Reply to editor and reviewers

Dear editor and reviewers,

We sincerely thank the reviewers for their careful reading of our manuscript and for providing insightful comments and suggestions. In response to the reviewers' feedback, we have addressed each point and revised the manuscript accordingly. We believe that these revisions have substantially improved both the clarity and overall quality of the manuscript. Detailed responses to each comment are provided below:

Reviewer #1

I appreciate the explanations provided by the authors. Generally, I accept them, though my preference is still for independence of the in the classification from the analyzed variables. We will have to agree to disagree on that point. On my second and third major points, it would be helpful if more of this can make its way into the manuscript. Not all of the studies referenced in the author response to my point #2 even appear in the manuscript. I understand that the authors did not produce the product that they analyze, but the question stands regardless. A few words on the limitations of the product with regard to ice microphysics is warranted in the data section - this is also relevant to a similar question in the CC review. As for my third point, I like this explanation from the authors, "Unlike vertically integrated variables (e.g., total water path), which aggregate information and may obscure vertical contrasts, PCA preserves key vertical structure signals that are physically relevant to precipitation processes." Might it be possible to work this into Section 2.3?

Reply: Thank you for your thoughtful and constructive comments. We appreciate your positive evaluation of our previous explanations.

Regarding your suggestion, we have incorporated additional clarifications from our response into the revised manuscript. The relevant discussions addressing your second and third major points have been expanded in the main text.

Furthermore, we agree that the explanation of the PCA methodology provided in our response improves the clarity of Section 2.3.

We hope these revisions adequately address your concerns and further strengthen the manuscript.

Reviewer #2

Specific comments - -

- The division into the eight categories seems somewhat arbitrary, and there is not a single clear takeaway from the paper; it would be difficult to summarise the importance of the results in one sentence. What's the most important thing learned here?

Reply: We thank the reviewer for this important comment. We acknowledge that the clarity of the clustering rationale and the key takeaway of the study can be further strengthened. The selection of eight clusters was determined based on the combined evaluation of the Davies–Bouldin (DB) index and the Calinski–Harabasz (CH) score. We acknowledge that the description in the original manuscript may not have been sufficiently clear. In the revised version, we have clarified the methodological explanation and explicitly stated the criteria used for determining the optimal number of clusters to improve transparency and readability.

Regarding the main takeaway, the primary contribution of this study is the objective, global-scale identification of distinct precipitation system (PS) types with multiple precipitation features derived from GPM DPR observations. The analysis reveals that these PS types exhibit systematic and physically interpretable differences in their structural and microphysical characteristics, and that continental–oceanic contrasts emerge naturally from the unsupervised framework without imposing land–ocean information. We have revised the Introduction and Conclusion sections to more clearly highlight this key finding and the overall significance of the work.

-This work is not clear as to whether it covers all precipitation phases or is specifically focused on rain. The focus on drop size distribution implies that rain is the only precipitation of interest, but if ice-phase precipitation is excluded from the data, this is never made clear.
Reply: Thank you for the helpful comment. This study does not focus solely on rain. The analysis includes precipitation systems containing liquid-, mixed-, and ice-phase hydrometeors derived from the GPM DPR observations. We agree that the previous description may have caused confusion because of the emphasis on drop size distribution (DSD). It should be clarified that the DSD parameters used in this study are intended to represent the integrated microphysical characteristics of precipitation systems rather than to denote rain-only processes. Meanwhile, we have clarified in the revised data description section that the precipitation systems analyzed in this study include liquid, mixed-phase, and ice-phase conditions.

-There is also a reference to warm rain vs cold rain processes, which implies that ice must be considered at some stage. There are a few claims that the microphysical information gathered will be useful for weather and/or climate prediction, but no specification of how. This seems especially odd in the context of having four ‘extreme’ categories, which it would be of obvious value to be able to more accurately predict.
Reply: Thank you for this thoughtful comment. Regarding the precipitation phase issue, we have clarified in the revised data description section that the precipitation systems analyzed in this study include liquid, mixed-phase, and ice-phase conditions.

Regarding the broader value of this work, we have revised the manuscript to better

articulate its scientific significance.

