed at the site, such as part of the ARM (Atmospheric Radiation Measurement) mobile facility AMF1 that took measurements during the most intensive period. An additional instrument and site (shown in Fig. 1 of Martin et al. (2017)) took measurements during two intensive operation periods (IOPs), with IOP1 being between February 1st and March 31st 2014 (wet season) and IOP2 between August 15th and October 15th 2014 (dry season). More details about the campaigns can be found in Martin et al. (2016, 2017).

[5] (1,71) : "such" is not needed

In order to create the convective systems database, radar volumes from SIPAM (*Sistema de Proteção da Amazônia*, System for the Protection of the Amazon) single polarization S-band radar located in Manaus (3.149°S, 59.991°W, 102.4 m altitude, Fig. 1) were selected in the period of GoAmazon experiment (2014-01-01 to 2015-12-31). These volumes consist of CAPPIs

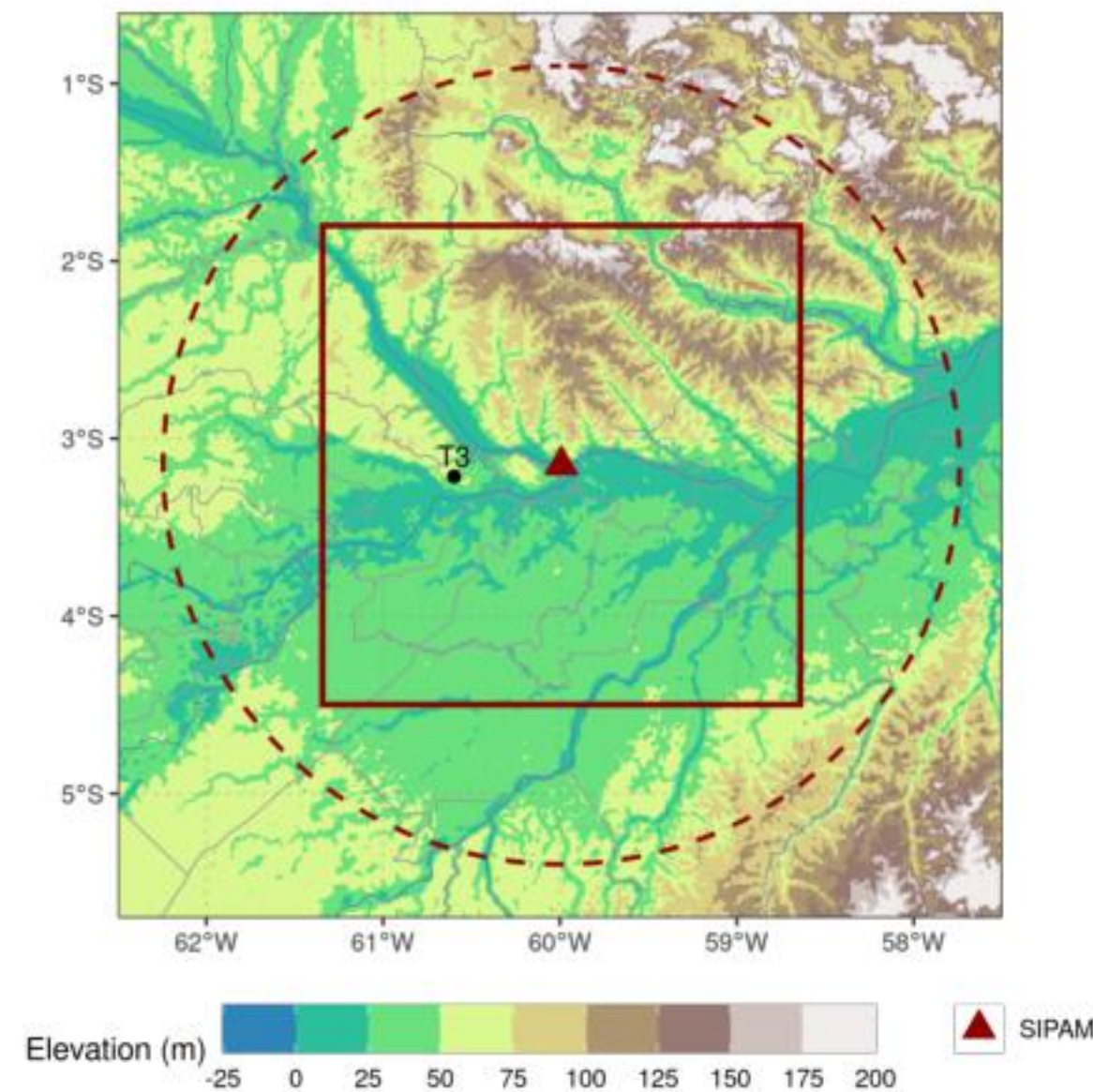


Figure 1. Amazonian region used in this study, with SIPAM radar location (red triangle), 250-km originally calculated coverage (dashed red lines), 150 km x 150 km bounding box (red square) used in the study and T3 site location in Manacapuru where surface data were collected (not shown in this study)

(Constant Altitude Plan Position Indicators) with a bias correction calculated by Schumacher and Funk (2018). Table 1 shows the data settings, including the bias values applied throughout the period.

This radar has a beamwidth of 1.8° and performs volumetric scans approximately every 12 minutes, covering a domain of 240 km from the radar location with a gate resolution of 500 m, azimuth resolution of 1° , and 17 elevation angles ($0.9\text{--}19.5^\circ$). Given the coarse characteristics of this radar and its scan strategy, such as varying numbers of elevation sweeps and antenna vibrations between volumes, the radar data used in this study consist of quality-controlled, bias-corrected CAPPIs, gridded onto a 1×1 km horizontal resolution. Following the approach of Saraiva et al. (2016), who limited the domain to 150 km radius to ensure better beam geometry and vertical resolution, we also restricted our analysis to a 150×150 km box centered on the radar location for consistent and reliable detection of convective structures while minimizing the effects of range-dependent radar beam geometry and incomplete vertical sampling at farther ranges.

A second data source was employed to calculate parameters related to lightning activity in the convective systems. The Vaisala GLD360 lightning network (Demetriades et al., 2010) measures cloud-to-ground return strokes in the VLF (Very Low Frequency) range using two location techniques, magnetic direction finding (MDF) and time-of-arrival (TOA), as well as lightning recognition algorithm. The lightning recognition algorithm in Brazil, which is 70% in average (Naccarato et al., 2010) but can be significantly lower in areas with deficient coverage such as Northern Brazil, including the Amazon region, as well as the definition of a measured stroke as a single or multiple real return strokes (Murphy and Nag,

[9] (2,89) : I suggest to shorten your sentences. Too many ideas expressed leading to awkward sentences

[10] (2,92) : A general comment throughout the paper.

Table 1. SIPAM radar data settings processed by CPTEC-INPE

Type	CAPPIs
Format	Binary, 15 x 500 x 500 elements
Resolution (vertical, horizontal, temporal)	1 x 1 x 1 km, 12 min
Min, max height	2 km, 16 km
CAPPI processing software	RSL (Radar Software Library)
Bias correction (Schumacher and Funk, 2018)	2014-01-03 to 2014-02-05: + 1.0 dB 2014-02-06 to 2014-08-19: + 3.0 dB 2014-08-20 to 2014-10-16: - 2.5 dB 2014-10-17 to 2015-03-06: - 5.5 dB 2015-03-07 to 2015-07-05: - 4.0 dB 2015-07-06 to 2015-10-28: - 1.5 dB 2015-10-29 to 2015-12-31: + 1.0 dB

2015). Strokes data were accumulated over 12 minutes in the same timestamps as the radar data and selected within the clusters polygons delimited by TATHU.

[11] (3,47) : First mention of TATHU. Need acronym spelled out and described. Then you say all this in next paragraph. Need to clean up.

2.2 Tracking methodology

The convective systems database was created with the TATHU (Tracking and Analysis of Thunderstorms) software package (Uba et al., 2022) applied to the radar data described in the previous section. The software is a free, open-source python pack-
age available at <https://github.com/uba/tathu>, which addresses convective system tracking as a multi-target tracking problem (Makris and Prieur, 2014). The main modules are observation, detection, description, tracking, and forecast (not used in this study). The algorithm detects agglomerates of pixels – called herein **clusters** – in an input field at a single time step according to the (one or more) threshold(s) and extracts its statistics such as size (in pixels), mean and maximum values of the input
field, among others. From subsequent time steps, based on spatial overlap, it tracks and names (via a universal unique identifier — *uuid*) the **convective systems** that occurred in the period and its status during the described life cycle, with status being “spontaneous generation” (new cluster), “continuity” (growing or decaying cluster), “split” (when a single cluster separates into two or more clusters after a time step) or “merge” (when two or more clusters merge into a single cluster after a time step). The identification of a split or a merge depends mainly on the clusters propagation between subsequent time steps, where
discrepancies can occur in situations such as clusters generating within the same relative overlap area of other propagating clusters.

[12] (3,85) : So agglomerate and cluster is the same. Convective system is a set of clusters with the same initial cluster. So a cluster is a "contiguous area" of radar pixels that satisfy a certain criteria. Usually it is a threshold of say 35 or 45 dBZ. What is used here? Why not just call it a storm?

[13] (3,87) : If agglomeration and clusters are the same, why introduce the two terms, keep is simple.

