Individual Project LTE

Characterizing the global variability of precipitating clouds observed by the space-borne GPM Dual-frequency Precipitation Radar

Final Report

Student:
Lorenzo Comi

EPFL Supervisor:
Gionata Ghiggi

EPFL Professor:
Alexis Berne

22.01.2024

Contents

List of Figures

List of Tables

Abstract

This project builds upon the insights gained from the Tropical Rainfall Measuring Mission (TRMM) and extends the analysis to the Global Precipitation Measurement (GPM) Dual-Frequency Precipitation Radar (DPR), launched in 2014 to enhance our comprehension of precipitation phenomena. Using the GPM-DPR database, we extracted storm patches and synthesized a new database by capturing 88 variables from these patches. These variables, categorized as "Spatial," "Vertical," and "Structural," are designed to be significant in discerning various storm structures. With a specific focus on high-intensity storms, we employed a Self-Organized Map to classify different phenomena based on both structural characteristics and geographical locations. Our initial findings revealed the ability to distinguish diverse storm structures and detect consistent seasonal and geographical variations in the classification. These results provide a solid foundation for further analysis, offering potential avenues to enhance our understanding of these phenomena.

1 Introduction

Studying precipitation is of great importance in gaining a more detailed knowledge of climate processes. Being able to recognize storm structures is also a key aspect in the understanding of the physics underlying these phenomena.

This project aims to investigate the precipitation patterns and their spatial and structural variability by employing the Global Precipitation Measurement (GPM) Dual-frequency Precipitation Radar (DPR) database from 2018 to 2023. This research builds on the insights provided by the Precipitation Radar (PR) of the Tropical Rainfall Measuring Mission Satellite (TMRR). With the launch of the GPM DPR in 2014, the fundamental precipitation observational data record has been extended, giving a more complete understanding of rainfall distribution, intensity and variability worldwide. The DPR is a key instrument on the collaborative NASA/Japan Aerospace Exploration Agency Global Precipitation Measurement (GPM) Core Observatory, shown in Figure 1. It incorporates a Ku-band precipitation radar (KuPR) and a Ka-band precipitation radar (KaPR). The KuPR, an enhanced version of the Precipitation Radar utilized in NASA's Tropical Rainfall Measuring Mission (TRMM), offers heightened sensitivity for measuring light rainfall and snowfall in mid-latitude areas [2] . The focus was to build upon the intriguing insights offered by the Precipitation Radar (PR) of the Tropical Rainfall Measuring Mission (TRMM) satellite, which captured the three-dimensional structure of highly precipitating clouds in the tropics and subtropics from 1997 to 2015 [3], but also spatial variability of rainfall intensity and distribution both globally and regionally [4][5][6].

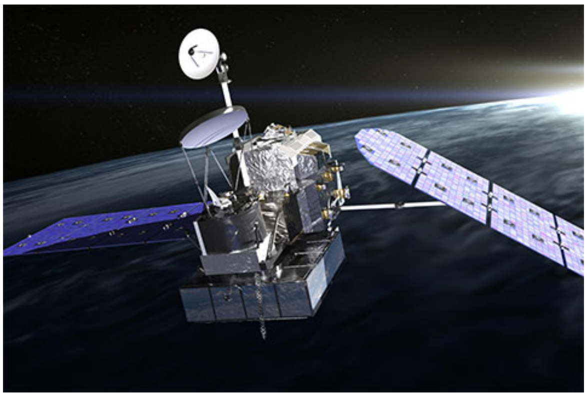


Figure 1: Global Precipitation Measurement (GPM) core satellite [1]

Now, building on these studies, the GPM-DPR database is used to identify and categorize storms to then uncover new insights into their structural and geographical variability.

The main goals of this analysis are:

- Use the new GPM-DPR data to uncover new insights into storm structures;
- Develop a database summarizing the precipitation events captured by GPM-DPR;
- Do an exploratory data analysis of the spatial distribution and variability of precipitation features;
- Analyze how the new database features vary on a global and seasonal scale;
- Look into the relation between the 2D spatial structure features and the Vertical information of precipitation.

In order to reach these goals the study searched into the spatial characteristics of storms, detected by radars through near-surface precipitation and reflectivity. Therefore, from the GPM DPR database, using near-surface precipitation data, the precipitation clouds were labeled before and then organized in patches, each one identifying one storm structure. For each storm patch (illustrated in Image 2), extensive statistics were extracted, that included spatial, vertical and structural features. This led to the creation of a new GPM-STORM database (STORMDB).

The newly derived database was then analyzed through self-organizing maps (SOM), aiming to gain a comprehensive understanding of global storm structure variability. The primary objectives were to discern potential correlations between spatial and vertical features while trying to classify recurring thunderstorm formations such as Cyclones, Mesoscale Convective Systems (MCS), which are structures made by converging multiple thunderstorms, and Squall Line, that is a group of storms arranged in a line that can reach hundreds of miles [7] based on their structural features.

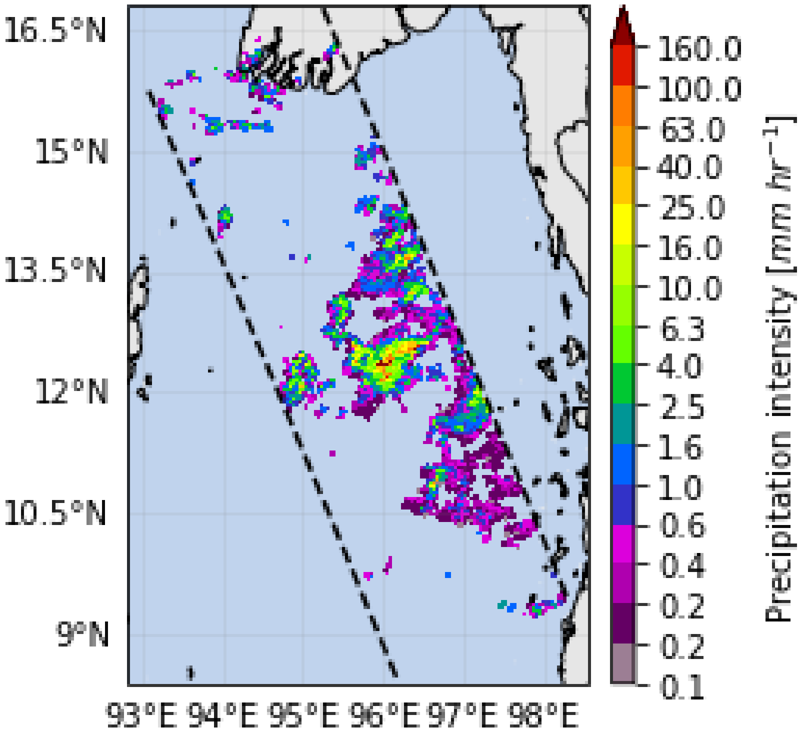


Figure 2: Example of extracted patch.

To this end, the analysis was focused on high-intensity precipitation events. Specifically, the patches were categorized as high-intensity precipitation events if they exhibited spatially structured near-surface precipitation intensities surpassing 10 mm/h.

2 Methods

In this section, the methodological approach is described. Figure 3 illustrates the project's structure. Starting from the GPM-DPR database, patches were extracted passing from precipitation pixels and labels, and then the GPM-STORM Database was created uniting the data from the GPM-DPR database and the patch identification. Following this, exploratory data analysis was performed on the STORMDB. A Self-Organizing Map (SOM) was then trained utilizing selected variables, and the obtained results were subsequently analyzed.

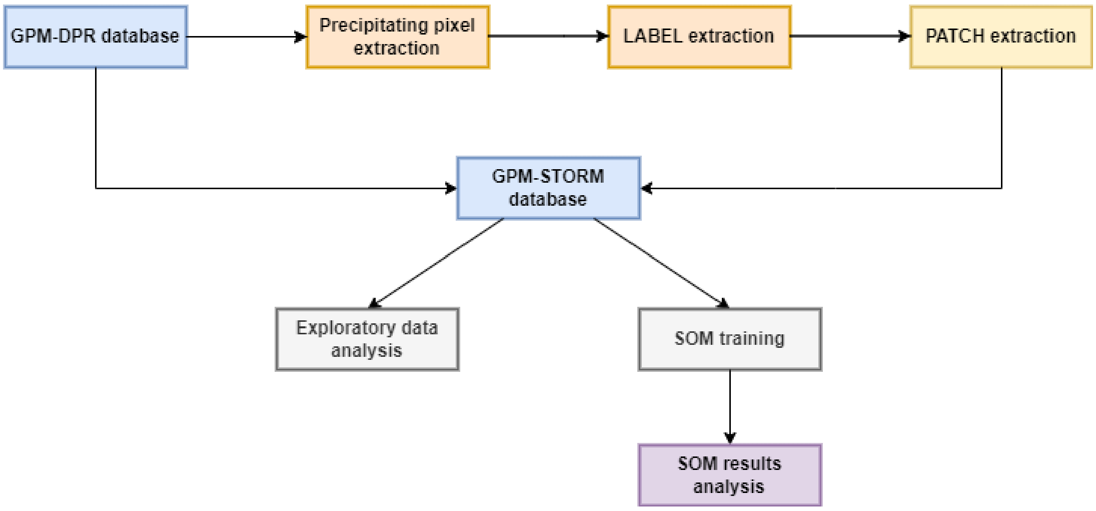


Figure 3: Flow chart.

2.1 GPM-STORM Database

STORMDB is made of 88 variables, divided into spatial, vertical, and structural features of the storms. Also, ID features are included to recover the storms in time and space.

2.1.1 Storm Identification

The data used to extract labels and patches are structured in granules that have a cross-track length of 49 pixels, equivalent to 245 km, and an along-track length of 20573 pixels which is equivalent to 102865 km. Considering this the minimum extension of the patches was set to 49 considering the cross-track dimension to avoid variability on this dimension and be able to then later identify storms that could have been cut. Regarding the along-track dimension, the minimum length was also set at 49 pixels, to be consistent with the previous length choice. However, this setting allows for expansion to larger sizes, enabling the adaptation to bigger storms when needed.