Line-by-line:

Lines 39–40: 'Moreover, disdrometers have not been deployed globally' – this is not correct as disdrometers have been deployed globally but are limited by being single, fixed points of observation; rephrase as something like 'Moreover, disdrometer observations are relatively sparse, especially over the ocean.'

Reply: Thank you for this suggestion. Revised.

Lines 50–52: 'Notably, DSD parameters were retrieved from the radar reflectivity measured by the PR with the assumption that the DSD can be characterized by the diameter parameter itself … As a result, the DSDs obtained via retrieval exhibited large errors.' Why and how did these errors result? This is not immediately clear even to another atmospheric scientist. Why was that approach inaccurate, and how does the data you use avoid this problem?

Reply: We thank the reviewer for the helpful comment. To avoid potential misunderstanding, the sentence in question has been removed in the revised manuscript. We note that the manuscript already describes the improved retrieval capability of the GPM DPR relative to the earlier PR.

Lines 56–57: 'The retrieved DSD parameters have been verified with ground-based observations and are better than those obtained via the TRMM PR algorithm' – quantify 'better', e.g. closer in value to ground-based observations?

Reply: Revised.

Lines 73–74: 'These studies revealed the variations in microphysical characteristics across different seasons and rainfall types.' This sentence does not add new information to an already long paragraph – delete.

Reply: Deleted.

Line 84: I suggest ending the paragraph at 'formation process of precipitation.' and starting a new paragraph from 'At present'.

Reply: Revised.

Lines 90–91: 'Although the dataset covered a wide range of precipitation regimes, it could not capture all rain regimes' – why not? What had to be omitted and why? Also, 'drop size distribution' in the context of precipitation systems necessarily must mean rain and not ice.

Reply: As shown in the Reply-Figure, the 12 disdrometers mainly located in the U.S., Darwin, Finland, and Indian ocean. None is found in the South America, Africa, and Asian continent. As we known, the precipitation over Amazon and Africa are typically different from the regions studied by Dolan et al. (2018) with disdrometers. Therefore, we conclude that "it could not capture all rain regimes". We revised this sentence to make it clear, for example, it should be climatic regimes but not rain regimes.

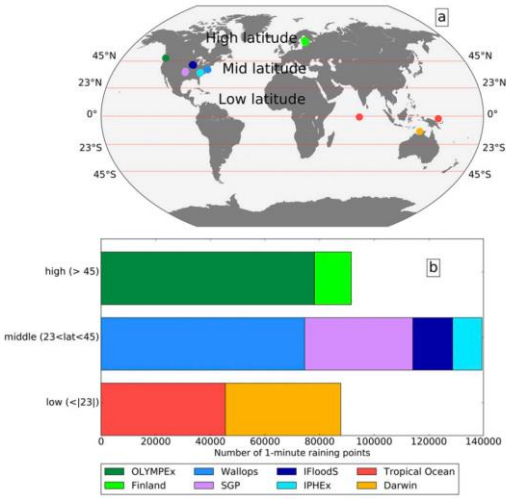

FIG. 1. (a) Locations of disdrometer observations used herein. See Table 1 for individual datasets that fit into the larger eight locations. (b) Distribution of raining minutes as a function of latitude band.

Reply-Figure 1. Adopted from Dolan et al. (2018)

Lines 112–114: 'The parameters used in this machine learning model include DSD parameters (Dm and Nw), near-surface precipitation rate (mm h-1), attenuation-corrected radar reflectivity (dBZ), reflectivity near the surface (Zsurf), and typeprecip (stratiform or convective precipitation), and airTemperature.' This sentence is a confusing mixture of variable names, long names, and units. A table with variable name offset in code formatting would help, e.g.: Variable name precipRateNearSurface Long name Unit near-surface precipitation rate mm h-1 Lines
Reply: We have revised the manuscript by summarizing the input variables in a new table (Table 1), where the variable names, full descriptions, and units are clearly listed using consistent formatting.