Table 2 shows the TATHU settings used in this study. We selected SIPAM radar reflectivity greater than 20 dBZ at the [15] (4,6) : OKThe problem with the terms that you use "agglomeration", "cluster" etc is that these are adjectives and not nouns. That is, "clusters of what", the [16] (4,9) : or precipitation cells or something like that. this altitude [14] (4,8) : OK...now I understand. With 20 dBZ threshold, you can have several thunderstorm cells. Explain this and I and your readers will understand what you mean. Please fix above.

maximum reflectivity is frequently observed in convective precipitation cores below freezing layer. Second, lower elevations near the radar are significantly affected by ground clutter and beam blockage (Giangrande et al., 2016; Schumacher and Funk, 2018), which compromises the reliability of reflectivity data closer to the surface. Additionally, the SIPAM radar's relatively coarse resolution limits the ability to resolve vertical structures accurately, particularly storm tilts. However, in the tropical environment of the Amazon, vertical wind shear is generally weak, reducing the likelihood of significant storm tilt or vertical displacement of convective cores. These factors together support the use of a horizontal 3-km CAPPI as a consistent and suitable input for identifying and tracking convective systems in this study. Other tracking algorithms consider composite reflectivity (e.g., Heikenfeld et al. (2019); Sokolowsky et al. (2024)) instead of 2D reflectivity fields to offer a more comprehensive depiction of the 3D structure of convective systems. However, considering the context of our study and the nature of this dataset just exposed above, the 3D structure of the convective systems should be captured. A comprehensive study of the implications for the reflectivity thresholds for tracking convective systems in CAPPI fields can be found in Leal et al. (2022), and it is not the scope of this study. Moreover, our approach is consistent methodologically with previous studies that have successfully employed single-level CAPPI fields to characterize convective systems in the Amazon (e.g., Laurent and Laurent (2002); Albrecht et al. (2011); Leal et al. (2022); Gupta et al. (2024)). These studies provided valuable insights into storm morphology, evolution, and vertical structure using similar techniques. Adopting a comparable methodology ensures continuity and comparability with the existing body of literature on Amazonian convection. [17] (4,40) : missing word ... "not within the scope"

The relative overlap area strategy considers two subsequent clusters in time as the same convective system when there's at least a 10% overlap between their areas (i.e., polygons). The maximum interval between images considers a data gap sufficient to ensure continuity of the convective systems, but can result in different convective systems being tracked as the same if they are in the overlap area, considering that the average lifecycle of tropical convection is smaller than 60 minutes. Within the main statistics, the clusters with number of layers equal to "0" (for only having the 20-dBZ reflectivity threshold) or "1" (for having both 20 and 40-dBZ reflectivity [18] (4,62) : what is "layers" - number of cappis? or multiple threshods? This is not described previously and needs to be explained here. deep convective cores.

For better illustration of the terminology of TATHU tracking system applied to SIPAM 3-km CAPPI fields, Fig. 2 shows a schematic of these definitions: a **cluster** is a contiguous region (polygon) of pixels within the 3-km CAPPI reflectivity field that exceeds 20 dBZ reflectivity threshold identified in a continuous sequence of clusters that are linked across radar volumes [21] (4,70) : OK.. this should have been described around line 78. In addition, you should indicate why you are tracking clusters and not cells - i.e. the scientific rationale. cores that exceed 40 dBZ reflectivity threshold, an estimation of deep convection, but these cores are not considered as separated clusters during the description and tracking of the convective systems. [20] (4,77) : Do you mean "these cores are considered as part of a single cluster for tracking."

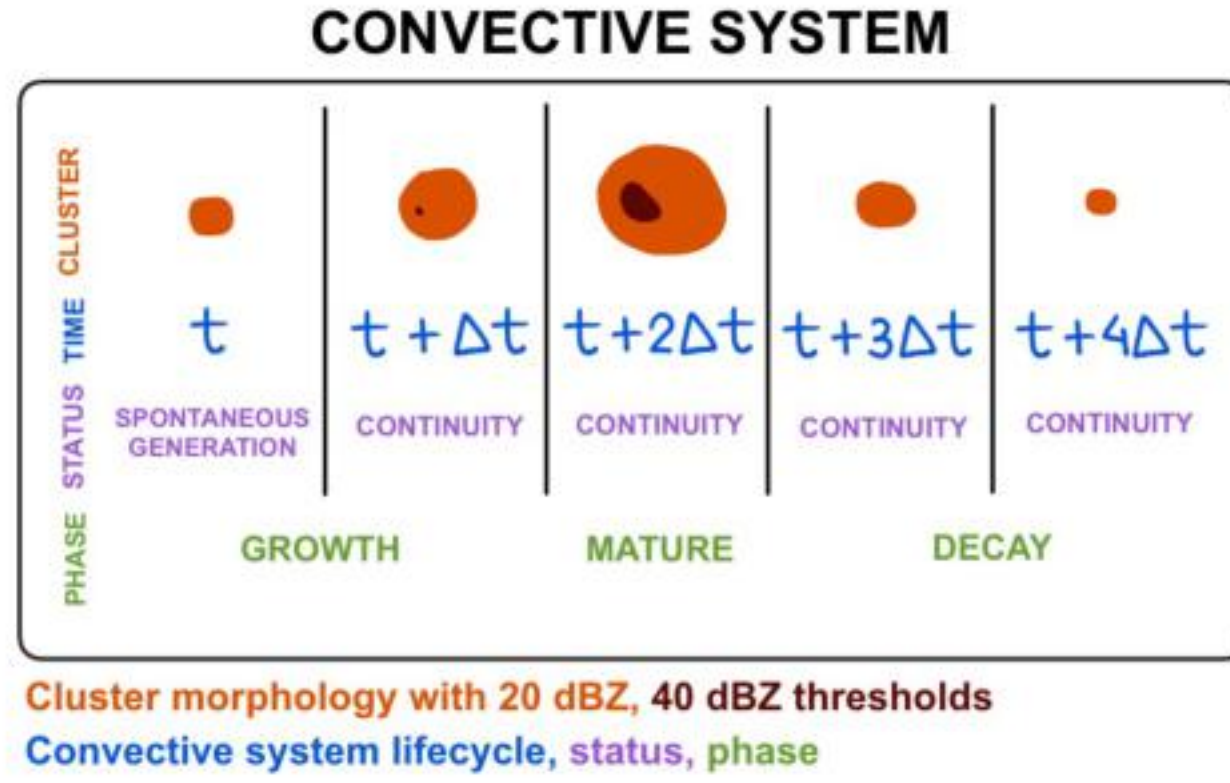
Output data was stored in a PostGIS database, a Data Base Management System (DBMS) with geospatial support that allows storage of large volumes of data in tabular form, including geolocated geometries. The database was converted to GeoJSON datasets in order to become easily available at Lopes (2024).

Table 2. TATHU parameters (original names in parentheses) and values chosen for the generation of *systems* (raw) and *systems_filtered* (filtered) datasets.

Input data	3-km CAPPI in a 150 x 150 km box (300 x 300 elements) centered at the radar location, between January 2014 and December 2015
Reflectivity thresholds (<i>threshold value</i>)	values greater than 20 dBZ, 40 dBZ
Minimum cluster sizes (<i>minarea</i>)	100 km ² , 40 km ²
Tracking technique (<i>trackers class</i>)	Relative overlap area (<i>RelativeOverlapAreaStrategy</i>)
Minimum cluster overlap area	10%
Maximum interval between images	60 min
Statistics (<i>stats</i>)	Maximum, mean and standard deviation of reflectivity, size (amount of pixels), number of layers (corresponding to having one or two reflectivity and minimum size thresholds)
Output	PostGIS database (<i>systems</i> and <i>systems_filtered</i> tables)
Filters applied	- Have at least one 40 dBZ/40 km²(corresponding to deep convection) pixel in any timestamp - Do not intersect the bounding box of the grid (corresponding to probably have part of the cluster outside the tracking region) - Last longer than 12 minutes (one timestamp) or have relation with other convective system (by split or merge)

Two datasets were defined based on the TATHU tracking. The original - called herein **raw** - dataset contains all the convective systems observed in the period, regardless of duration, size and status during the life cycle. The **filtered** dataset contains only the convective systems that met the filtering criteria described in Table 2. The second threshold criteria (40 dBZ in a 40 km² minimum area) is used here as a definition of deep convective cores present during the lifecycle. Clusters intersecting the borders of the grid had their associated convective systems discarded to exclude systems without full lifecycle within the study area. Systems with only one timestep were also discarded. We created this filtered dataset to provide a subset of deep convective systems with full life cycle, important criteria for several convection studies.

The original *systems* dataset contains 91609 convective systems and 322896 clusters, while the filtered *systems_filtered* dataset contains 5976 convective systems (6.5% of the original dataset) and 40394 clusters (12.5% of the original dataset). Using the filtered dataset with the additional lightning data, several other parameters were calculated for each cluster (Table 3) related to storm morphology. The following equations were applied for VIWL (Vertically Integrated Warm Liquid), VII (Vertically Integrated Ice) and VIL (Vertically Integrated Liquid), respectively:



[23] (6,11) : Good. A suggestion is to include $t - \Delta t$ and $t+5 \Delta t$ to show initiation and then complete decay

Figure 2. Illustrative deep convective system as observed by TATHU: clusters morphology with 20-dBZ (orange) and 40-dBZ (dark red) thresholds in 5 subsequent timestamps with TATHU status (spontaneous generation or continuity) and different phases (growth, mature and decay) of a full lifecycle.

$$VIWL = \sum_{i=2km}^{5km} 3.44 \times 10^{-6} \left(\frac{Z_i + Z_{i+1}}{2} \right)^{\frac{4}{7}} \quad (1)$$

$$VII = \sum_{i=7km}^{16km} \pi \rho_i N_o^{\frac{3}{7}} \left[\frac{5.28 \times 10^{-18}}{720} \left(\frac{Z_i + Z_{i+1}}{2} \right) \right]^{\frac{4}{7}} \quad (2)$$

$$135 \quad VIL = \sum_{i=2km}^{16km} 3.44 \times 10^{-6} \left(\frac{Z_i + Z_{i+1}}{2} \right)^{\frac{4}{7}} \quad (3)$$

3 Results

3.1 Raw and filtered datasets characteristics

This subsection compares the physical characteristics between the raw and filtered datasets in order to define what type of convective system each dataset represents. Fig. 3 shows the size — represented by area in km^2 — distribution of clusters and maximum size of convective systems. Both distributions have maximum frequency in the smallest (1000 km^2) area. The maximum area of the raw systems is 6 times larger (60000 km^2 vs 10000 km^2) and the distribution drops faster in the filtered systems. These characteristics show the filtering effect around the border of the 150 km bounding box (Fig. 1), which excluded very large clusters: a 200×200 points clusters with a 40000 km^2 area, for example, can be considered too large because it occupies 2/3 of the grid and probably intercepts the border of the bounding box in a given time stamp.