The single storms have been identified as patches. The patches have been created from the labeling of the storms based on the storms' near-surface precipitation. The threshold was set to be 0.05 mm/h to identify precipitating clouds, therefore all the pixels with values above the threshold were considered part of the precipitation clouds. The threshold was set to a low value to be as comprehensive as possible when analyzing the characteristics of the storms in a second moment. The possible negative outcome of this choice is that noise could be detected as part of the storm, therefore this was taken into account when analyzing the data. Then each label, identifying a storm, was created setting a minimum area of 49*49 pixels, and the precipitation clouds were defined as a part of a single storm if they were not separated by more than 25 km (5km * footprint) as part of the storm. Eventually, the patches are built based on the labeled storm and consider also what is around the labels to then incorporate what was mistakenly considered as a different storm in the first place when labeling.

2.1.2 Feature Extraction

Between 2018 and 2023, a total of 1,483,827 patches were identified. For each patch, 88 features were computed, and they are detailed in Table 2. These variables can be categorized into four groups.

The "Spatial" variables statistically describe the extent and intensity of near-surface precipitation. For example, features such as "*Precipitation Average*" (the mean precipitation in the considered patch) and "*Precipitation Area*" (counting pixels with precipitation for each patch) enable to partially separate between storms with high-intensity precipitation and those with extensive coverage.

Next are the "Structure" variables, which aim to provide descriptive features of storm organization. For instance, "*Aspect Ratio*" is defined as the aspect ratio of an ellipse fitted to the largest connected cell in the patch area. "*Count Rainy Areas Over Threshold*" indicates the number of cells in the patch exceeding a specified precipitation intensity threshold. These variables assist in identifying storm structures, like Squall Lines with elongated shapes or equatorial thunderstorm formations characterized by disconnected cells, hence a higher number of areas.

Moving on to "Vertical" variables, these are based on reflectivity and include REFCH, EchoDepth, and Echotopheight. REFCH, or "Reflectivity at the Echo Top", is a meteorological parameter measured in meters (m), representing the height where maximum reflectivity is observed in the column. EchoDepth represents the vertical extent of a radar echo with reflectivity specified above a certain threshold, while EchoTopHeight indicates the height of maximum height where the radar echo can be detected when reflectivity is above a threshold. These terms are employed in meteorology to characterize the vertical structure of convective weather systems, such as thunderstorms. Understanding EchoDepth and EchoTopHeight is essential for evaluating the intensity and organization of convective storms [8].

Finally, the ID variables, determined the position of the patch in space and time, to be able to locate the patch in space and time but also to recover it from the GPM-DPR database.

2.2 Self-Organizing-Maps

Self-Organizing Maps (SOMs), developed by Teuvo Kohonen in the 1980s [9], are a type of artificial neural network designed for unsupervised learning and dimensionality reduction. SOMs serve as a powerful tool for organizing and visualizing high-dimensional data in a more accessible manner.

The SOM comprises a grid of neurons, each representing a potential cluster for the input data samples. After feeding SOMs with the defined input data, they go through a training process where neurons with weight vectors most similar to the input vectors are selected as winners. The weights of these winning neurons, along with their neighbors, are then adjusted to align more closely with the input data.

One notable feature of SOMs is their topology-preserving property [10], ensuring that nearby neurons respond to similar input vectors. This intrinsic characteristic aids in maintaining spatial relationships between data points. Once trained, the SOM provides a condensed, lower-dimensional representation of the input data, facilitating effective clustering, and data visualization. These characteristics of SOMs align with the objectives of the project, in which Somoclu was used. Somoclu is a parallel tool for training self-organizing maps on large data sets [11].

The efficiency of SOMs becomes evident in their ability to exhibit similar structures side by side. This is particularly valuable when dealing with dynamic structures that undergo constant changes, making it challenging to achieve a clear-cut differentiation. The SOM's capability to preserve the spatial relationships between similar structures facilitates a more insightful and coherent representation, offering a deep understanding of the intricate patterns emerging from the multitude of variables.

An in-depth analysis was done on what can be referred to as the "shape SOM". The SOM was trained using "Surface" and "Structural" Variables (shown in detail in Table 2), all derived from Near-surface precipitation data. The SOM effectively organized patches based on these training variables. First, the SOM was initialized setting the size of the matrix as 10 times 10 (100 neurons per node). The grid type and map type were respectively chosen to be 'rectangular' and 'planar' as these parameters are suitable for our application and assure a regular and visually interpretable map. Once initialized, the training was done setting the epochs to 100 as it was considered enough for the model to converge without risking overfitting. The radius and scale parameters, shown in table 1, balance the need to capture global patterns in the early stages and to train and refine the map with smaller, more local adjustments as training progresses.

Table 1: SOM Parameters

radius0	**radiusN**	**scale0**	**scaleN**
0	1	0.5	0.001

The training was conducted incorporating almost all "Surface" and "Structure" variables. Only variables that were created with statistics using precipitation threshold over 20 mm/h were excluded due to their high selectivity. To better understand why they were excluded we must consider that during SOM training, patches containing NaN (Not a Number) values in any of the training set variables were omitted, as the algorithm cannot process such values. Therefore, using variables such as *"Minor Axis Largest Patch Over 20"* would lead to only considering less than 1% of the extracted patches. In our analysis, for instance, the variable "*Minor Axis Largest Patch Over 10*" was the most selective because only 5.4% of patches had enough organized values over 10 mm/h to fit an ellipse. Consequently, the classification focused exclusively on this subset of high-intensity storms.

3 Results and Developments

3.1 GPM-STORM

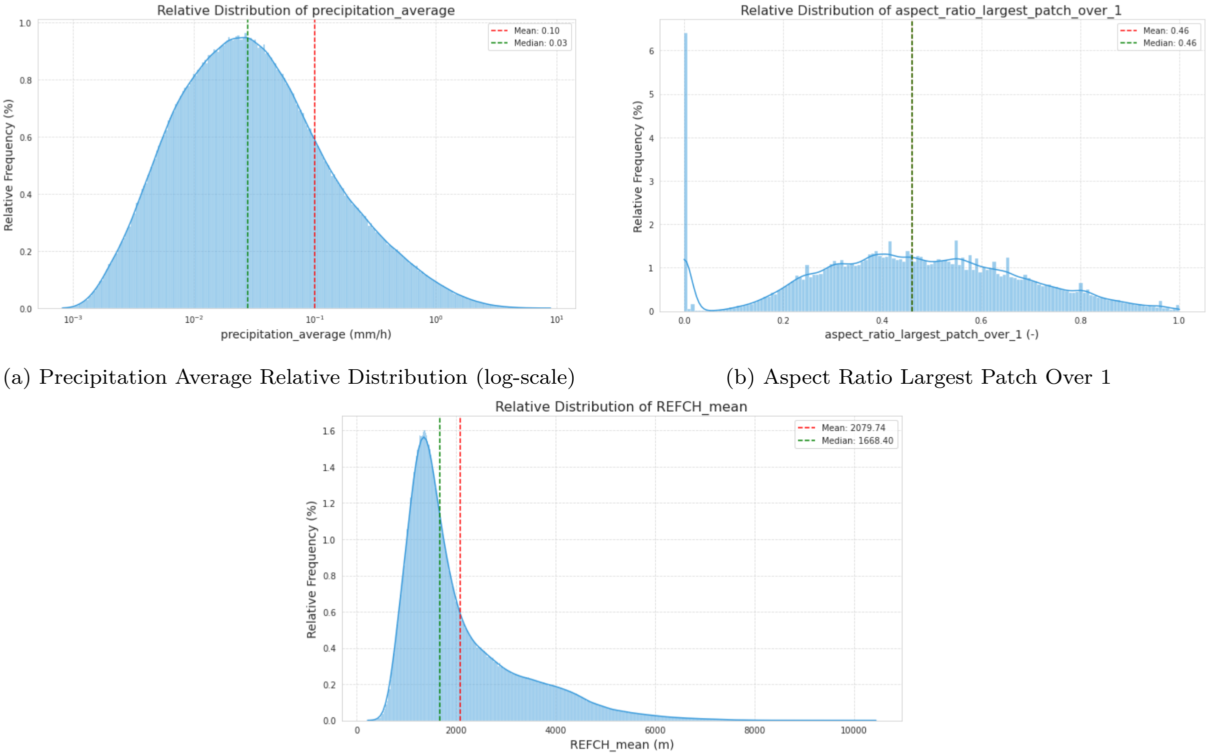
3.1.1 Features Distribution

The STORMDB, derived from patches extracted from GPM-DPR data, comprises a total of 88 variables, as previously mentioned. In our analysis, we began by examining the relative distributions of these variables, focusing on showcasing exemplary distributions to illustrate key findings. Notably, a subset of the "Surface" variables exhibited strongly positively skewed distributions, although none of them conformed to the hypothesis of a log-normal distribution. Here, three selected variable distributions, are illustrated as representative of their group.

One representative distribution is that of "*Precipitation Average*" (Figure 4a) which is a positively skewed distribution, which represents the trend observed in most "Surface" variables. In such distributions, the median possesses a considerably smaller value than the mean.

On the other hand, the distribution of "*Aspect Ratio Largest Patch Over 1*" (Figure 4b) is significantly different from the previous one. It displays a peak for very low values, followed by a distribution resembling normality. This peculiar shape raises suspicions about the accuracy of the initial peak values and prompts further investigation.

Finally, we examined the distribution of "*REFCH Mean*" (Figure 4c), representing "Vertical" features. Despite also showcasing a positively skewed distribution, it is notably less skewed than the "Surface" variables. This observation underscores the distinct nature of "Vertical" features in contrast to their "Surface" counterparts.



(a) Precipitation Average Relative Distribution (log-scale)

(b) Aspect Ratio Largest Patch Over 1

(c) REFCH Mean Relative distribution

Figure 4: Distribution Plots

3.1.2 Variables Significance

The chosen variables, elaborated in Table 2, played a significant role in training the SOM to recognize structures, geographical locations, and the seasonality of storms. However, several considerations arise from both the variable distributions and boxplots.

Concerning the "Surface" variables, those without thresholds are notably susceptible to noise, leading to distributions that are heavily compressed towards lower values, exemplified by the Precipitation Average Relative Distribution in Figure 4a. This compression poses challenges when training the SOM using the entire STORMDB, potentially causing subtle distinctions in high-intensity storms to be overlooked. Consequently, the selected SOM was tailored to exclusively identify high-intensity storms.