115–117: 'The precipitation type helps distinguish between stratiform and convective precipitation pixels, while air temperature is used to separate snow from rain' – why not use the phase information already present in the DPR data?
Reply: According to GPM document, the determination of phase relies on bright band as shown in the screenshot. The phase of ground surface precipitation is determined by node D in the situation of liquid precipitation. In the stratiform precipitation without bright band and convective precipitation, the node D is set according to 0℃. Therefore, if the ambient temperature at ground level is 2 ℃, the phase will be defined as liquid. In reality, even if the ground level is 2 ℃, the precipitation could also be solid or mixed phase. Therefore, to be strict, we used the threshold (surface air temperature larger than 10 ℃) to exclude those precipitation systems with solid precipitation on the ground.

3.6.2.2.1 Stratiform precipitation with bright band

The range bin that corresponds to the upper edge of bright band is node B, the range bin at the peak of bright band is node C, and the range bin at the lower edge of bright band is node D. Between node B and node D, particle temperature is set to 0°C. Above node B, the particle temperature is basically the same with ambient air temperature, but is 0°C if the air temperature is higher than 0°C. Below node D, the particle temperature is basically set to the same with the air temperature, but 0°C if the air temperature is lower than 0°C. The range bin with particle temperature closest to −20°C corresponds to node A, and that closest to 20°C is node E. At and above node A, the phase of precipitation is *Solid*. At and below node D, the phase of precipitation is *Liquid*. At other range bins, below node A and above node D, the phase of precipitation is *Mixed* (Solid and Liquid).

3.6.2.2.2 Stratiform precipitation without bright band

The process is the same as Section 3.6.2.2.1, but the nodes B, C, and D are at the range bin corresponding to 0°C.

3.6.2.2.3 Convective precipitation or other-type precipitation

The process is the same as Section 3.6.2.2.2.

Lines 126–128: 'The average Dm and Nw profiles were used for each PS, and if the profiles of the maximum Dm and Nw values in each layer were involved, MAX-Dm and MAX-Nw, respectively, were used' – what does 'involved' mean here?
Reply: Thank you for pointing out the ambiguity. Here, "involved" was intended to mean that when the analysis refers to the layer-wise maximum values, we use the variables MAX-Dm and MAX-Nw rather than the mean profiles. To avoid confusion, we have revised this sentence in the manuscript to make the distinction clearer.

Line 128: what is Ze? This is never defined. (Reflectivity by analogy with Zsurf?)
Reply: Thank you for pointing this out. Here, Z_e refers to the radar reflectivity factor (dBZ). We have now explicitly clarified its definition in the newly added variable table to avoid confusion.

Lines 155–156: 'In summary, PCA aims to transform numerous pertinent features into a comparatively limited number of irrelevant ones, thereby retaining as much of the informational content of the original data as possible' – it's the other way around: 'In summary, PCA aims to transform numerous irrelevant features into a comparatively limited number of pertinent ones, thereby retaining as much of the informational content of the original data as possible'. Re: section 2.3 generally, your explanation in the response to reviewer comments was exceptionally clear: The purpose of PCA was to reduce the vertical dimension, compressing the original 176 height levels into a single representative component for each variable. Only one principal component was retained for each profile, as it explains the dominant variance of the vertical structure and serves as a compact descriptor for clustering. The PCA was applied once,

independently for each parameter, and no temporal dimension was involved in the PCA procedure. Time was not treated as an input dimension; only the vertical (height) dimension was reduced. This is an excellent, straightforward explanation, which I found easier to understand than the text of your actual paper. Consider editing
Reply: Thank you for the helpful suggestion. We have corrected the description of PCA and revised Section 2.3 followed the suggestion.

lines 155–163 to something like the following: In summary, PCA aims to transform numerous irrelevant features into a comparatively limited number of pertinent ones, thereby retaining as much of the informational content of the original data as possible. In our case, we used PCA to reduce the vertical dimension, compressing the original 176 height levels into a single representative component for each variable (Ze, Dm, or Nw). Only one principal component per variable was retained for each profile, as this sufficiently captures the dominant variance of the vertical structure and served as a compact descriptor for clustering. The PCA was applied once, independently, for each parameter. Time was not treated as an input dimension; only the vertical (height) dimension was reduced.
Reply: Thank you for the helpful suggestion. We have revised this part followed your suggestion.