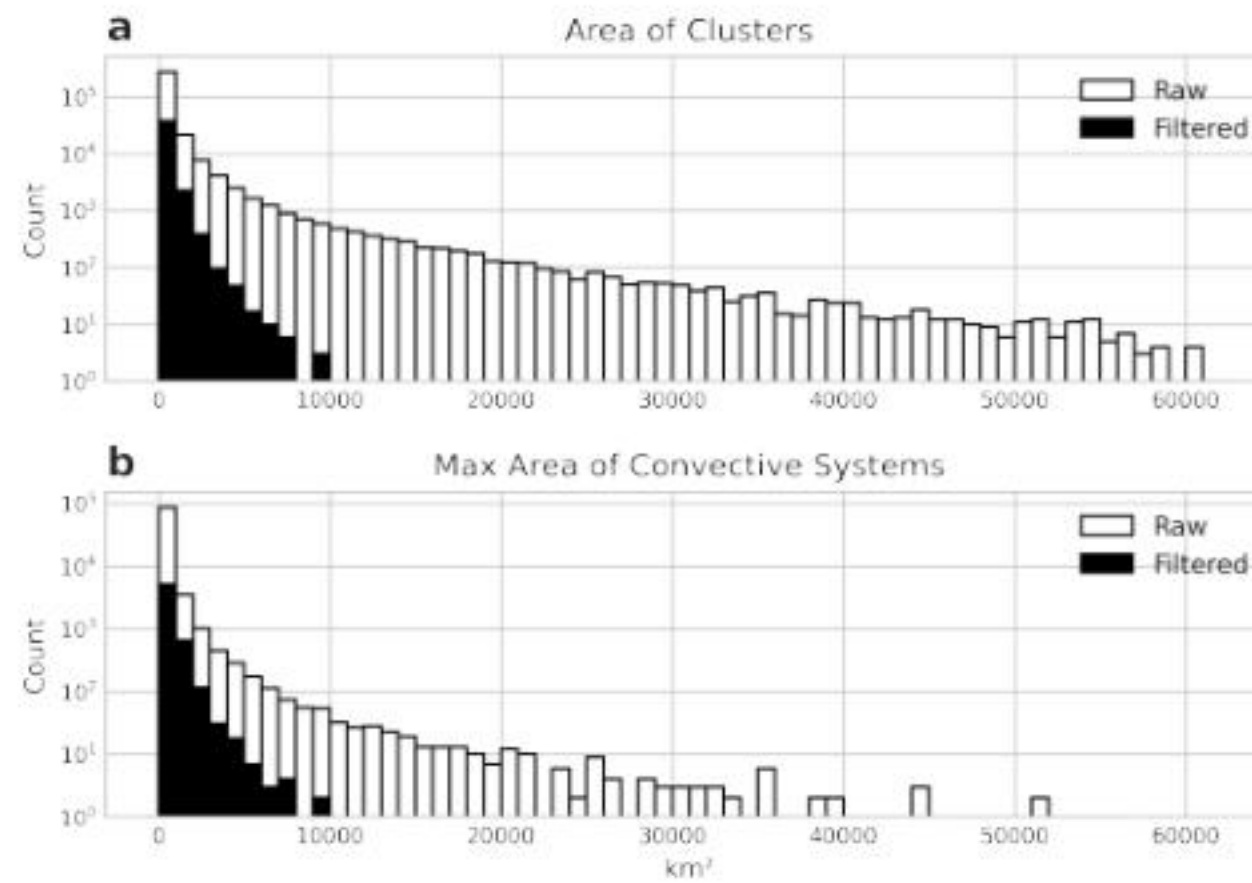
[24] (6,90) : replace : with . and start a new sentence.

Table 3. Clusters additional parameters calculated on the filtered dataset

Variable	Description	Reference
<i>gld</i>	GLD strokes within cluster area detected within 12 min (interval between scans)	
<i>echotop_0</i> , <i>echotop_20</i> , <i>echotop_40</i>	0, 20 and 40 dBZ echo top heights	
<i>z_freq</i>	Reflectivity frequencies per height; 15 x 16 matrices with reflectivities between -10 and 70 dBZ (every 5 dBZ) and heights between 2 and 16 km	Yuter and Houze (1995)
<i>viwl_kgm2</i>	VIWL (Vertically Integrated Warm Liquid) of cluster in kg/m ² , Equation 1	
<i>vii_kgm2</i>	VII (Vertically Integrated Ice) of clusters in kg/m ² , Equation 2	Petersen and Rutledge (2001)
<i>vil_kgm2</i>	VIL (Vertically Integrated Liquid) of cluster in kg/m ² , Equation 3	Greene and Clark (1972)
<i>nae_s_1</i>	Normalized area expansion in s ⁻¹ between clusters of a same convective system	Machado and Laurent (2004)
<i>gld_strmin</i>	Strokes rate between clusters of a same convective system in strokes/min	
<i>echotop0_kmin</i> , <i>echotop20_kmin</i> , <i>echotop40_kmin</i>	0, 20 and 40 dBZ echo top rate between clusters of a same convective system in km/min	

145 Fig. 4 shows the mean and maximum clusters reflectivity distribution of raw (a) and filtered (b) datasets. On raw data, the distributions show no significant frequency peaks, with mean reflectivity distributed mainly (frequency above 20%) between 20 and 40 dBZ and maximum reflectivity between 35 and 50 dBZ (frequency above 15%). Oppositely, on the filtered data, peaks (above 35%) can be found between 25 and 40 dBZ and between 50 and 55 dBZ of mean and maximum reflectivity, respectively. These differences between the distributions represent the filtering effect on the type of the selected clusters: the
150 filters ended up selecting more intense clusters (larger mean and maximum reflectivity) and excluded mainly the ones that did not exceed 40 dBZ — observe that the clusters with maximum reflectivity below 40 dBZ are significantly less frequent (below 5%) compared to the raw data ones.

[27] (7,86) : So, it is one filter that makes all the difference. The one requiring a pixel > 40 dBZ. And not the boundary filter. Correct/ ??



[26] (8,10) : The shape of the PDF of the cluster area and max convective system area look similar, with the maximum values the same. Does this make sense? Should the max value of the convective systems be larger?

Figure 3. Distribution of clusters area (a) and convective systems' max area (b) of raw (white) and filtered (black) datasets.

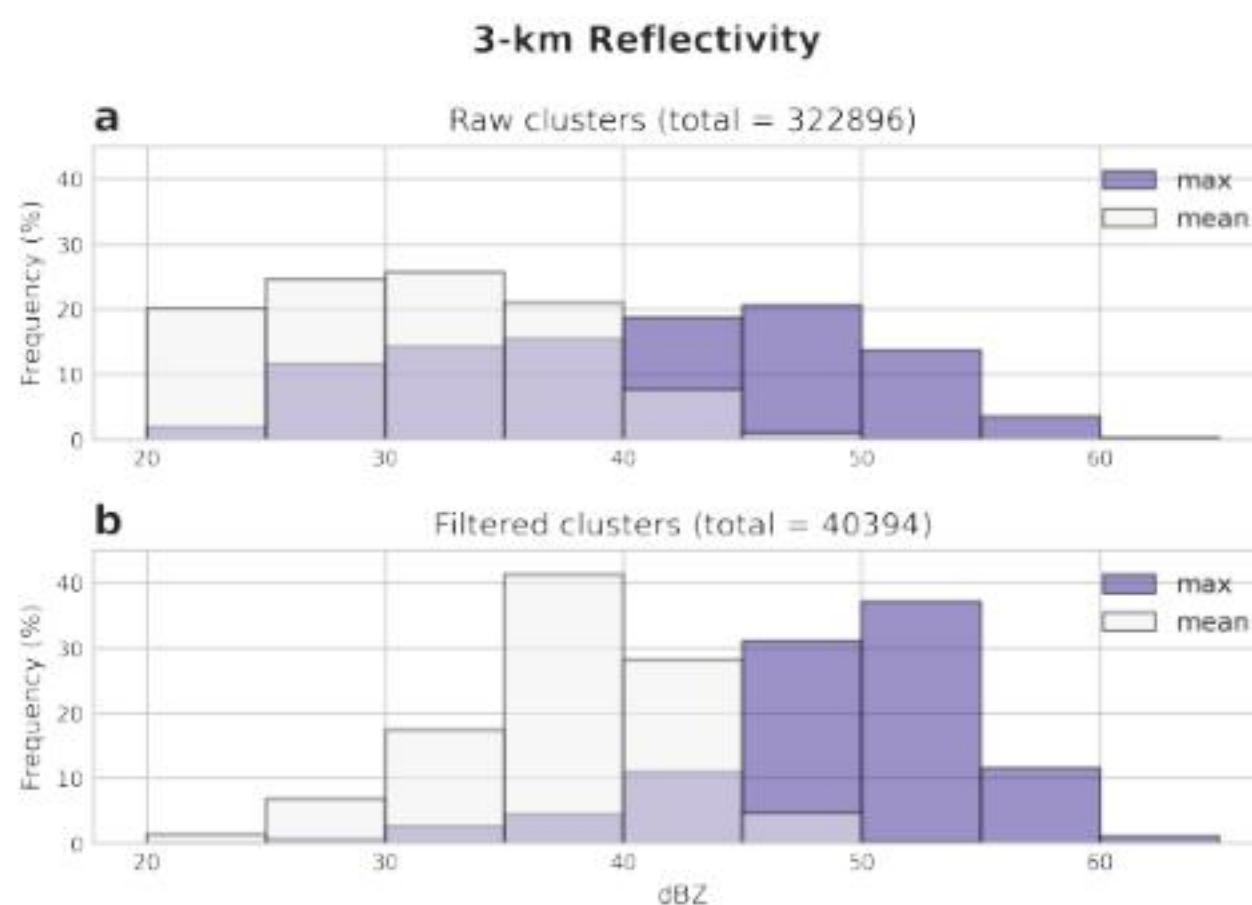


Figure 4. Distribution of clusters max (purple) and mean (white) reflectivity of raw (a) and filtered (b) datasets.

Table 4 shows some characteristics of the convective systems of raw and filtered datasets using the clusters classifications on each time step. On both datasets, the percentage of spontaneously generated convective systems was similar (above 70%); on the raw data, not all these systems had their full lifecycle covered, while on the filtered data this is true because one of the filtering criteria is to exclude convective systems that leave the radar coverage area. 53% of the filtered convective systems had split or merge during their lifecycle, compared to 37% of the raw systems; an important point here is that 31% of the raw systems and only 2% of the filtered systems lasted only one time step (12 min) (not shown), meaning that raw convective systems did not last long enough to go through split or merge. The majority (80%) of the filtered and half (49%) of the raw convective systems were considered with full lifecycle (last time step was “continuity”), which also is explained by the high

[30] (8,86) : There is a mesmerizing number of facts here. Consider making bullet points so that the reader does not have to do so much mental gymnastics.

Table 4. Frequency of spontaneously generated (did not generate from splits or first steps of algorithm rounds), with split and/or merge and full lifecycle (last time step in continuity) convective systems (CS) of raw and filtered datasets.

	Raw	Filtered
Convective Systems (total)	91609	5976
CS spontaneously generated (%)	72	73
CS with split/merge (%)	37	53
CS with full lifecycle (%)	49	80

percentage of raw systems with only one time step (i.e., only “spontaneous generation” status). The 20% of filtered convective systems without full lifecycle are within the not spontaneously generated systems, products of split or merge of other systems.