On the other hand, variables associated with EchoTopHeight and REFCH are profoundly influenced by topography, particularly by the height of the radar beam above the ground [12]. However, in this case, the influence primarily stems from the measurements themselves, which are impacted by the proximity of elevated areas. This dynamic introduces a potential source of confusion when training the SOM using these variables, as regional patterns may be more attributed to the height above ground rather than the intrinsic characteristics of the storms.

Considering these two factors, we can then exclude the influence of these biases on the SOM that was considered here for the analysis, since only high-intensity patches were considered and vertical features were not used for training.

3.1.3 Global Statistics

Spatial distributions of various variables were examined to gain insights into their geographical patterns. Of particular interest are two key variables, namely "*Precipitation Max*" and "*REFCH Mean*".

The spatial distribution of the feature "*Precipitation Max*", illustrated in Figure 5, is depicted by calculating the average intensity distribution of maximum precipitation in every 1° time 1° geographical bin. The visualization highlights regions that are particularly prone to high-intensity precipitation, such as the oceanic zones surrounding the equator, especially just above it, and the areas covered by dense rainforests.

Turning attention to "*EchoTopHeight30 Mean*", as presented in Figure 6, this represents the mean binned values for EchoTopHeight with a set threshold of 30 dbZ. Analyzing this spatial distribution reveals essential topographical information associated with "Vertical" variables. This validates the quality of the data observed as it was expected from this variable to be strongly influenced by topography. It is also consistent with the distribution results of previous studies [13]. Notably, areas corresponding to significant mountain ranges, such as the Himalayas or the Andes, exhibit markedly higher mean heights. This observation underscores the influence of topography and latitude on the vertical characteristics captured by the database.

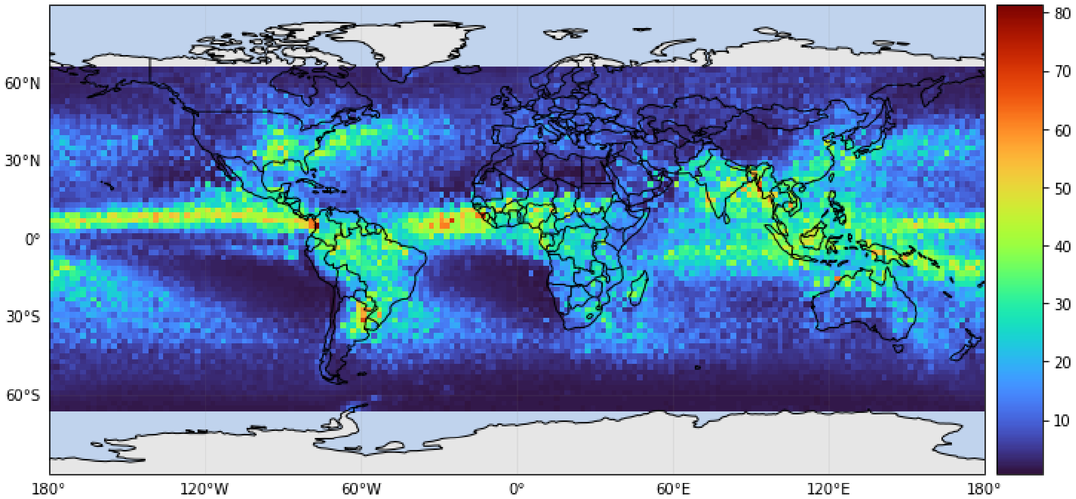


Figure 5: Precipitation Max (unit mm/h) global distribution calculated as the mean for each single bin.

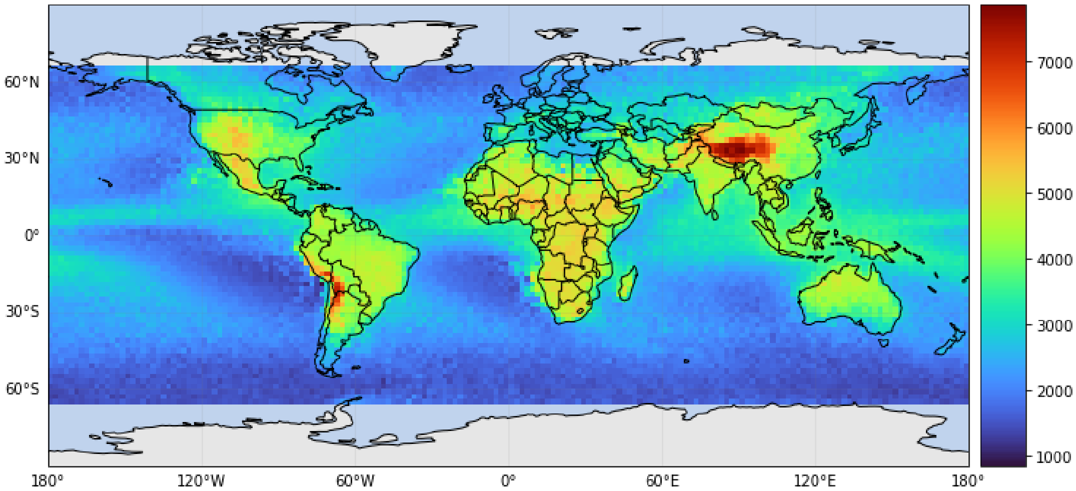


Figure 6: EchoTopHeight30 Mean (unit m) global distribution calculated as the mean for each single bin.

3.2 Storm Structures

Post-training, a random patch was extrapolated and plotted for each node, providing a visual representation of the SOM's learned patterns. Additionally, statistical analysis was performed for each node, involving the calculation of the mean for each variable within the subset of patches belonging to that node. This analysis aimed to provide insights into the characteristics of each node.

To further understand the SOM's representation, 25 random exemplary patches were extracted for each node. Subsequently, all patches belonging to a particular node were geographically located on a world map. The color coding of these patches was based on the season, offering a visual summary of the spatial distribution of high-intensity storms across different seasons. The seasons are the meteorological seasons for the Northern Emisphere, therefore Winter is December, January and February, Spring is March, April and May, Summer is June, July and August, and Fall is September, October and November. Eventually, another SOM was trained with the same parameter to discuss the intrinsic stochasticity of this method.

3.2.1 Storm Organization

In this section, we unveil the diversity of storm spatial patterns captured by SOM training. Figure 7 shows a sample patch that has been assigned to a given node of the SOM. The randomly plotted patches, as seen in Figure 7, are organized in terms of structure and intensity. In the lower part, the precipitation is more sparse and less intense, as confirmed by the statistics on Precipitation Average, as seen in Figure 12a. Having a closer look at samples from each node, it is possible to distinguish similar precipitation structures recurring in each node. In the last column of the SOM, it is possible to notice that elongated storm structures are recurrent and even tough, it is not possible to precisely classify them as Squall Lines, they present similar patterns. As shown in Figures 15 and 16, it is evident that the structures in the nodes are similar, therefore the classification was effective. Moreover, since they represent two confining nodes, this also shows how the structures are similar for close nodes, as it was expected.

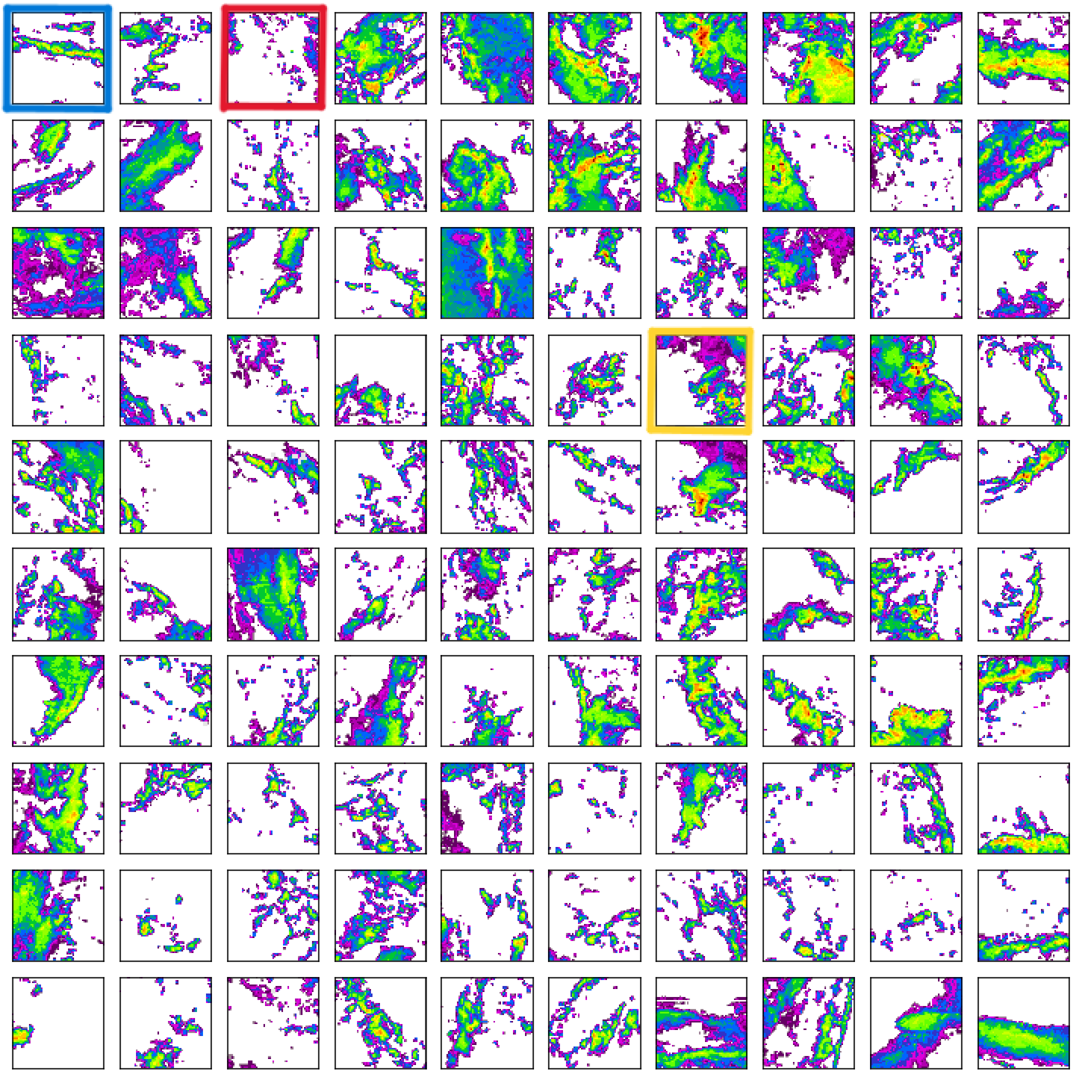
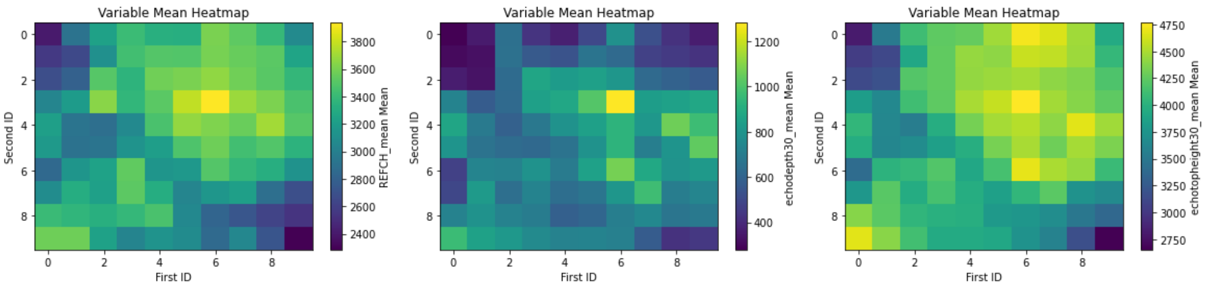