Lines 186–187: 'the precipitation area directly characterizes the spatial differences in both the vertical and horizontal distributions of the system' – how does the horizontal area characterise the vertical distribution of the system? Explain.
Reply: We thank the reviewer for pointing this out. We apologize for the unclear and inaccurate wording in the original manuscript. The precipitation area is intended to characterize the horizontal spatial extent of the system rather than its vertical distribution. We have revised the sentence in the manuscript to clarify this point.

Lines 177–178: 'Collectively, constructing a multidimensional precipitation feature space with these parameters enhances the accuracy and robustness of the clustering analysis' – like the sentence in lines 73–74, this sentence is not adding new information and can be deleted.
Reply: Deleted.

Lines 202–204: 'Geographically, high-latitude shallow precipitation systems are mainly distributed north of 40°N and south of 35°S, subtropical shallow precipitation systems occur primarily within about 20° on both sides of the equator, and marine extreme convection is mostly located over the ocean' – this doesn't make sense to me. Wouldn't all cases of marine extreme convection be located over the ocean? Similarly, while I understand that you might have a 'deep' PS at high latitudes (which would therefore not therefore be classified as 'high-latitude shallow'), I don't understand how all cases of 'high-latitude shallow' are not located at high latitudes.
Reply: Thank you for the careful reading. We agree that the original wording was potentially confusing. Our intention was to describe the observed geographical

distribution characteristics of each PS type rather than to imply that the classification itself was defined by location. The PS types were obtained objectively from the clustering algorithm without using geographic information. To describe these eight PS types, we named them according to the spatial distribution patterns and vertical structures. The first PS type is named as high-latitude shallow PS because they mostly located in high-latitude regions and echo top height is low. We have revised the text to clarify this point and to avoid redundant or ambiguous wording.

Lines 209–211: Like RC2, I expected the 'extreme' cases to be defined by extremes in physical variables, like a very high amount of precipitation in mm or a very high rain rate in mm h-1. It's not clear in a physical sense why the extreme cases are extreme. Figures 4 through 6 make the case for 'extremity' clearer, so perhaps reference those here.
Reply: Thank you for this helpful comment. We agree that the physical meaning of the term "extreme" should be clarified. In the revised manuscript, we have refined the description to explicitly state that the "extreme" PS types refer to precipitation systems exhibiting markedly stronger precipitation intensity, radar reflectivity, and vertical development relative to other PS categories.

Lines 266–267: 'the most severe convective storms' – what does 'severe' mean in this context? Greatest convective intensity? 'Severe' would usually be used to describe consequences, like flooding or hail damage.
Reply: Thank you for this helpful comment. We agree that the term "severe" may be ambiguous and is more commonly associated with hazardous impacts rather than convective intensity. In the revised manuscript, we have replaced "the most severe convective storms" with "the most intense convective storms" to better reflect the physical meaning.

Line 320: 'Additionally, land indicates a dry adiabatic lapse rate' – land surface does not indicate or guarantee a dry adiabatic lapse rate, only a comparatively drier one than over ocean.
Reply: Thank you for the helpful comment. The statement has been revised in the manuscript to reflect this more accurate description.

Lines 374–375: 'more concentrated distributions' – what does 'concentrated' mean? Narrower distributions?
Reply: Here, "more concentrated distributions" was intended to indicate narrower distributions. We have revised the wording in the manuscript.

Line 384: 'showed a significant abnormal value of Nw' – abnormal how? High or low, or both?
Reply: Here, "abnormal" specifically refers to anomalously low values of Nw. We have clarified this in the revised manuscript to avoid ambiguity.

Lines 406–407: ‘Ryu et al. (2021) analyzed DSDs during three types of heavy rainfall events with different rain intensities. They also reported that Dm increases with increasing rainfall intensity, whereas Nw decreases with increasing rainfall intensity.’ Avoid literature review in the results – this could be cut or moved back to the introductory material.
Reply: Deleted.