Table 5 shows the distribution of raw and filtered (47%) systems lasting between 1 and 3h, compared to 36% of filtered systems. The majority of convective systems were short-lived, with 40% of raw and 47% of filtered systems lasting up to 1 hour, highlighting the prevalence of isolated convective systems. A notable difference is observed in the 1–3h range, where only 10% of raw systems persisted, compared to 36% in the filtered dataset. For long-duration systems, a significant disparity emerges: nearly twice as many raw systems (631) lasted longer than 6h compared to those in the 5–6h range (318). In contrast, the filtered dataset shows fewer systems exceeding 6h (18) than those lasting between 5–6h (27), indicating a stronger filtering effect on prolonged convective systems.

Fig. 5 shows the monthly distribution of raw and filtered clusters and convective systems. Seasons and intensive operation periods (IOPs) were defined according to Machado et al. (2018): dry season between August and October, dry-to-wet season between November and December, wet season between January and March, IOP1 between February 1st and March 31st 2014 and IOP2 between August 15th and October 15th 2014. In general, a larger frequency of clusters and convective systems occurred in the wet and transition (wet-to-dry and dry-to-wet) seasons, with a peak of filtered clusters/systems in November 2015 (almost two times more than the raw clusters/systems). The proportion between raw and filtered systems changes over the months, with a larger frequency of raw clusters and convective systems on the wet seasons and the opposite on the dry seasons. This difference can be explained by the filter of very large clusters cited previously, which are more common in the wet season. Considering the climatological characteristics of each season, it is expected that more clusters/convective systems occur during the wet season than of the dry season, which was the case for both raw and filtered data.

Table 6 presents the duration of raw and filtered convective systems by IOP. The distribution of system durations is generally consistent across seasons, with similar proportions observed between raw and filtered datasets. In contrast to Table 5, where less than half of the raw systems lasted up to 1h, seasonal distributions show that more than 80% of raw systems and 50% of

[28] (9,35) : How can this be? The definition of the filtered dataset is that you have full lifecycle. So, you must have another category or you define a convective system as "new" if it originates from a split. Not sure what you do with a merge. Do you kill one off and declare complete life cycle or ???

[29] (9,55) : Which filter? Please provide insight on which filter to provide a sense of science.

[31] (9,90) : The way you compare raw vs filtered system implies scientific or meteorological value. While it serves as a warning to users, you are much better off if you focus on the filtered results. The only conclusion that you can come to is the filters are the reason for the difference. You can say this once and then there is nothing interesting.

Table 5. Frequency of spontaneously generated (did not generate from splits or first steps of algorithm rounds), with split and/or merge and full lifecycle (last time step in continuity) convective systems (CS) of raw and filtered data tables.

	Raw	Filtered
CS duration $\leq 1h$	79301 (86,6%)	3448 (57,7%)
$1h < \text{CS duration} \leq 2h$	6950 (7,6%)	1535 (25,7%)
$2h < \text{CS duration} \leq 3h$	2597 (2,8%)	635 (10,6%)
$3h < \text{CS duration} \leq 4h$	1170 (1,3%)	239 (4%)
$4h < \text{CS duration} \leq 5h$	641 (0,7%)	74 (1,2%)
$5h < \text{CS duration} \leq 6h$	318 (0,3%)	27 (0,5%)
CS duration $> 6h$	631 (0,6%)	18 (0,3%)
Total	91609 (100%)	5976 (100%)

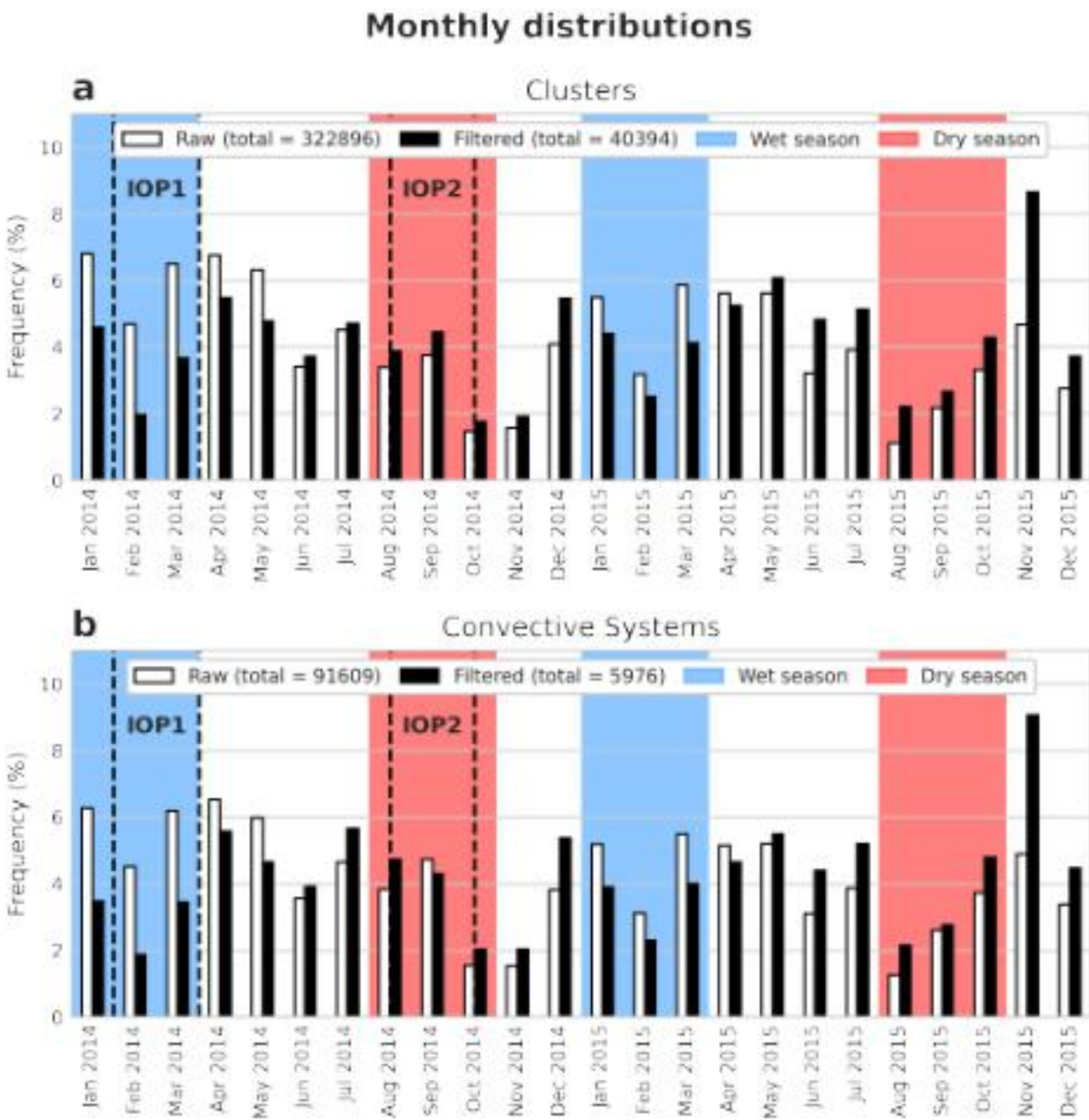


Figure 5. Monthly relative frequency distribution of clusters (a) and convective systems (b) of raw (white) and filtered (black) datasets. The blue and red areas delimit wet and dry seasons, respectively, and dashed lines delimit the intensive operation periods IOP1 and IOP2.

185 filtered systems persisted for no more than 1h, indicating the predominance of isolated convective systems. Additionally, 97% of raw and 94% of filtered systems had durations of 3h or less.

Table 6. Distribution of convective systems (CS) durations of raw and filtered datasets separated by intensive operation periods IOP1 and IOP2.

	IOP1		IOP2	
	Raw	Filtered	Raw	Filtered
CS duration $\leq 1h$	8427 (86,1%)	176 (55,7%)	6565 (90%)	272 (55,5%)
$1h < \text{CS duration} \leq 2h$	772 (7,9%)	88 (27,8%)	432 (5,9%)	130 (26,5%)
$2h < \text{CS duration} \leq 3h$	277 (2,8%)	34 (10,8%)	146 (2%)	53 (10,8%)
$3h < \text{CS duration} \leq 4h$	118 (1,2%)	10 (3,2%)	60 (0,8%)	24 (4,9%)
$4h < \text{CS duration} \leq 5h$	66 (0,7%)	6 (1,9%)	41 (0,6%)	7 (1,4%)
$5h < \text{CS duration} \leq 6h$	40 (0,4%)	0 (0%)	19 (0,3%)	2 (0,4%)
CS duration $>\leq 6h$	91 (0,9%)	2 (0,6%)	31 (0,4%)	2 (0,4%)
Total	9792 (100%)	316 (100%)	7294 (100%)	490 (100%)

[36] (11,44) : "divided" not needed

Fig. 6 shows the hourly distribution of raw and filtered clusters divided by IOP. Comparing the raw and filtered clusters, in all seasons there are a larger frequency of raw clusters during late night/dawn and a larger frequency of filtered clusters during late morning/afternoon. In the dry season (IOP2), this difference is more pronounced: the filtered clusters are more frequent (above 15%) between 1400 and 1500 local time compared to the raw clusters (below 10%). These differences between raw and filtered clusters indicate a more diurnal characteristic of the filtered clusters, while the raw clusters are more nocturnal, probably represented by the very large and long-lasting (above 6h) clusters described previously.

[37] (11,50) : "occurrence" not "frequency" or "the frequency during late ..."

[38] (11,58) : This is scientifically misleading - makes no sense.

Fig. 7 shows the hourly distribution of raw and filtered convective systems initiation (defined here as the first timestamp) divided by IOP. Comparing the raw and filtered systems, in both seasons there are a larger frequency of raw systems initiating during dawn, while a larger frequency of filtered systems initiate during morning/afternoon. This difference is even greater (almost 10%) in the dry season. Comparing the seasons, both in the dry and wet seasons the initiation peak occurs only at 1400 local time. Specifically about the filtered systems, the 1400 local time peak during the wet season is not highlighted in IOP1, with another peak at 1200 local time.