Figure 7: SOM grid samples, with interesting nodes highlighted in blue (1st row, 1st column), red (1st row, 3rd column) and yellow (4th row, 7th column).

3.2.2 From Precipitation to Reflectivity

Upon closer examination of the data, it becomes evident that the "Vertical" variables underwent reorganization within the SOM. Figure 8 illustrates the mean values in each node of three representative "Vertical" variables. It can be seen that even though these variables were not part of the training subset of the SOM they present specific patterns in the SOM. Interestingly, the variability between the nodes presents similarities for the nodes next to each other, showing that even for these variables the relation with the training set was strong enough to preserve the topography-preserving feature of the SOMs. Some specific nodes have close characteristics in all the "vertical" variables. For example, the top left and bottom right corners of the SOM display low mean values for all concerned vertical variables, indicating a reciprocal relationship.

On the other hand, establishing a direct link between a single variable and the distribution of "Vertical" variables is challenging. For example, even though, it is already known the link between strong reflectivity and intense precipitation [14], the strict correlation between them is weak [15]. This is confirmed considering the mean values distribution of the nodes for "Vertical" features compared to the mean values distribution of "*Precipitation Average*", shown in Figure 12a. This can be explained, when considering REFCH and EchoTopHeight, by the fact that other important factors play an important role, as they are hugely influenced by topography and latitude, as previously seen also through their global distribution. On the other hand, the good reorganizations of storms based on "Vertical" features that were not included in the training of the SOM, uncover a direct relationship between recurrent structures and EchoTopHeight, EchoDepth and REFCH values. Notably, Echodepth exhibits a distinct peak in the 4th row and 7th column (yellow square in Figure 7), mirroring similar peaks observed for REFCH mean and Echotopheight. This structure will be analyzed more in-depth in Section 3.2.3, where the node presenting the highest peak in reflectivity will be analyzed. Moreover, EchoTopHeight exhibits multiple peaks, suggesting potentially stronger influences from topographical features when compared to the other "Vertical" variables.



(a) REFCH Mean (b) Echodepth30 Mean (c) Echotopheight30 Mean

Figure 8: Mean for each node of the selected variables.

3.2.3 Spatial and Seasonal Distributions

The geographical distribution of patches across nodes revealed distinct patterns, indicating that storms with a given spatial structure are characteristic only of specific geographic regions. Illustrated in Figure 9, nodes associated with high-intensity, high-reflectivity storms predominantly occur in the equatorial area. Notably, these storms exhibit a seasonality prevalence in Summer and Fall for the Northern Hemisphere and Winter and Spring for the Southern Hemisphere.

In contrast, the corner on the upper left side of the SOM, depicted in Figure 10, demonstrates an inverse seasonality pattern, with a prevalence in the mid-upper latitudes of the oceanic regions. This suggests a distinct storm pattern compared to the equatorial storms.

Furthermore, Figure 11 reveals a unique storm pattern prevalent in the Amazon Rainforest, with recurrent occurrences in the Central West Pacific and Southeast Asia. This distinctive distribution highlights the localized nature of storms in specific regions, showcasing the versatility and regional variations in storm characteristics captured by the SOM.

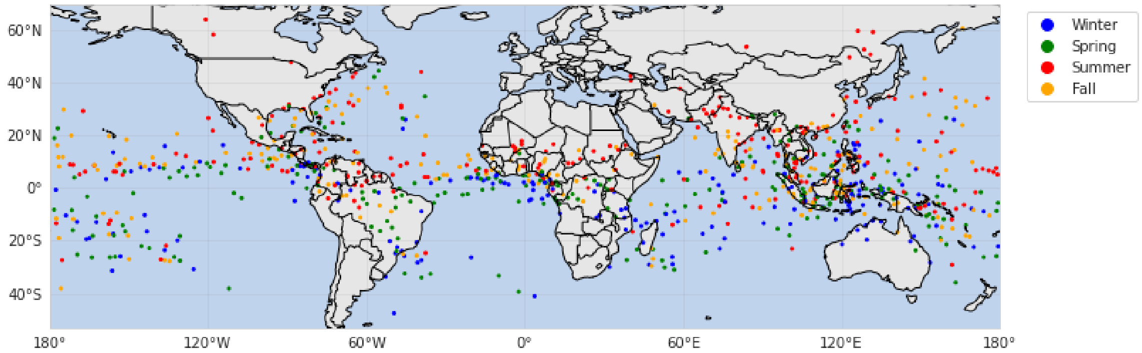


Figure 9: Node 4th row, 7th column (yellow square) patch global distribution, color-coded by season.

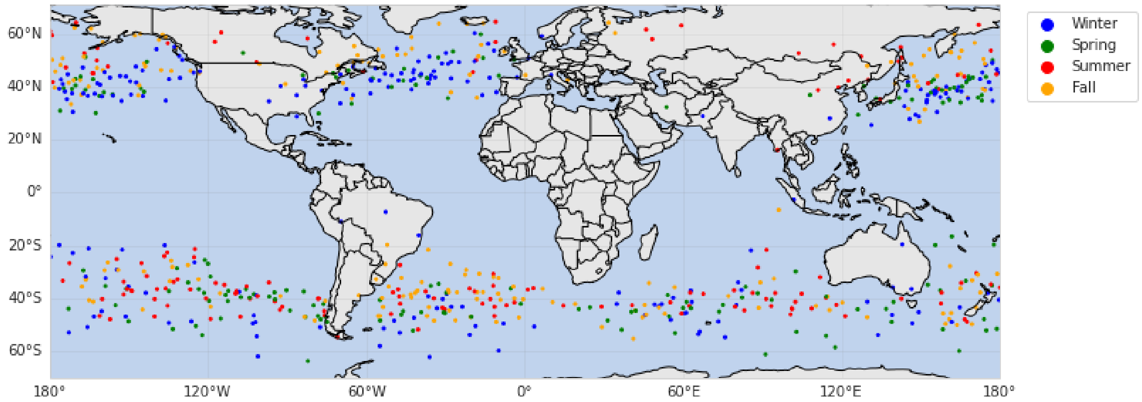


Figure 10: Node 1st row, 1st column (blue square) patch global distribution, color-coded by season.

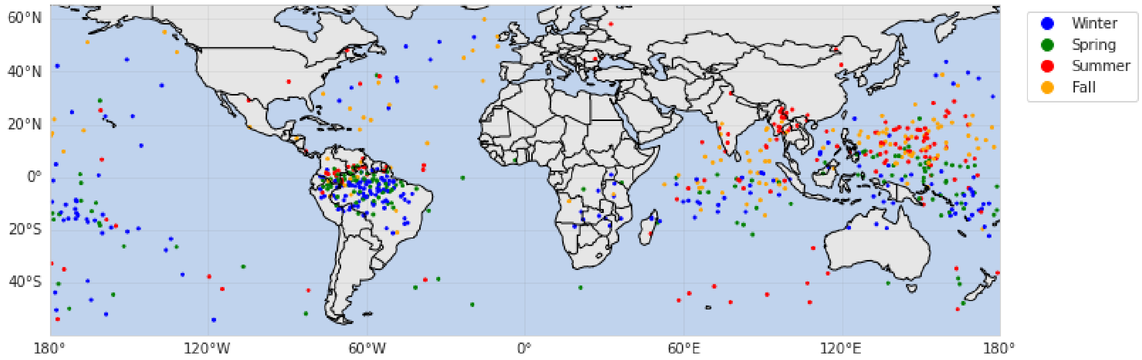


Figure 11: Node 1st row, 3rd column (red square) patch global distribution, color-coded by season.

This emphasizes the potential to classify unique storms that form specifically in certain geographical areas. While the classification didn't precisely isolate storms to individual regions, it did showcase latitude differences, as evident in Figure 10, and specific regional occurrences, as seen in Figure 11.

The non-random nature of this differentiation is further supported by the observed seasonality in these occurrences. For instance, in Figure 11, it's evident that in the Amazon, these storms are prevalent during winter, while in India, Myanmar, and the Central West Pacific Ocean, they occur predominantly in Fall and Summer. The recurring appearance of the same structure in the same area during the same season suggests that the node identified a distinct type of storm.

Taking a closer look at samples from this specific node (Figure 13), it becomes apparent that these storms exhibit a multi-cell formation with locally intense peaks in precipitation. Interestingly, they also feature medium average values of Echodepth and Echotopheight.