Lines 411–412: ‘Notably, in extreme convection, with strong convection at the top of the storm, attenuation becomes notable at low storm levels, which might influence the retrieval of microphysical parameters.’ Expand on this briefly and specify that you are referring to rain attenuation of radar. Even within the atmospheric science community, it is helpful to clarify the physical mechanisms.
Reply: In the revised manuscript, we have explicitly clarified that the attenuation refers to rain-induced radar attenuation and have added a brief explanation of how strong low-level precipitation can weaken the radar signal and potentially affect the retrieval of microphysical parameters.

Line 426: ‘were significantly distinct’ – do not use ‘significantly’ qualitatively; add quantitative measure of significance, or rephrase.
Reply: Thank you for the helpful suggestion. The sentence has been rephrased to avoid implying statistical significance and to improve clarity.

Line 449: ‘occurred mostly between 0 and 5 a.m.’ – is this local time? If so, specify; perhaps rephrase as ‘occurred mostly between 00:00 and 05:00 local time’.
Reply: Revised.

Line 469: ‘The diurnal cycles of Nw were basically different with those of Dm and occurrence’ – what does this mean? Rephrase.
Reply: Thank you for the helpful comment. In the original sentence, “basically different” was intended to indicate that the overall diurnal variation pattern of N_w differs from those of D_m and occurrence. To improve clarity and precision, we have revised the sentence accordingly in the manuscript.

Lines 504–506: ‘Additionally, continental rainfall is associated with lower Nw values due to the cold rain mechanism, whereas oceanic rainfall is associated with higher Nw values resulting from a warm rain regime’ – these mechanisms have not previously been discussed/outlined in this paper, so this line is unexpected. Add a line or two of explanation.
Reply: Added.

Lines 507–509: ‘PS with a higher land proportion exhibit more concentrated Nw values, whereas those with a greater ocean proportion exhibit larger Nw values. However, the distribution of Dm is the opposite: PS with a higher ocean proportion exhibit more concentrated Dm values than land-dominated PSs do’ – as in lines 374–

375, it's not clear what 'concentrated' means here. Smaller? More tightly clustered?
Reply: We have clarified the meaning of "concentrated" by replacing it with statistical descriptions where appropriate.

Technical corrections Issues throughout paper:- - consistency in verb tense to discuss your work: is it 'This study focused on …' (as in line 43) or 'all the ~9 million PSs derived in section 2.2 are used' (as in line 165)? Stardardise. consistency in pluralising 'PS' as 'PSs'. Some sentences (e.g. 'their ratios of the total PS are less than' in line 210 or 'Subtropical shallow PS primarily occurred' in line 229) read as though the plural is supposed to be understood to exist within the abbreviation: 'their ratios of the total precipitation systems are less than' and 'Subtropical shallow precipitation systems primarily occurred'. But in other places, like in line 207, 'the various types of PSs differed', the plural 's' is put outside the acronym. In other places, the PS seems to be singular, e.g. line 310: 'Extreme strong PS attained an echo top height greater than 18 km, and it also exhibited' … This needs to be standardised across the whole paper.
Reply: Thank you for this helpful suggestion. In addition, we have carefully reviewed the entire manuscript and standardized the verb tense usage and the pluralization of "PS/PSs" to ensure consistency and clarity throughout the paper.

Line 5: 'Mou ntain' → 'Mountain'
Reply: Revised.

Line 18: 'mainly locate over tropical continent' → 'are mainly located over tropical continents'
Reply: Revised.

Lines 21–22: 'In contrary' → 'In contrast'
Reply: Revised.

Lines 69 and 70: 'Precipitation Rate' → 'precipitation rate' (no need to capitalise)
Reply: Revised.

Line 72: 'demonstrate obvious seasonal variations' → 'demonstrated obvious seasonal variations'
Reply: Revised.

Line 126: 'Nw profiles' → 'Nw profiles' (space between words has been accidentally subscripted)
Reply: Revised.

Line 166: 'the maximum precipitation rate near the surface, Htop, the precipitation area' → 'the maximum precipitation rate near the surface, the echo top height of the PS, the precipitation area'

Reply: Revised.

Line 196: ‘convective/stratiform fractions and so on.’ → ‘convective/stratiform fractions.’ (‘And so on’ is too informal.)
Reply: Revised.