When comparing the physical characteristics of the raw and filtered datasets, the filters significantly decrease the sampling size yet defines a specific subset of convective systems: in general, they have areas up to 10000 km², maximum reflectivity mainly above 50 dBZ, full lifecycle (growth, mature and decay phases) and last up to 3h. The raw dataset includes these same systems as well as diverse convective types such as mesoscale (considering the areas above 50000 km²), both short-lived (12 min duration) and long-lived (duration above 6h), and stratiform (reflectivity below 30 dBZ) systems. Both datasets are

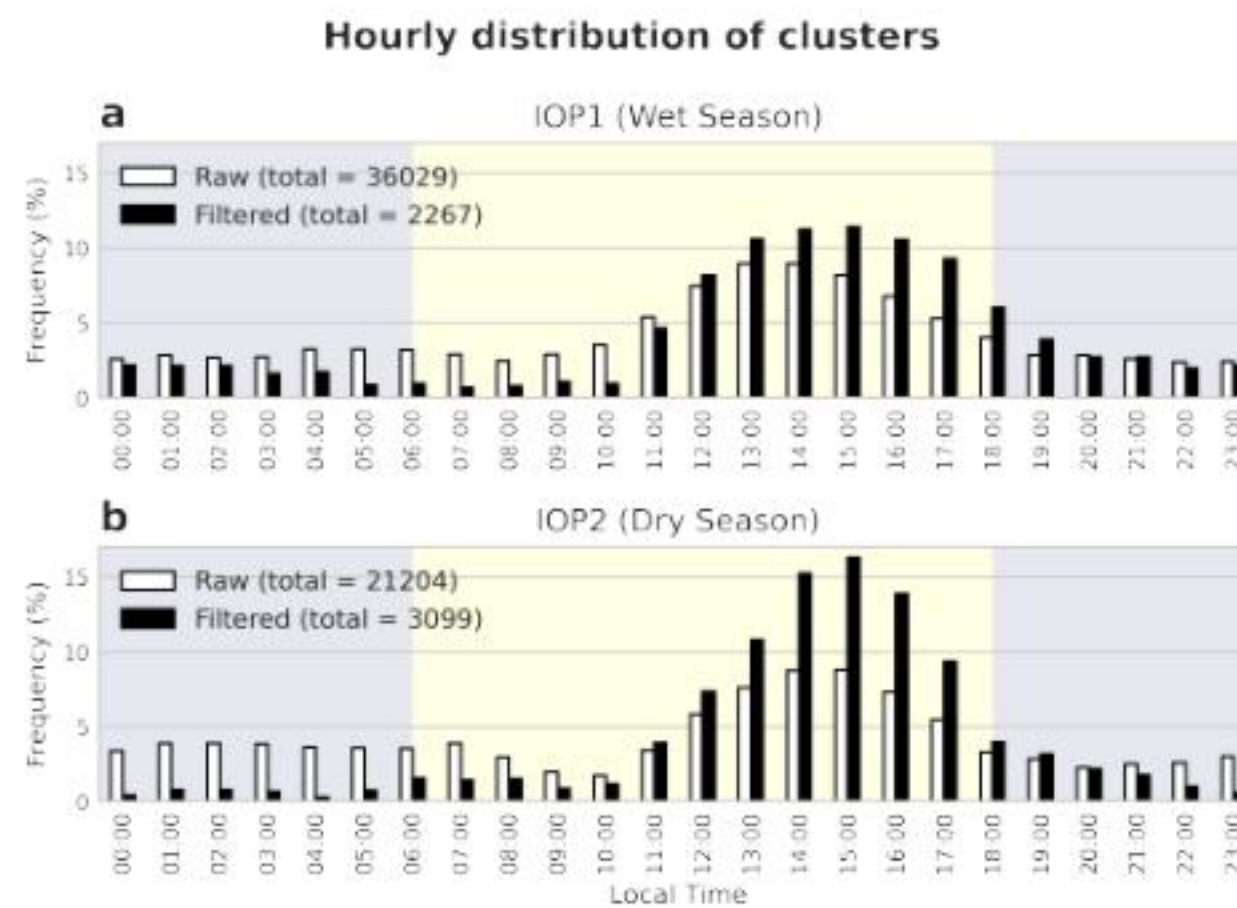


Figure 6. Hourly relative distribution of clusters of raw (white) and filtered (black) datasets separated by intensive operation periods IOP1 (a) and IOP2 (b). The blue and yellow areas delimit diurnal and nocturnal periods, respectively.

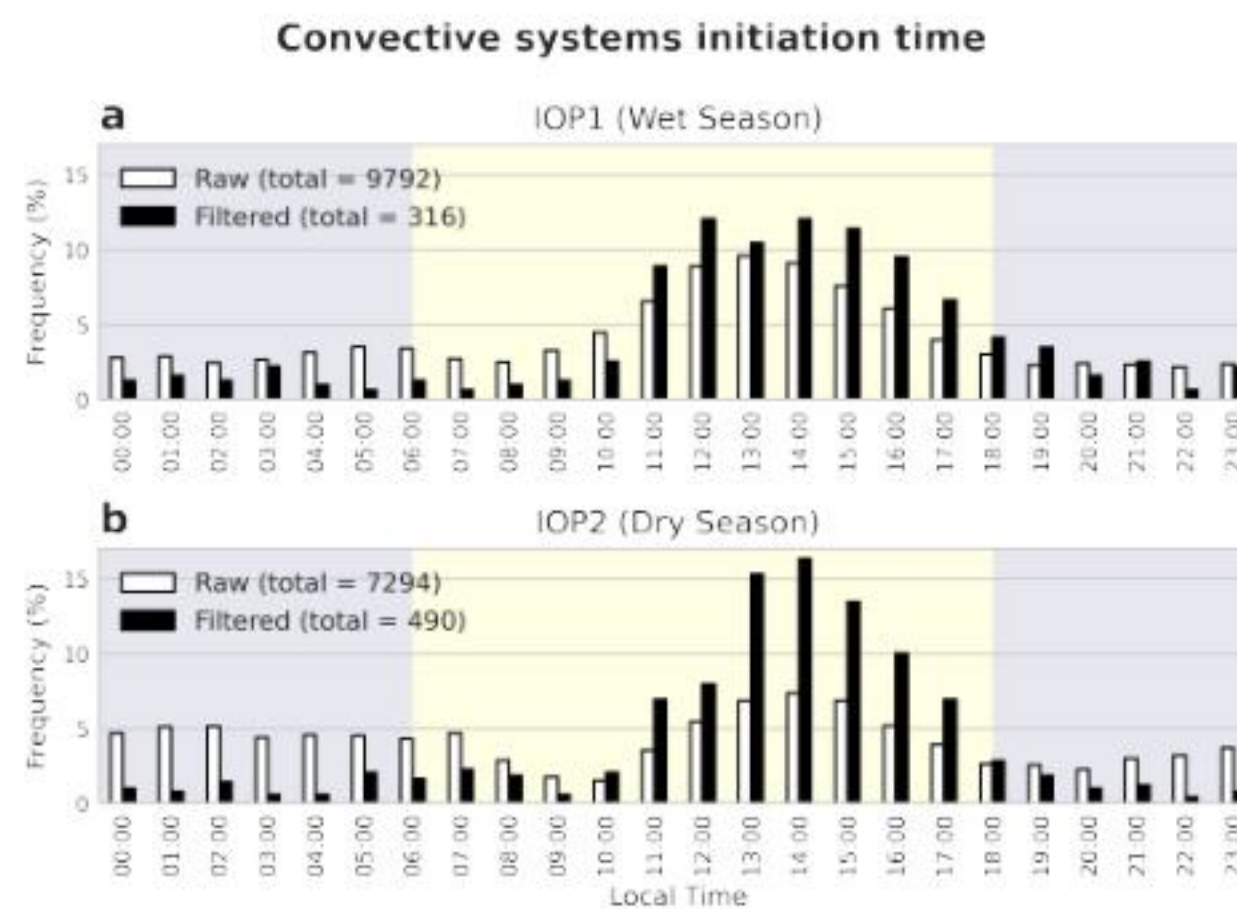


Figure 7. Hourly relative distribution of convective systems initiation time of raw (white) and filtered (black) datasets separated by intensive operation periods IOP1 (a) and IOP2 (b). The blue and yellow areas delimit diurnal and nocturnal periods, respectively.

useful in further convection studies but should be chosen rigorously depending on their objectives: if the convection phase is
 205 important, for example, the filtered dataset will be more appropriate.

3.2 Ch. [39] (12,75) : All of this makes no scientific sense. All you are showing is that don't use raw results. Have to use scientific judgement - I would shorten tis section significantly. Summarize in a table and then say users beware" or you are discrediting other studies that did not filter as you did.

Starting with the monthly distribution, Fig. 8 shows the clusters and convective systems separated by lightning activity, as well as the distribution of GLD strokes. More than double the clusters (27214 vs. 13180) had no electrical activity, while more convective systems (3758 vs. 2218) showed lightning, which means that, in general, the convective systems with lightning