On the other hand, when delving into a specific subset characterized by high values for the "Vertical" variables, particularly the node depicted in Figure 9, a consistent pattern emerges. This pattern is observable across most nodes and it shows that there is a notable prevalence of occurrences in the equatorial region, with sporadic instances also at medium latitudes. The SOM adeptly distinguishes between storms in medium-high latitudes and those in equatorial regions. While some nodes exhibit specific regional occurrences, the majority primarily differentiate based on latitude.

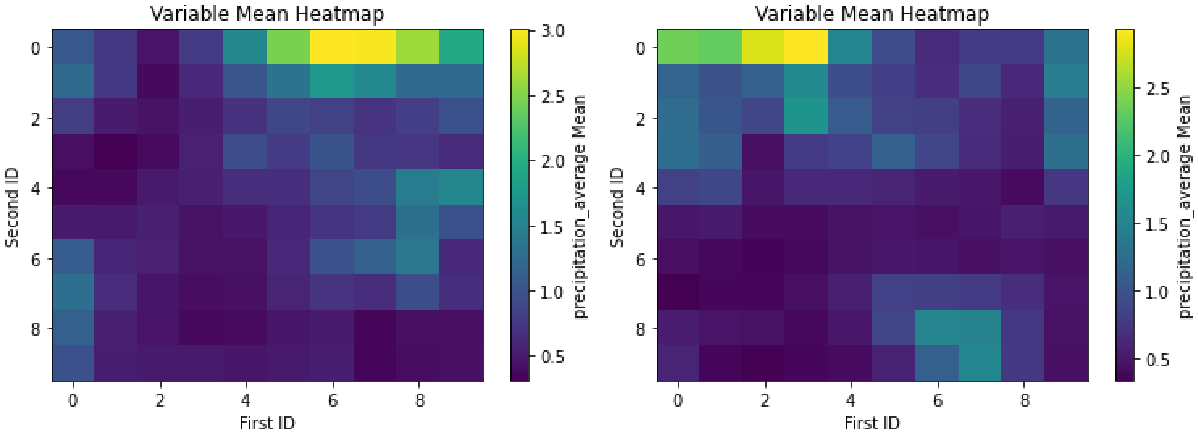
Moreover, most nodes exhibit consistent seasonal variation, alternating between the northern and southern hemispheres. This enhances the analysis's potential to categorize and classify specific storm formations, opening avenues for more in-depth investigations. For instance, increasing the grid size could significantly impact the identification of more regionally specific storm formations in the nodes. This suggests a promising direction for future research and refinement of the analysis.

3.2.4 SOMs' Stochasticity

Self-Organizing Maps (SOMs) exhibit inherent stochasticity despite their deterministic learning algorithms. The stochastic nature arises from the randomized initialization of weights, influencing the network's response to the input data during training. This randomness introduces variability in the convergence process, leading to different outcomes when the same dataset is processed multiple times. The influence of stochasticity becomes particularly evident in the formation of map structures, with slight variations observed in the final topology.

In this project, two SOMs were trained using identical parameters. Figure 12 illustrates that when comparing the node mean precipitation average, the SOMs display different structures. It is crucial to note, however, that while the structures differ, they exhibit similar patterns. The scale and peak similarities in both SOMs suggest a consistent representation of the data.

Upon closer examination of the second trained SOM's structure, the recognition of seasonality and regionality in the nodes becomes evident, as it was already seen and discussed for the previous SOM. Figures 17 and 14 showcase the division of different structures in nodes, highlighting interesting patterns in specific nodes.



(a) Mean of Precipitation Average for first trained SOM.

(b) Mean of Precipitation Average for second trained SOM.

Figure 12: Mean of Precipitation Average for each patch of each node (unit mm/h).

4 Conclusion

This project has set the basis to further understand precipitation patterns and their spatial variability by employing the Global Precipitation Measurement (GPM) Dual-frequency Precipitation Radar (DPR) database from 2018 to 2023.

STORMDB was successfully created, and the variables were confirmed to be significant. When examining the global distribution patterns of the variables, it was observed that they aligned with anticipated outcomes, consistent with findings from prior studies or as expected based on the inherent characteristics of the variables. On the other hand, some variables presented incoherences, that would need further investigations. . By focusing on the spatial characteristics of storms and employing self-organizing maps (SOMs) to structure our data, the project revealed an intrinsic relationship between "Spatial" features and "Vertical" features. The SOM analysis also showcased the ability to organize storms based on their structure, even though it is not able to precisely distinguish between specific storm formations just by using a single SOM. On the other hand, it was possible to capture the variability between nodes in structure, intensity, and seasonality shaped by the SOM. The nodes effectively categorized storms based on their geographical locations and exhibited consistent seasonal variations.

In conclusion, the results obtained by this first analysis are already showing interesting insights into the possible recognition of specific high-intensity storm structures. However, it must be considered that this project will function as the basis for further analysis that could be done starting from the STORMDB, which could lead to a future better understanding of these phenomena.

4.1 Next Steps

The STORMDB provides a solid foundation for further refinement and enhancement of storm classification. The next first step should be expanding the database with additional parameters, such as cloud temperature and the presence of hail. This would contribute to a more comprehensive understanding of the physical processes underlying different storm types. These additions could facilitate more precise classifications, enabling the identification of specific cloud formations and distinguishing between various meteorological phenomena.

Additionally, by taking a closer look at the global distribution of storms in the node it was clear how the SOM could still not classify the storms to more specific world regions. It could be helpful to use bigger grids when training the SOM to possibly find more detailed storm structure separation. This could help identify distinct regional storm formations more accurately, giving better insights into the specific geographical characteristics of storms.

In summary, expanding the database with additional parameters, and conducting more detailed regional analyses would undoubtedly strengthen the project's outcomes.

5 Finding Data

STORMDB can be found in the LTE server number 8 (ltesrv8). The path to follow is "home\comi\Projects\dataframe2.parquet", it is referred to as dataframe2 as it is an uploaded version of the first STORMDB. All the images for each node of the trained SOM can be found on the GitHub GPM Storm

References

[1] *JAXA | Global Precipitation Measurement/Dual-frequency Precipitation Radar (GPM/DPR).* https://global.jaxa.jp/projects/sat/gpm/ (cit. on p. 2).

[2] NASA Earth Science Data Systems. *DPR | Earthdata.* en. Publisher: Earth Science Data Systems, NASA. Apr. 2022. https://www.earthdata.nasa.gov/sensors/dpr (cit. on p. 2).

[3] Chuntao Liu, Edward J. Zipser, and Stephen W. Nesbitt. "Global Distribution of Tropical Deep Convection: Different Perspectives from TRMM Infrared and Radar Data". EN. In: *Journal of Climate* 20.3 (Feb. 2007). Publisher: American Meteorological Society Section: Journal of Climate, pp. 489–503. ISSN: 0894-8755, 1520-0442. DOI: 10.1175/JCLI4023.1. https://journals.ametsoc.org/view/journals/clim/20/3/jcli4023.1.xml (cit. on p. 2).

[4] Chuntao Liu et al. "A Cloud and Precipitation Feature Database from Nine Years of TRMM Observations". EN. In: *Journal of Applied Meteorology and Climatology* 47.10 (Oct. 2008). Publisher: American Meteorological Society Section: Journal of Applied Meteorology and Climatology, pp. 2712–2728. ISSN: 1558-8424, 1558-8432. DOI: 10.1175/2008JAMC1890.1. https://journals.ametsoc.org/view/journals/apme/47/10/2008jamc1890.1.xml (cit. on p. 2).

[5] Yunfei Fu et al. "Climatological characteristics of summer precipitation over East Asia measured by TRMM PR: A review". en. In: *Journal of Meteorological Research* 31.1 (Feb. 2017), pp. 142–159. ISSN: 2198-0934. DOI: 10.1007/s13351-017-6156-9. https://doi.org/10.1007/s13351-017-6156-9 (cit. on p. 2).

[6] Nan Li et al. "Studies of General Precipitation Features with TRMM PR Data: An Extensive Overview". en. In: *Remote Sensing* 11.1 (Jan. 2019). Number: 1 Publisher: Multidisciplinary Digital Publishing Institute, p. 80. ISSN: 2072-4292. DOI: 10.3390/rs11010080. https://www.mdpi.com/2072-4292/11/1/80 (cit. on p. 2).

[7] NOAA. *Thunderstorm Types.* EN-US. text. https://www.nssl.noaa.gov/education/svrwx101/thunderstorms/types/ (cit. on p. 3).

[8] Charlotte A. DeMott and Steven A. Rutledge. "The Vertical Structure of TOGA COARE Convection. Part I: Radar Echo Distributions". en. In: *Journal of the Atmospheric Sciences* 55.17 (Sept. 1998), pp. 2730–2747. ISSN: 0022-4928, 1520-0469. DOI: 10.1175/1520-0469(1998)055<2730:TVSOTC>2.0.CO;2. https://journals.ametsoc.org/doi/10.1175/1520-0469(1998)055%3C2730:TVSOTC%3E2.0.CO;2 (cit. on p. 5).

[9] T. Kohonen. "The self-organizing map". In: *Proceedings of the IEEE* 78.9 (Sept. 1990). Conference Name: Proceedings of the IEEE, pp. 1464–1480. ISSN: 1558-2256. DOI: 10.1109/5.58325. https://ieeexplore.ieee.org/abstract/document/58325?casa_token=MeAOk0caY9MAAAAA:Pvjjv5avUA1ZdKrgb5gSXayu3j7COGHOiEfyhJFYYI7RZqLRg7G2YXVS1jC5oVJC7tJhwfRexcs (cit. on p. 5).

[10] Balazs Feil and János Abonyi. "Introduction to Fuzzy Data Mining Methods". In: Jan. 2008, pp. 55–95. DOI: 10.4018/978-1-59904-853-6.ch003 (cit. on p. 5).

[11] Peter Wittek et al. "Somoclu: An Efficient Parallel Library for Self-Organizing Maps". In: *Journal of Statistical Software* 78.9 (2017). arXiv:1305.1422 [cs]. ISSN: 1548-7660. DOI: 10.18637/jss.v078.i09. http://arxiv.org/abs/1305.1422 (cit. on p. 5).