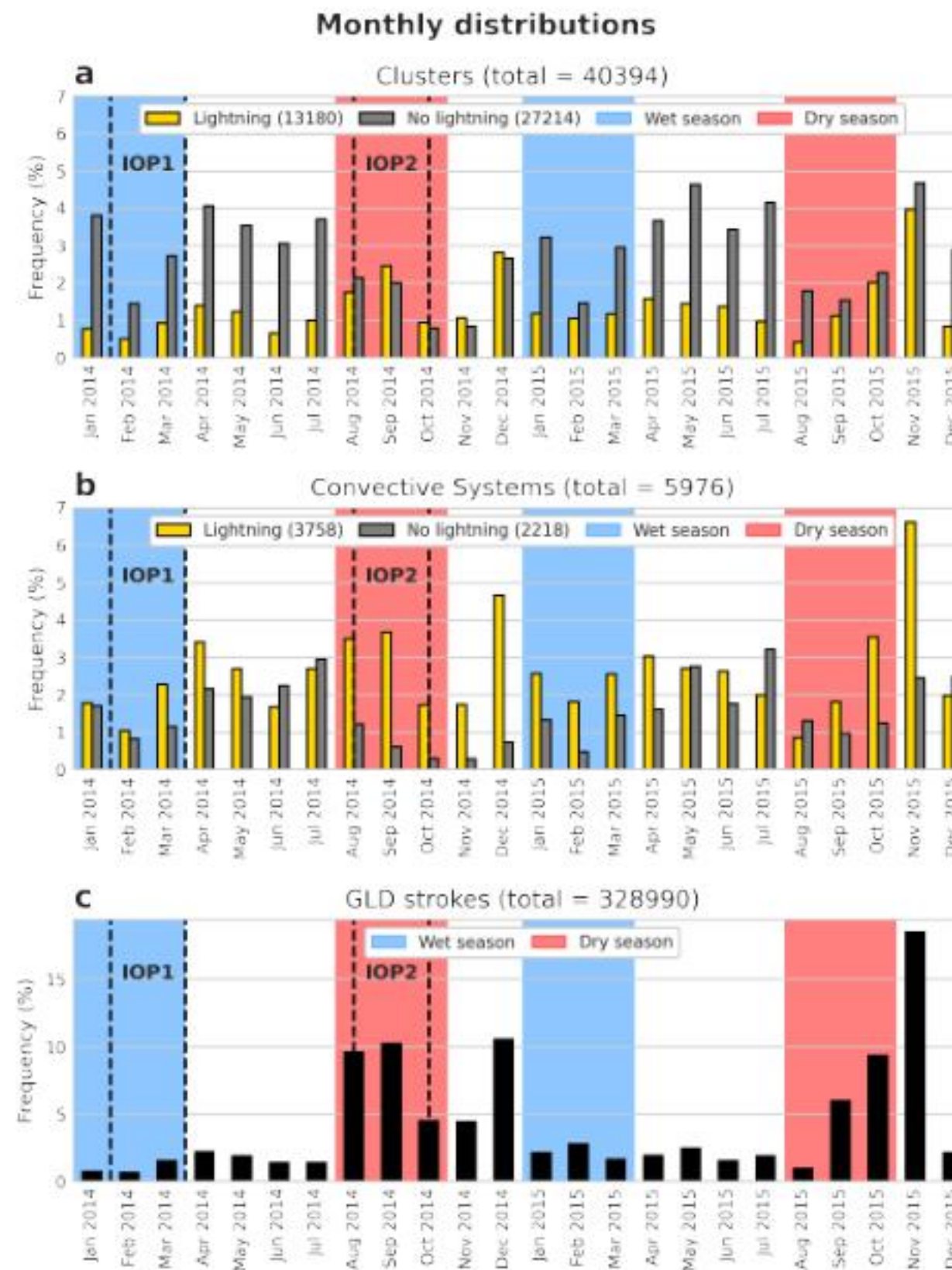


Figure 8. Monthly relative distribution of clusters (a), convective systems (b) and GLD strokes (c) of filtered datasets separated by lightning (yellow bars) or no lightning (gray bars) occurrence. The blue and red areas delimit wet and dry seasons, respectively, and dashed lines delimit the intensive operation periods IOP1 and IOP2.

consists of only a few clusters with lightning. There are significant differences between clusters and systems frequency: a larger frequency of clusters and systems without lightning occurs between wet and dry seasons, while peaks of clusters and systems frequency occurs in the dry-to-wet season, comparable to the peak strokes frequency. These findings are similar to what is found in previous works such as Albrecht et al. (2011, 2016).

Separating the analysis in systems with and without lightning, Fig. 9 shows the hourly distribution of clusters and initiation (first timestamp) of convective systems with lightning and GLD strokes separated by IOP. All variables are more frequent during late morning/afternoon, with significant differences between IOPs: on IOP2, the clusters and strokes peaks occurs at 1500 local time, which is also the preferential time of convective systems initiation; on IOP1, the clusters and strokes peaks occur around the same time, but the systems initiation occurs preferentially earlier, at 1200 local time. It's important to note that the low amount of systems influenced the frequency distribution, but the results were similar to their corresponding dry and wet seasons (not shown). For the clusters and initiation of convective systems without lightning (not shown), the largest frequencies occur during late morning/afternoon. The initiation frequency peaks on the dry and wet seasons occur at 1500 local

[35] (13,88) : Awkward, rewrite.

[34] (13,90) : Replace with "Convective systems without lightning initiated within the filter area, the largest..."

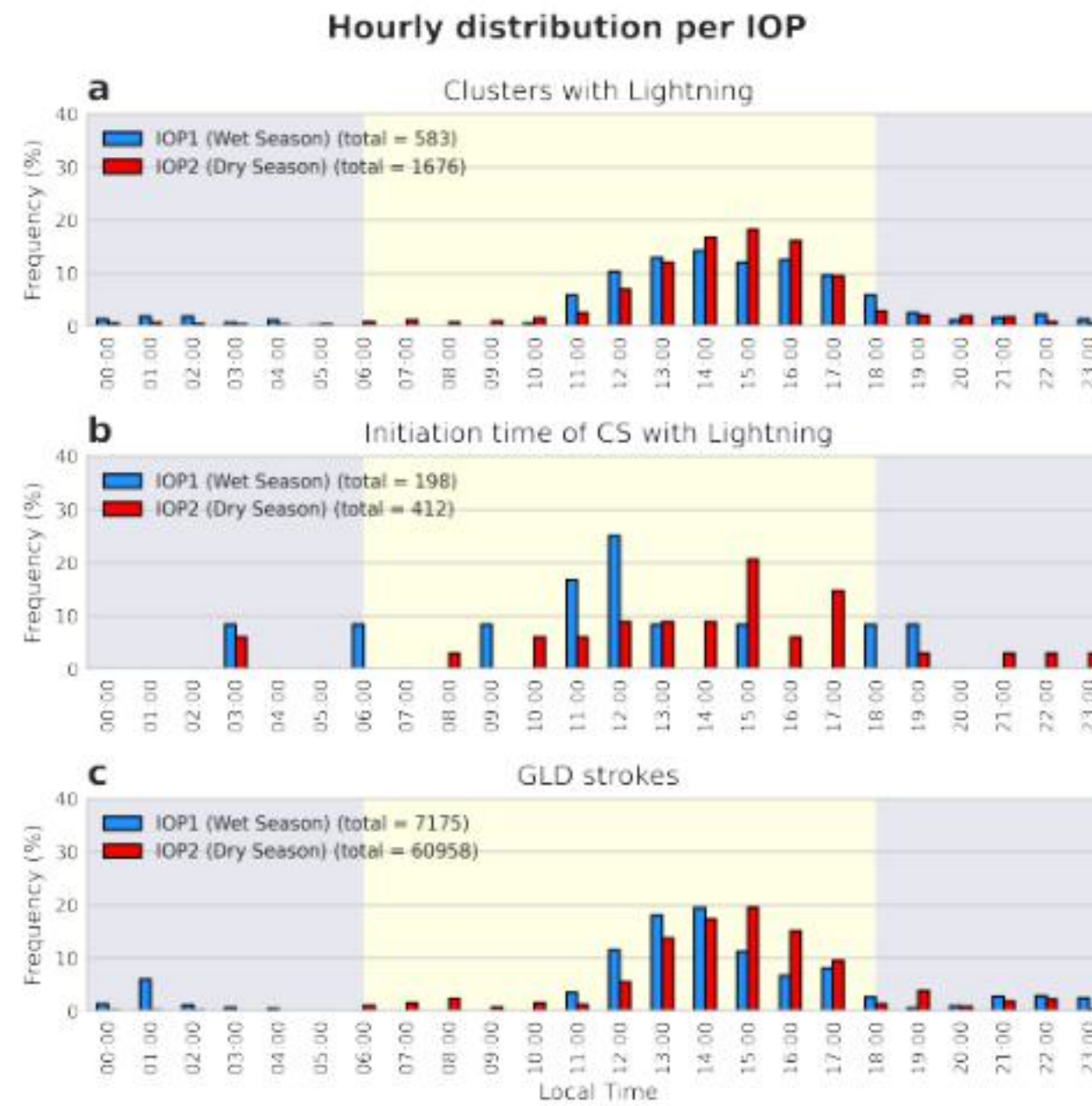


Figure 9. Hourly relative distribution of clusters with lightning (a), initiation time of clusters with lightning (b) and GLD strokes (c) of filtered dataset during intensive operation periods IOP1 (blue) and IOP2 (red). The blue and yellow areas delimit diurnal and nocturnal periods, respectively.

time, while on the dry-to-wet season the distribution is more dispersed, with approximately equal peaks at 1200, 1300 e 1600 local time.

In order to analyze the propagation direction of the convective systems, Fig. 10 shows the frequency distribution of the propagation direction separated by IOP. The predominant direction of the convective systems is from the east, consistent with the main dynamic forcing in the Amazon: moisture flux from the tropical Atlantic by the easterlies influenced by the position of the Intertropical Convergence Zone (ITCZ) (Silva Dias and Carvalho, 2016). Comparing the IOPs, in the wet season, about 25% propagated from the east and east-northeast, while in the dry season more than 30% of the systems propagated from the east and east-southeast. This direction shift from east-northeast to east-southeast is related to the shift in the position of the ITCZ and the cold fronts propagation in the South American continent, which affect the zonal winds regime in the Amazonian region (Rickenbach et al., 2002). These results are also consistent with Gupta et al. (2024), which focused on isolated convective systems near Manacapuru (T3) site.

In order to analyze the intensity of the convective systems, Fig. 11 shows the frequency distribution of maximum heights and variation rates (considering the 12-min interval between radar scans) of the 0, 20 and 40 dBZ echoes, associated with the cloud top height, precipitable hydrometeors height and intense precipitation heights, respectively. The maximum top heights are more frequent above 10 km, with peaks in 11 km (dry season) and 15 km (wet season). The maximum precipitable height was more frequent between 7 and 11 km. The maximum intense precipitation height was also more frequent between 4 and 7