[12] Tim Marshall. "TOPOGRAPHIC INFLUENCES ON AMARILLO RADAR ECHO CLIMATOLOGY". In: (1980). https://www.academia.edu/34321624/TOPOGRAPHIC_INFLUENCES_ON_AMARILLO_RADAR_ECHO_CLIMATOLOGY (cit. on p. 8).

[13] Lei Ji et al. "Consistency of Vertical Reflectivity Profiles and Echo-Top Heights between Spaceborne Radars Onboard TRMM and GPM". en. In: *Remote Sensing* 14.9 (Jan. 2022). Number: 9 Publisher: Multidisciplinary Digital Publishing Institute, p. 1987. ISSN: 2072-4292. DOI: 10.3390/rs14091987. https://www.mdpi.com/2072-4292/14/9/1987 (cit. on p. 8).

[14] Jian Zhang et al. "Multi-Radar Multi-Sensor (MRMS) Quantitative Precipitation Estimation: Initial Operating Capabilities". en. In: *Bulletin of the American Meteorological Society* 97.4 (Apr. 2016), pp. 621–638. ISSN: 0003-0007, 1520-0477. DOI: 10.1175/BAMS-D-14-00174.1. https://journals.ametsoc.org/doi/10.1175/BAMS-D-14-00174.1 (cit. on p. 11).

[15] Tsechun Wang and Guoqiang Tang. "Spatial Variability and Linkage Between Extreme Convections and Extreme Precipitation Revealed by 22-Year Space-Borne Precipitation Radar Data". en. In: *Geophysical Research Letters* 47.12 (2020). _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1029/2020GL088437, e2020GL088437. ISSN: 1944-8007. DOI: 10.1029/2020GL088437. https://onlinelibrary.wiley.com/doi/abs/10.1029/2020GL088437 (cit. on p. 11).

A Appendix

A.1 Dataset Variables

Tab available from next page.

Table 2: Variables of the Dataset, in bold the variables used to train the SOM.

Variable	Group	Units	Explanation
Precipitation Average	Spatial	mm/h	Average of near-surface precipitation of all the pixels in the patch.
Precipitation STD	Spatial	mm/h	Standard Deviation of all the pixels in the patch
Precipitation Area	Spatial	pixels	Number of Pixel with near-surface precipitation over zero
Precipitation Pixel Center	Spatial	-	Number of Pixel in the central part of the patch.
Precipitation Sum	Spatial	pixels	Sum of all the near-surface precipitation value of the patch
Precipitation Max	Spatial	mm/h	Max near surface precipitation value
Count Rainy Areas Over 0	Structure	pixels	Number of connected precipitating areas when the threshold is 0 mm/h
Count Rainy Areas Over 1	Structure	pixels	Number of connected precipitating areas when the threshold is 1 mm/h
Count Rainy Areas Over 2	Structure	pixels	Number of connected precipitating areas when the threshold is 2 mm/h
Count Rainy Areas Over 5	Structure	pixels	Number of connected precipitating areas when the threshold is 5 mm/h
Count Rainy Areas Over 10	Structure	pixels	Number of connected precipitating areas when the threshold is 10 mm/h
Count Rainy Areas Over 20	Structure	pixels	Number of connected precipitating areas when the threshold is 20 mm/h
Count Rainy Areas Over 50	Structure	pixels	Number of connected precipitating areas when the threshold is 50 mm/h
Count Rainy Areas Over 80	Structure	pixels	Number of connected precipitating areas when the threshold is 80 mm/h
Count Rainy Areas Over 120	Structure	pixels	Number of connected precipitating areas when the threshold is 120 mm/h
Mean for Rainy Pixels Over 0	Spatial	mm/h	Average of only pixels with values of near-surface precipitation over 0 mm/h
Mean for Rainy Pixels Over 1	Spatial	mm/h	Average of only pixels with values of near-surface precipitation over 1 mm/h
Mean for Rainy Pixels Over 2	Spatial	mm/h	Average of only pixels with values of near-surface precipitation over 2 mm/h
Mean for Rainy Pixels Over 5	Spatial	mm/h	Average of only pixels with values of near-surface precipitation over 5 mm/h
Mean for Rainy Pixels Over 10	Spatial	mm/h	Average of only pixels with values of near-surface precipitation over 10 mm/h
Mean for Rainy Pixels Over 20	Spatial	mm/h	Average of only pixels with values of near-surface precipitation over 20 mm/h
Mean for Rainy Pixels Over 50	Spatial	mm/h	Average of only pixels with values of near-surface precipitation over 50 mm/h
Mean for Rainy Pixels Over 80	Spatial	mm/h	Average of only pixels with values of near-surface precipitation over 80 mm/h
Mean for Rainy Pixels Over 120	Spatial	mm/h	Average of only pixels with values of near-surface precipitation over 120 mm/h
Count Rainy Pixels Over 0	Spatial	pixels	Number of pixel over 0 mm/h
Count Rainy Pixels Over 1	Spatial	pixels	Number of pixel over 1 mm/h
Count Rainy Pixels Over 2	Spatial	pixels	Number of pixel over 2 mm/h

Continued on next page

Table 2 – Continued from previous page

Variable	**Group**	**Units**	**Explanation**
Count Rainy Pixels Over 5	Spatial	pixels	Number of pixel over 5 mm/h
Count Rainy Pixels Over 10	Spatial	pixels	Number of pixel over 10 mm/h
Count Rainy Pixels Over 20	Spatial	pixels	Number of pixel over 20 mm/h
Count Rainy Pixels Over 50	Spatial	pixels	Number of pixel over 50 mm/h
Count Rainy Pixels Over 80	Spatial	pixels	Number of pixel over 80 mm/h
Count Rainy Pixels Over 120	Spatial	pixels	Number of pixel over 120 mm/h
Major Axis Largest Patch Over 0	Structure	pixels	Major Axis of the ellipse fitted on the largest precipitating cloud over 0 mm/h
Major Axis Largest Patch Over 1	Structure	pixels	Major Axis of the ellipse fitted on the largest precipitating cloud over 1 mm/h
Major Axis Largest Patch Over 2	Structure	pixels	Major Axis of the ellipse fitted on the largest precipitating cloud over 2 mm/h
Major Axis Largest Patch Over 5	Structure	pixels	Major Axis of the ellipse fitted on the largest precipitating cloud over 5 mm/h
Major Axis Largest Patch Over 10	Structure	pixels	Major Axis of the ellipse fitted on the largest precipitating cloud over 10 mm/h
Major Axis Largest Patch Over 20	Structure	pixels	Major Axis of the ellipse fitted on the largest precipitating cloud over 20 mm/h
Minor Axis Largest Patch Over 0	Structure	pixels	Minor Axis of the ellipse fitted on the largest precipitating cloud over 0 mm/h
Minor Axis Largest Patch Over 1	Structure	pixels	Minor Axis of the ellipse fitted on the largest precipitating cloud over 0 mm/h
Minor Axis Largest Patch Over 2	Structure	pixels	Minor Axis of the ellipse fitted on the largest precipitating cloud over 0 mm/h
Minor Axis Largest Patch Over 5	Structure	pixels	Minor Axis of the ellipse fitted on the largest precipitating cloud over 0 mm/h
Minor Axis Largest Patch Over 10	Structure	pixels	Minor Axis of the ellipse fitted on the largest precipitating cloud over 0 mm/h
Minor Axis Largest Patch Over 20	Structure	pixels	Minor Axis of the ellipse fitted on the largest precipitating cloud over 0 mm/h
Aspect Ratio Largest Patch Over 0	Structure	-	Minor over Major Axis
Aspect Ratio Largest Patch Over 1	Structure	-	Minor over Major Axis
Aspect Ratio Largest Patch Over 2	Structure	-	Minor over Major Axis
Aspect Ratio Largest Patch Over 5	Structure	-	Minor over Major Axis
Aspect Ratio Largest Patch Over 10	Structure	-	Minor over Major Axis
Aspect Ratio Largest Patch Over 20	Structure	-	Minor over Major Axis
Count Rainy Pixels in Patch Over 0	Structure	pixels	Number of pixels in the largest patch with a threshold of 0 mm/h
Count Rainy Pixels in Patch Over 1	Structure	pixels	Number of pixels in the largest patch with a threshold of 1 mm/h
Count Rainy Pixels in Patch Over 2	Structure	pixels	Number of pixels in the largest patch with a threshold of 2 mm/h

Continued on next page

Table 2 – Continued from previous page

Variable	**Group**	**Units**	**Explanation**
Count Rainy Pixels in Patch Over 5	Structure	pixels	Number of pixels in the largest patch with a threshold of 5 mm/h
Count Rainy Pixels in Patch Over 10	Structure	pixels	Number of pixels in the largest patch with a threshold of 10 mm/h
Count Rainy Pixels in Patch Over 20	Structure	pixels	Number of pixels in the largest patch with a threshold of 20 mm/h
Percentage Rainy Pixels Between 0 and 1	Structure	%	
Percentage Rainy Pixels Between 1 and 2	Structure	%	
Percentage Rainy Pixels Between 2 and 5	Structure	%	
Percentage Rainy Pixels Between 5 and 10	Structure	%	
Percentage Rainy Pixels Between 10 and 20	Structure	%	
REFCH mean	Vertical	m	Mean value of REFCH in the patch
REFCH max	Vertical	m	Max value of REFCH in the patch
REFCH std	Vertical	m	Standard deviation of REFCH in the patch
EchoDepth18 mean	Vertical	m	Mean value of EchoDepth (threshold = 18dBz) in the patch
EchoDepth18 max	Vertical	m	Max value of EchoDepth (threshold = 18dBz) in the patch
EchoDepth18 std	Vertical	m	Standard deviation of EchoDepth (threshold = 18dBz) in the patch
EchoDepth30 mean	Vertical	m	Mean value of EchoDepth (threshold = 30dBz) in the patch
EchoDepth30 max	Vertical	m	Max value of EchoDepth (threshold = 30dBz) in the patch
EchoDepth30 std	Vertical	m	Standard deviation value of EchoDepth (threshold = 30dBz) in the patch
EchoDepth50 mean	Vertical	m	Mean value of EchoDepth (threshold = 50dBz) in the patch
Echodepth50 max	Vertical	m	Max value of EchoDepth (threshold = 50dBz) in the patch
Echodepth50 std	Vertical	m	Standard deviation value of EchoDepth (threshold = 50dBz) in the patch
Echotopheight30 mean	Vertical	m	Mean value of EchoTopHeight (threshold = 30dBz) in the patch
Echotopheight30 max	Vertical	m	Max value of EchoTopHeight (threshold = 30dBz) in the patch
Echotopheight30 std	Vertical	m	Standard Deviation value of EchoTopHeight (threshold = 30dBz) in the patch
Echotopheight50 mean	Vertical	m	Mean value of EchoTopHeight (threshold = 50dBz) in the patch
Echotopheight50 max	Vertical	m	Max value of EchoTopHeight (threshold = 50dBz) in the patch
Echotopheight50 std	Vertical	m	Standard deviation of EchoTopHeight (threshold = 30dBz) in the patch
Along Track Start	ID	-	Pixel in which the patch start in the granule