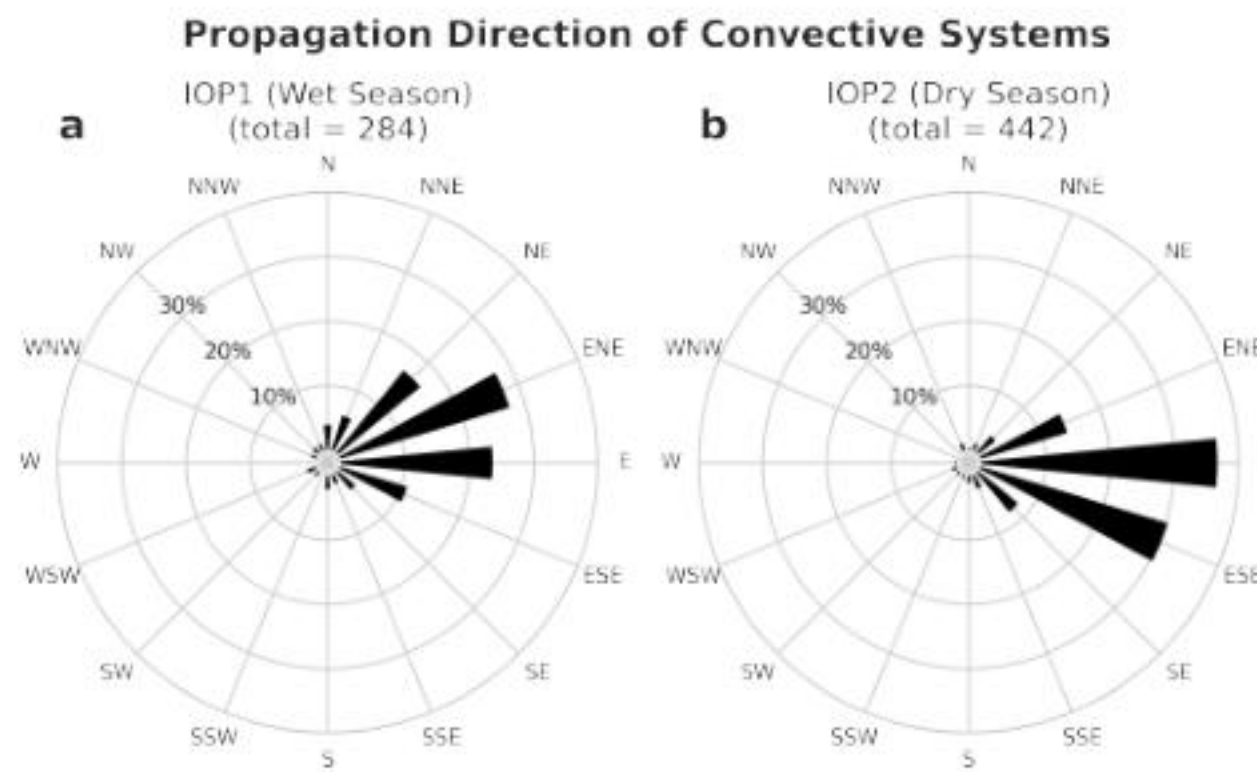


Figure 10. Convective systems' relative propagation direction distribution in the filtered dataset separated by intensive operation periods IOP1 (a) and IOP2 (b). The direction was defined by the distance between the first and last centroids of the convective system.

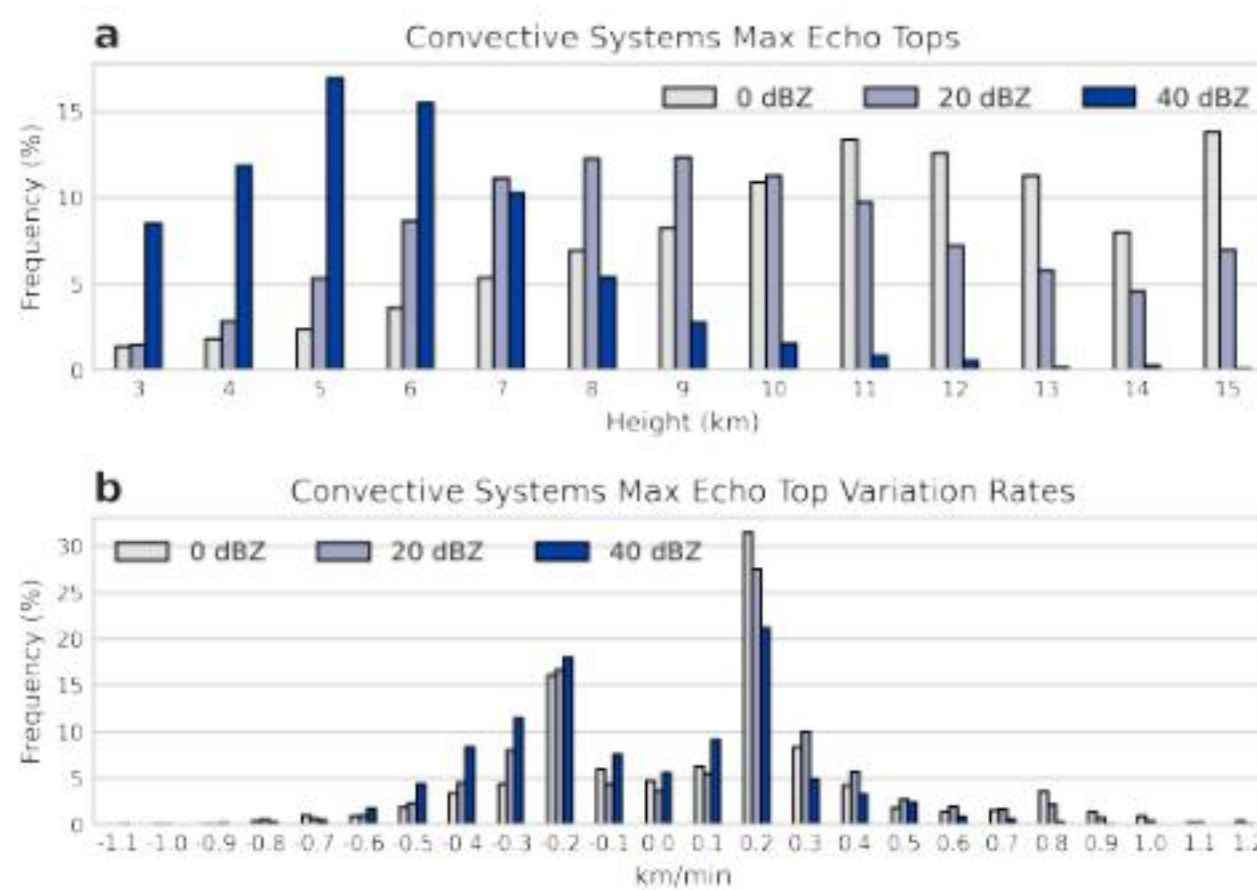


Figure 11. Convective systems' 0 (gray bars), 20 (light blue bars) and 40 dBZ (dark blue bars) max echo tops (a) and variation rates (b).

km, with peaks in 5 km (both dry and wet seasons). These high tops, complemented by maximum precipitation height more frequently between 7 and 11 km and maximum intense precipitation height more frequently between 4 and 7 km, show how these systems were predominantly deep in its most intense moment. The variation rates (Fig. 11b) were similar between the echoes, with frequency peaks in -0,2 and 0,2 km/min, indicating significant fluctuations of the echo tops throughout its life cycle.

Fig. 12 shows the frequency distribution of the maximum echo tops of 0, 20 and 40 dBZ separated by IOPs. As in the complete time series, the maximum top heights are more frequent above 10 km, with peaks in 11 km (IOP2) and 15 km (IOP1). The maximum precipitable height was also more frequent between 7 and 11 km. The maximum intense precipitation height was also more frequent between 4 and 7 km, with peaks in 5 km.

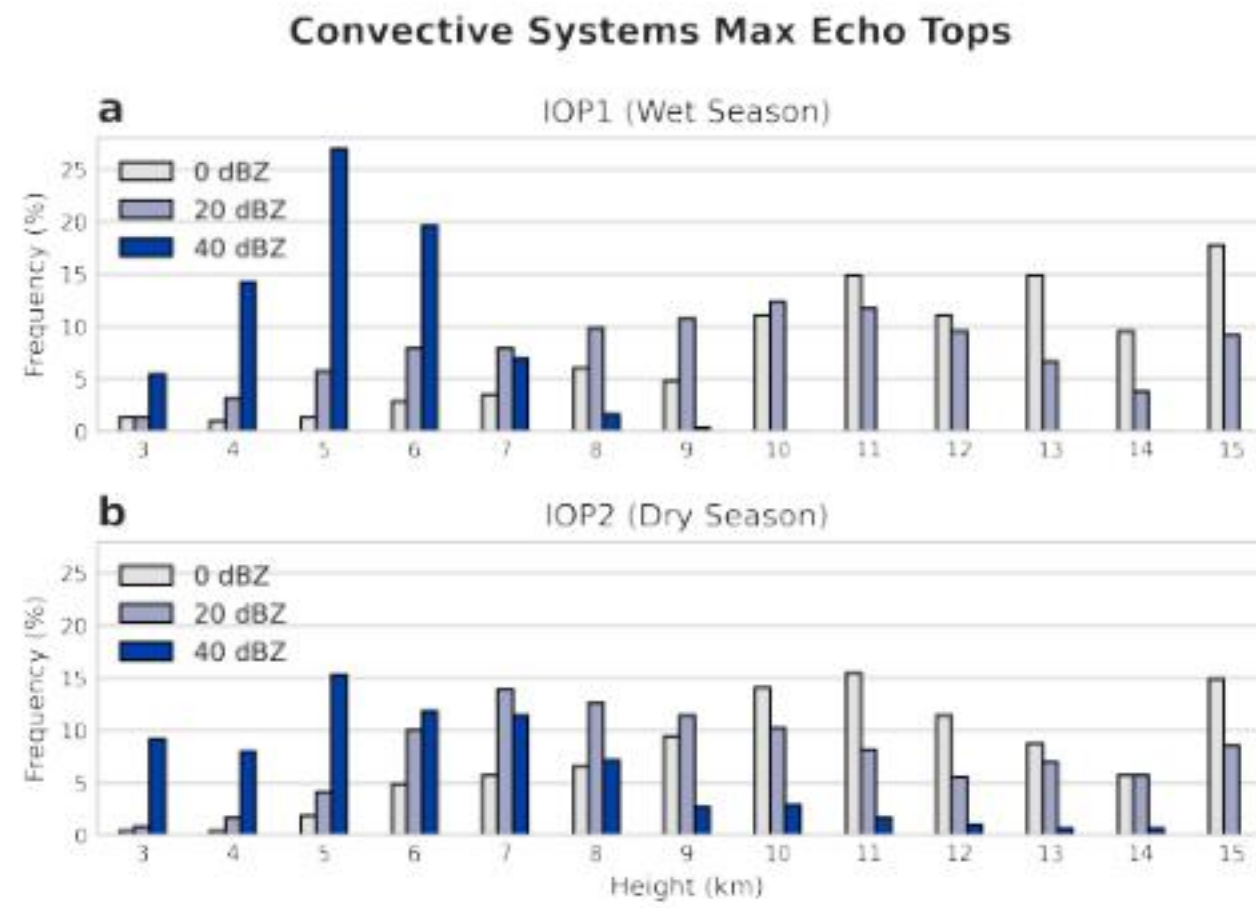


Figure 12. Convective systems' 0 (gray bars), 20 (light blue bars) and 40 dBZ (dark blue bars) max echo tops separated by intensive operation periods IOP1 (a) and IOP2 (b).