Continued on next page

Table 2 – Continued from previous page

Variable	**Group**	**Units**	**Explanation**
Along Track Stop	ID	-	Pixel in which the patch stop in the granule
Gpm Granule ID	ID	-	Granule ID number
Time	ID		time dd/mm/yy hh:mm
sunLocalTime	ID	-	sun local time
Longitude	ID	°	
Latitude	ID	°	
Flag Granule Change	ID	-	Is the patch at the beginning or end of the granule

A.2 Nodes samples

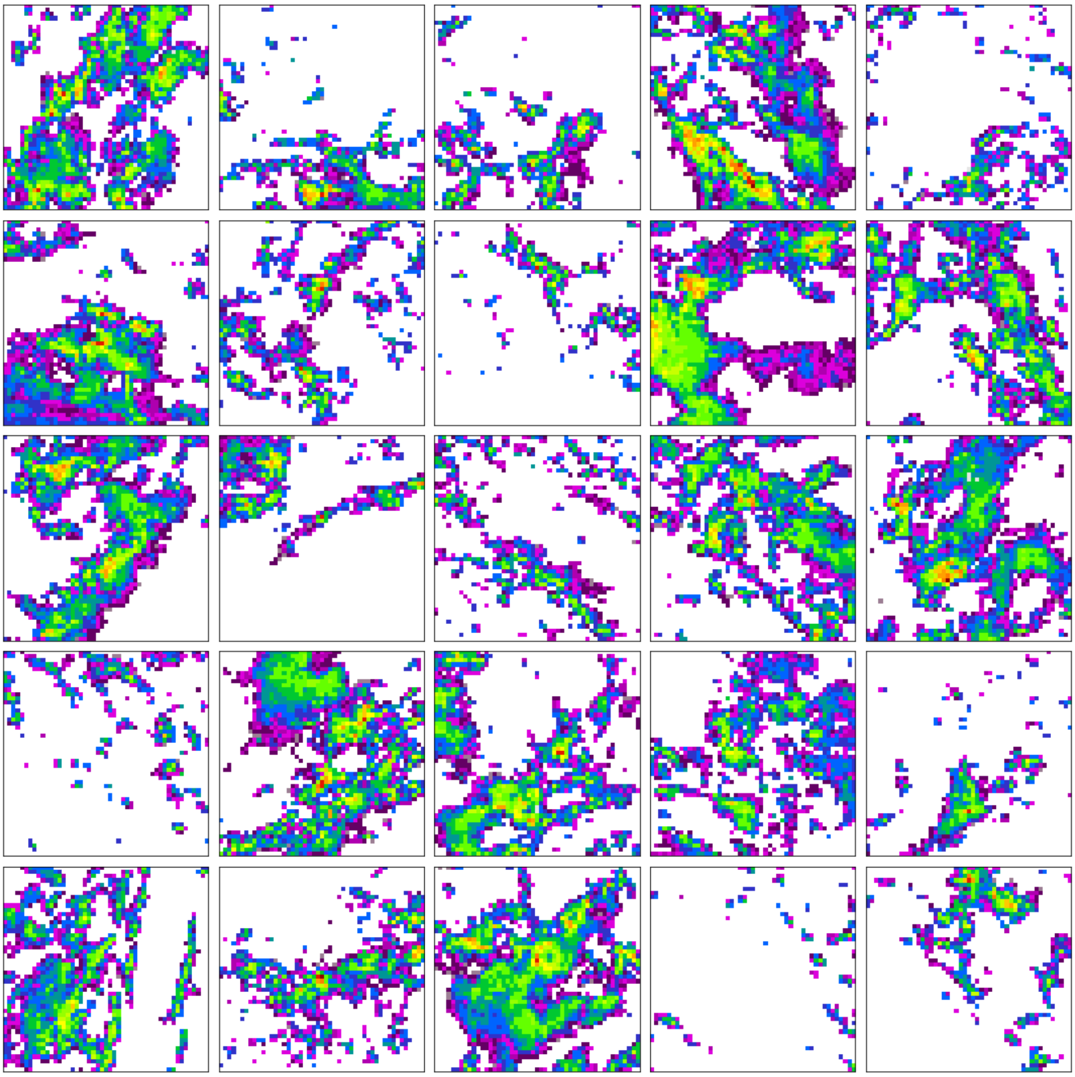


Figure 13: SOM grid samples.

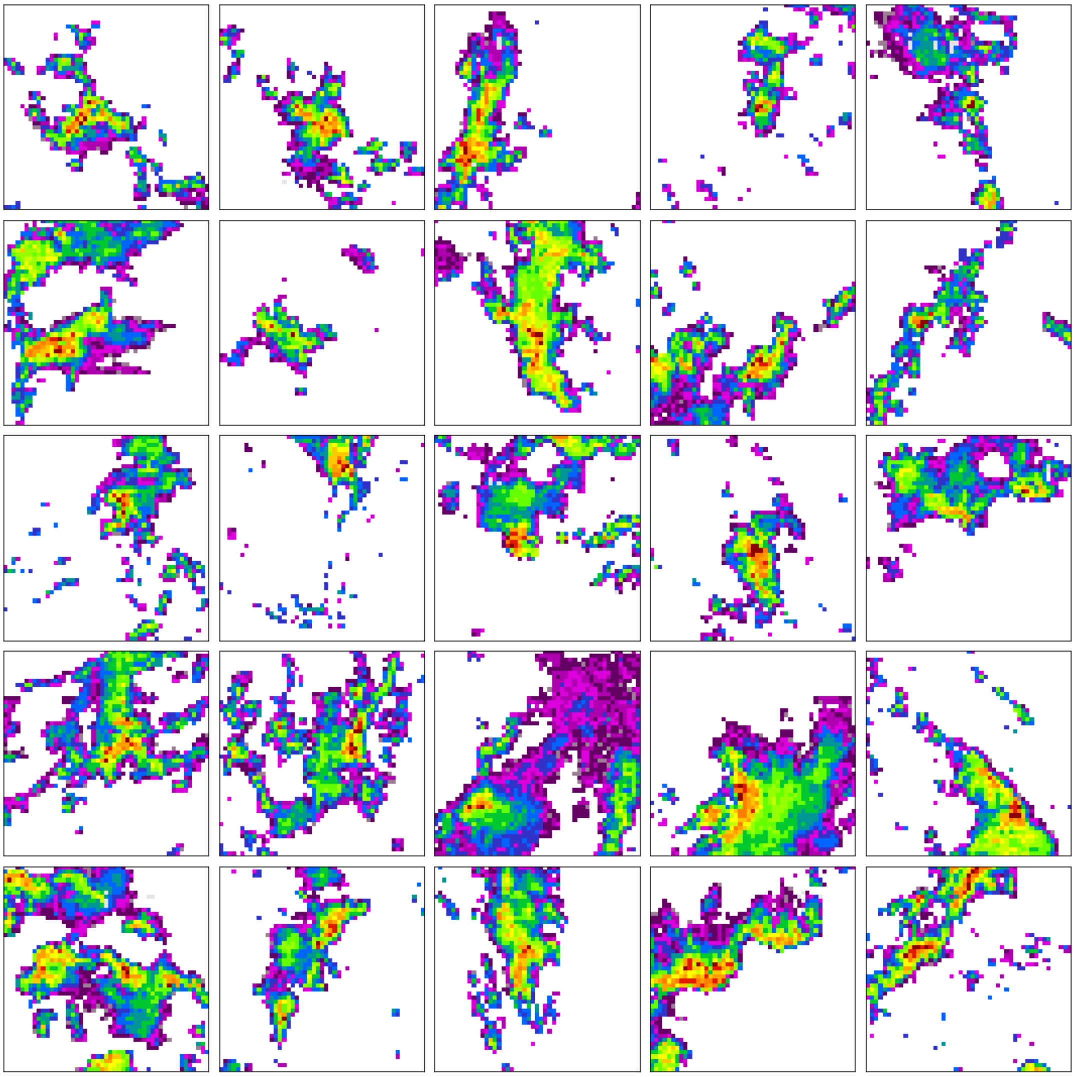


Figure 14: SOM grid samples for node in the first line and fifth column.

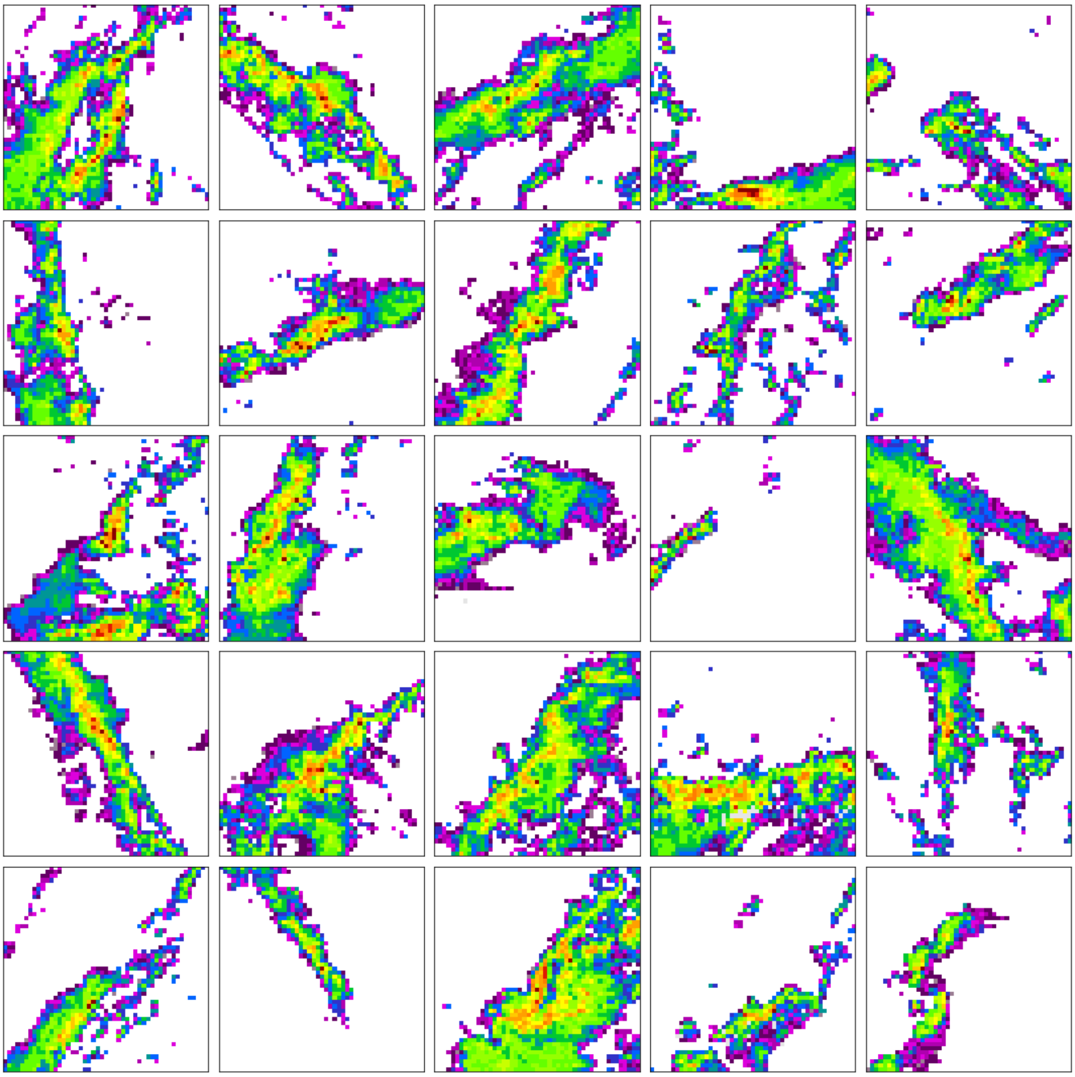


Figure 15: SOM grid samples for node in the 5th Line and last Column.

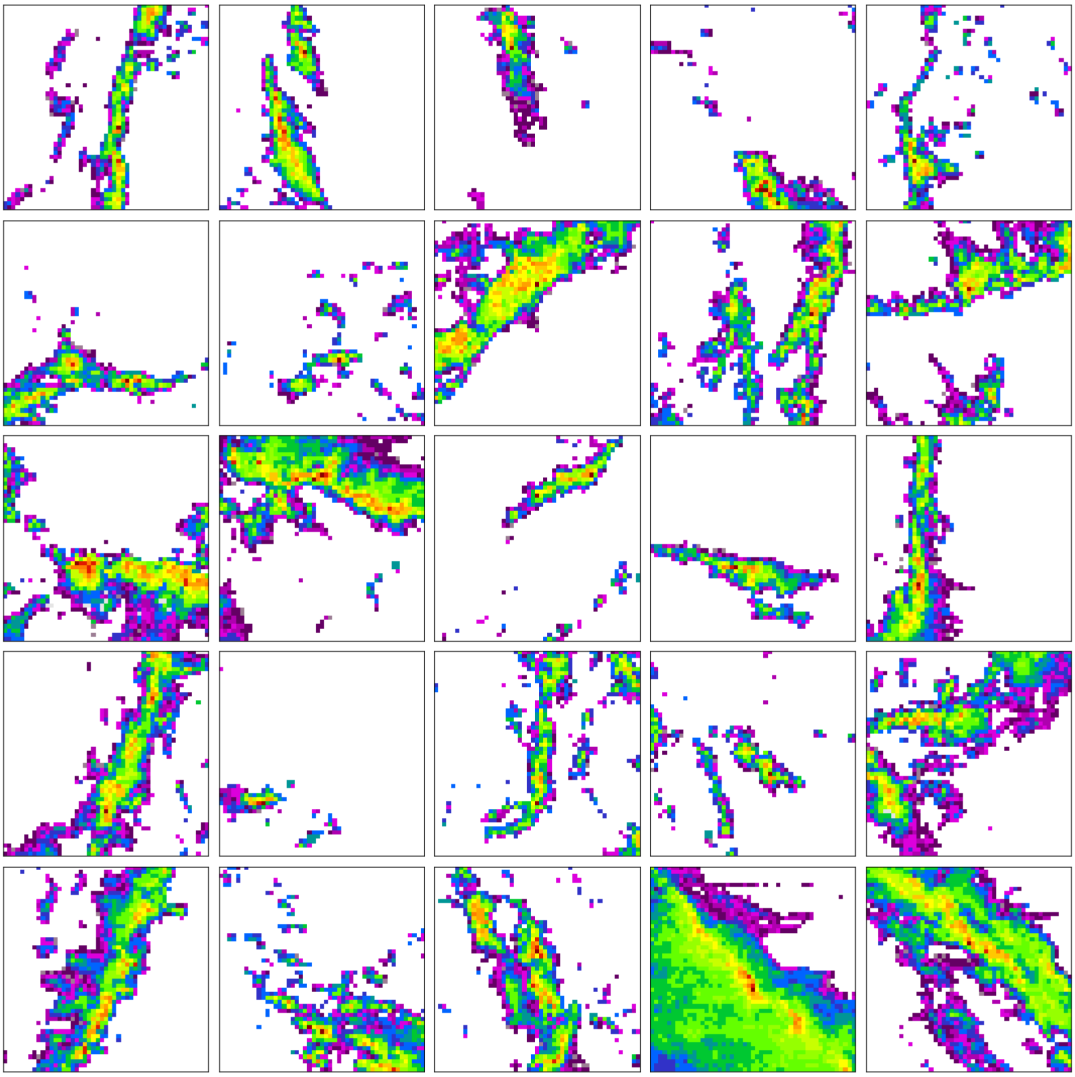


Figure 16: SOM grid samples for node in the 6th Line and last Column.

A.3 Nodes samples second SOM

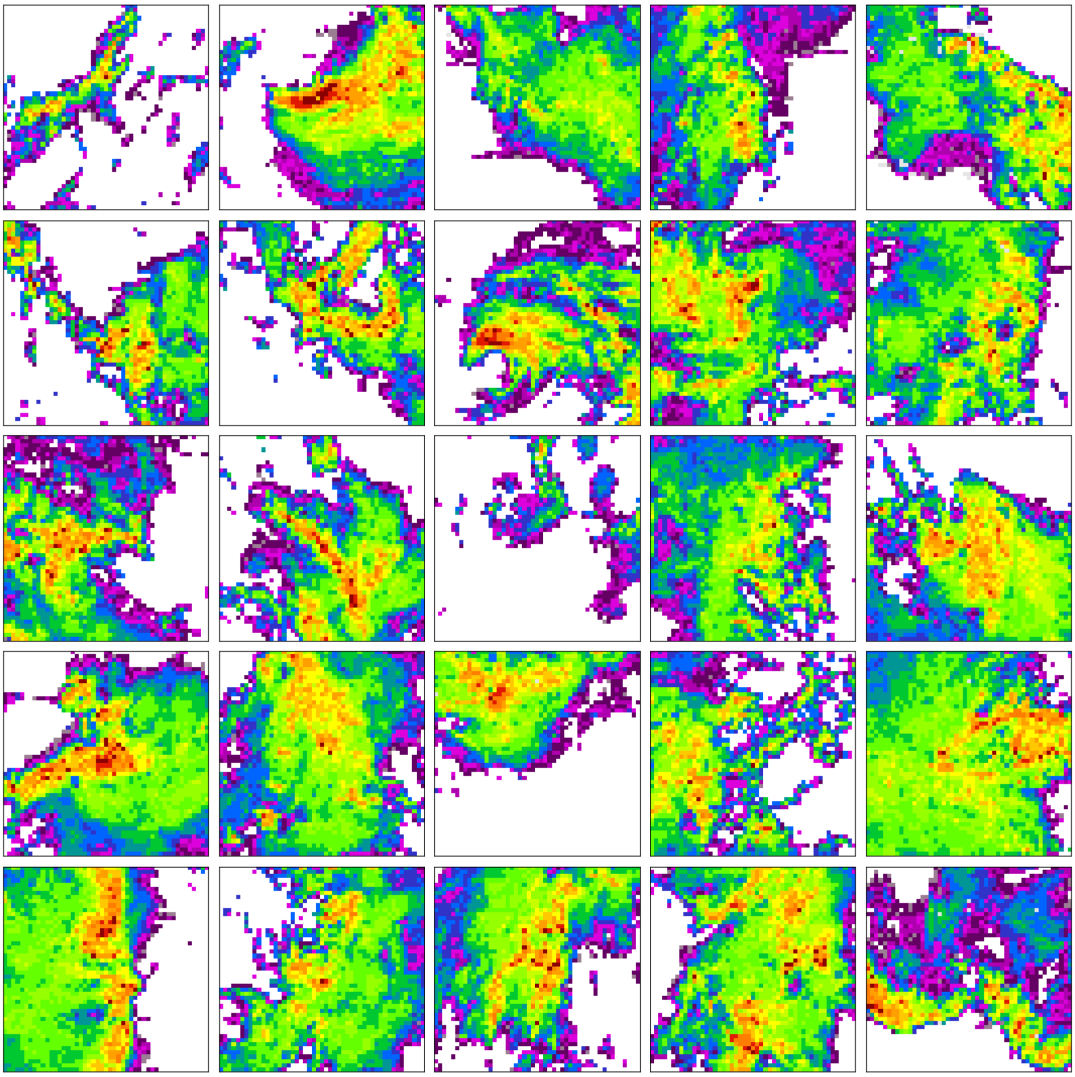


Figure 17: SOM grid samples for node in the first Line and Third Column.