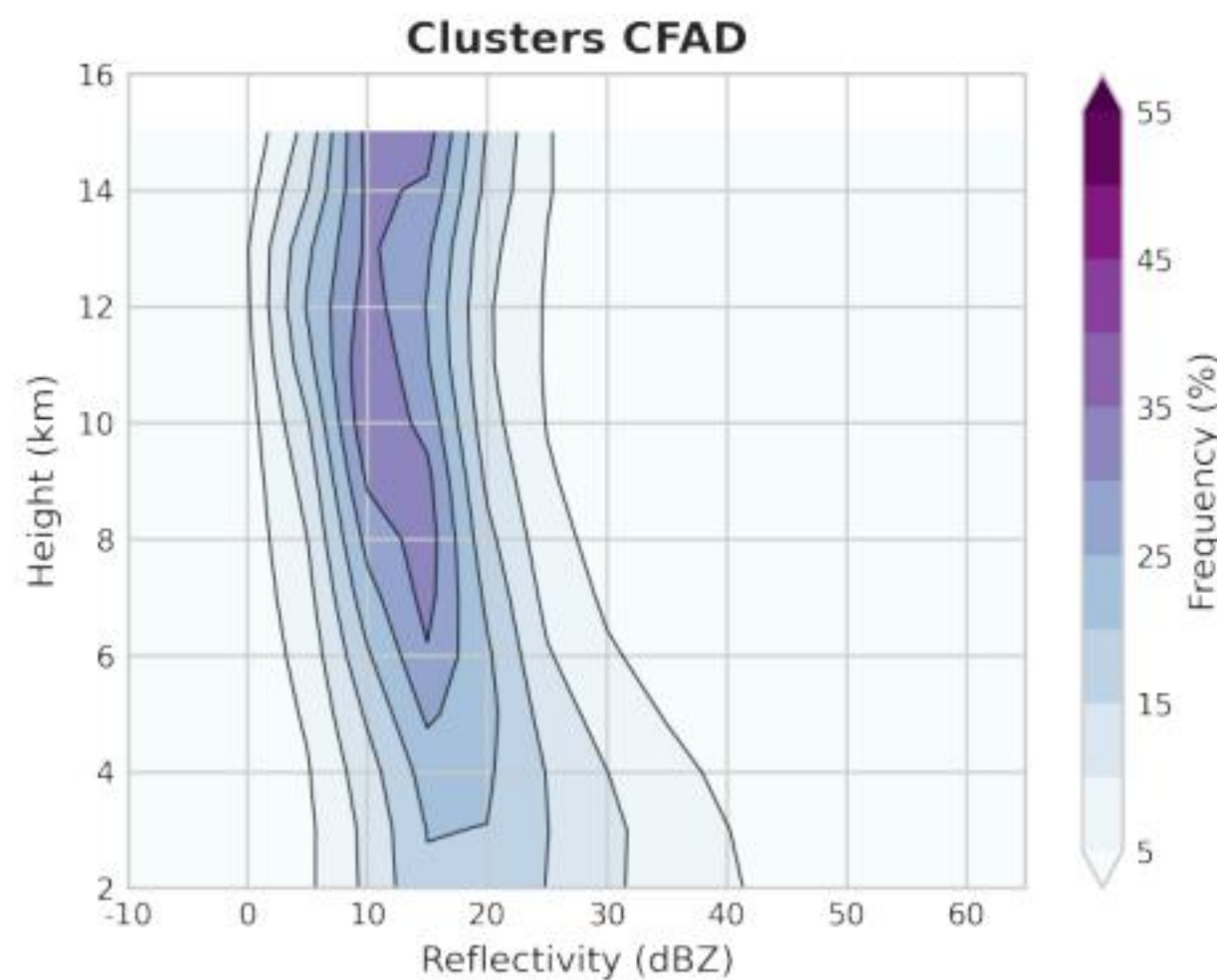


Figure 13. Mean Contoured Frequency by Altitude Diagram (CFAD) of clusters' reflectivity with 5-dBZ bins.

In order to analyze the clusters vertical profile, Fig. 13 shows the frequency diagram by altitude of the clusters reflectivity of the complete time series. The most frequent profile is of a small reflectivity variation with height, 25 dBZ on the surface to 10 dBZ on 15 km height (i.e., a -1 dBZ per km variation rate). Less frequent (between 5 and 10%) profiles have very low (5 to 10 dBZ) or high (40 dBZ) reflectivity on the surface, and up to 25 dBZ on 15 km height, with a minimum of 0 dBZ on 12 km height.

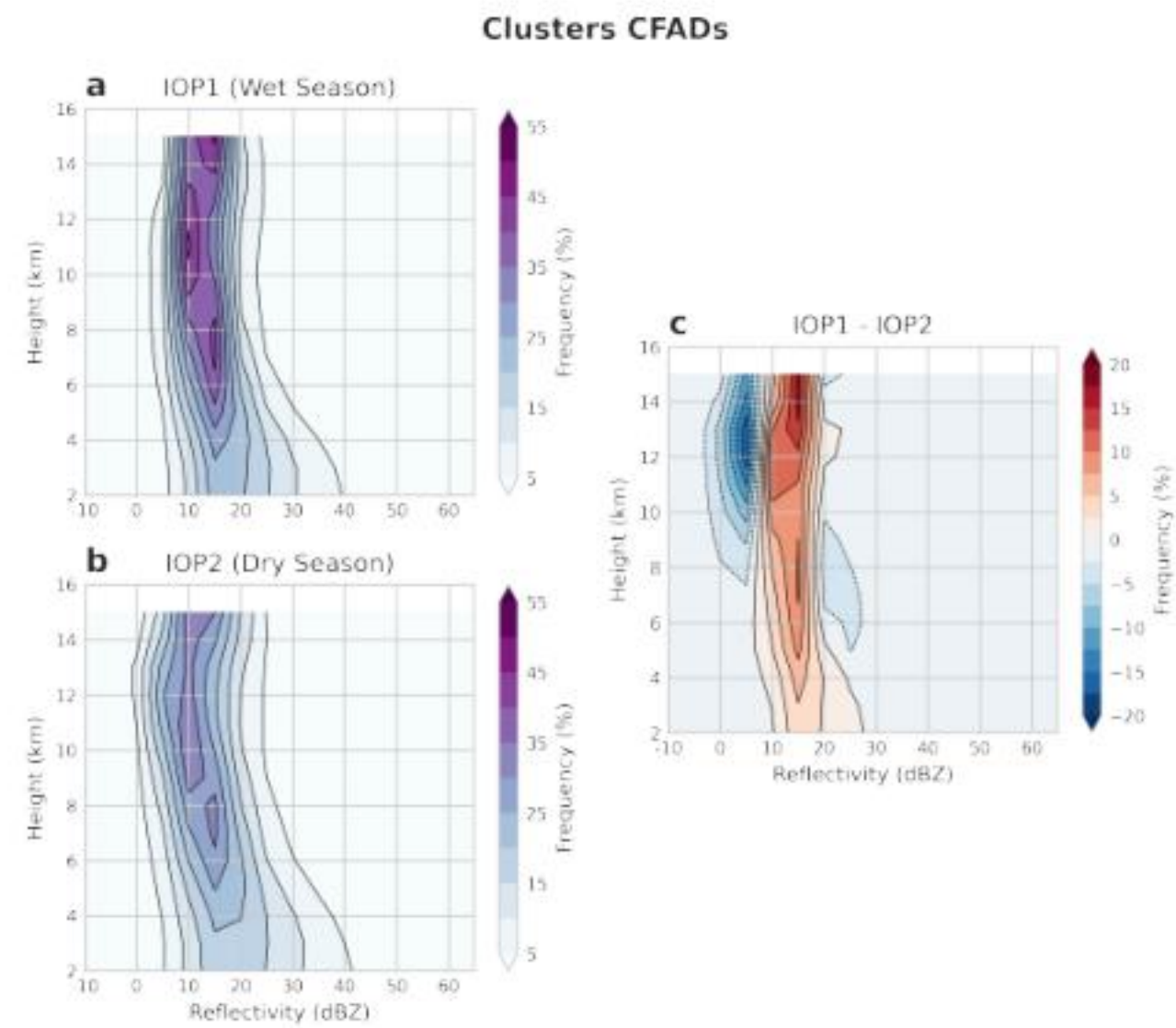


Figure 14. Mean Contoured Frequency by Altitude Diagram (CFAD) of clusters' reflectivity with 5-dBZ bins separated by intensive operation periods IOP1 (a) and IOP2 (b) as well as IOP1 - IOP2 anomaly (c).

Fig. 14 shows the clusters frequency by altitude diagram separated by IOP as well the different between them. Considering the largest frequencies, the profiles are similar between them (and with the complete time series profile), but the IOP1 profile (wet season) is more intense (up to 20%) than the IOP2 (dry season), specially between 10 and 20 dBZ.

[40] (17,51) : This section contains the relevant scientific results. While the sentences could be simplified, otherwise I have no substantive comments.

255 4 Conclusions

A database of convective systems that occurred during GoAmazon experiment was created to provide comprehensive convection data for future GoAmazon studies. The *systems* and *systems_filtered* datasets cover the main convective characteristics, including morphology and intensity, as well as electrical activity. The filtered dataset is shown to be an acceptable sample of the convective systems within the research area. Convection seasonality is and dry-to-wet seasons and less intense in the wet season, typically occurring during late morning/early afternoon. The preeminent propagation direction of these systems are associated with easterlies with a transition from slightly north to slightly south associated with the ITCZ position.

It is important to consider the limitations in the convection description when using these data for future research. Since the SIPAM radar main role is operational, its settings are not optimal for convection research: low spatial (1 km) and temporal (12 min) resolution (considering it is a weather radar), beam blockage during the experiment (Giangrande et al., 2016; Tian et al., 2021), radar software settings change during the experiment. Another limitation is in the tracking itself, specially when dealing with system split/merge (which occurs in a significant portion of the convective systems database) that can be more complex in large, mesoscale systems. characteristics for cloud-aerosol-precipitation research.

270 5 Code availability

The TATHU software package is available at <https://github.com/uba/tathu> (Uba et al., 2022). The code developed to create the datasets with TATHU is available at <https://github.com/cclopes/tathu/tree/sipam-tracking/sipam-tracking>.

6 Data availability

The *systems* and *systems_filtered* datasets are available at <https://doi.org/10.5281/zenodo.13732692> (Lopes, 2024). SIPAM radar data are available at https://ftp.cptec.inpe.br/chuva/goamazon/experimental/level_0/eq_radar/esp_band_s/st_sipam/.

Author contributions. CL and RA designed the study. CL processed the data and wrote the manuscript advised by RA. DU developed the tracking software, helped during data processing, and reviewed the manuscript. TB processed the radar data and reviewed the manuscript. IS helped in providing the radar data and reviewed the manuscript.

Competing interests. No competing interests are present.