Line 239: ‘There regions’ → ‘These regions’ Lines
Reply: Revised.

250–251: ‘which isa hallmark’ → ‘which is a hallmark’
Reply: Revised.

Line 253: ‘Marine extreme PS was primarily’ – see note above about plurals; suggest ‘Marine extreme PSs were primarily’
Reply: Revised.

Line 254: ‘with only 943 PSs and 90% is over the ocean’ → ‘with only 943 PSs, of which 90% were over the ocean’
Reply: Revised.

Lines 261–262: ‘Oceanic extreme PS (extreme deep PS and marine extreme PS) with a high fraction of ocean pixels, have mean precipitation coverage area exceeding 36000 km2, significantly larger than continental extreme PS (strong PS and extreme strong PS)’ → ‘Oceanic extreme PS (extreme deep PS with a high fraction of ocean pixels and marine extreme PS) have mean precipitation coverage areas exceeding 36000 km2, significantly larger than continental extreme PS (strong PS and extreme strong PS)’
Reply: Revised.

Line 264: ‘Zhang and Wang (2021) . Furthermore’ → ‘Zhang and Wang (2021). Furthermore’ Figure 1: caption should have a period at the end Table 1: ‘PS’ is not necessary in the categories; they are all PSs
Reply: Revised.

Line 345: ‘the routinely retrieval algorithm’ → ‘the routine retrieval algorithm’
Reply: Revised.

Lines 346–347: ‘the DPR observations cannot quantify ice particle content and their size above the stratiform rain regions’ → ‘the DPR observations cannot quantify ice particle distribution and ice crystal size above the stratiform rain regions’? or ‘the DPR observations cannot quantify ice water content and ice crystal size above the stratiform rain regions’?
Reply: Thank you for the helpful suggestion. We have revised the statement accordingly and adopted the second formulation in the revised manuscript.

Line 367: ‘stratiform and convection samples’ → ‘stratiform and convective samples’
Reply: Revised.

Figure 7: unnecessary parentheses: ‘(the blue and orange rectangles denote the maritime and continental convective clusters, respectively, in Dm and log10(Nw) space from Bringi et al. (2003))’ → ‘The blue and orange rectangles denote the maritime and continental convective clusters, respectively, in Dm and log10(Nw) space from Bringi et al. (2003)’
Reply: Revised.

Line 388: ‘Ni et al. (2019) revealed’ → ‘Ni et al. (2019) revealed’
Reply: Revised.

Line 392: ‘Note that although the mean Dm and Nw values do not capture the variety of DSDs in each PS’ → ‘Note that the mean Dm and Nw values do not capture the variety of DSDs in each PS’; ‘although’ doesn’t make sense here
Reply: Revised.

Line 424: ‘due to predominate evaporation’ → ‘due to predominantly evaporation’
Reply: Revised.

Lines 430–432: ‘Balanced breakup and coalescence processes in the microphysical processes of extreme PS accounted for more than 40% of the total microphysical processes, significantly exceeding other three types of microphysical processes’ → ‘Balanced breakup and coalescence processes in extreme PS accounted for more than 40% of the total microphysical processes, significantly exceeding the other three types of microphysical processes’
Reply: Revised.

Line 455: ‘but shown a peak’ → ‘but showed a peak’
Reply: Revised.

Line 464: ‘have peaks in the around 15 pm’ → ‘have peaks around 15:00 local time’
Reply: Revised.

Line 473–474: ‘The extreme strong PS shown low values of Nw in the afternoon and little variations at night’ → ‘The extreme strong PS showed low values of Nw in the afternoon and little variation at night’
Reply: Revised.

Line 492: ‘unique microphysical and convection properties’ → ‘unique microphysical and convective properties’
Reply: Revised.

Lines 519–520: ‘Continental convection Cluster peak in the afternoon and summer’ → ‘Continental convection clusters peak in the afternoon and summer’
Reply: Revised.

Line 529: ‘treated as integrated entity’ → ‘treated as an integrated entity’
Reply: Revised.

Line 540: ‘to address this issue. By analyzing the interactions’ → ‘to address this issue. Nonetheless, by analyzing the interactions’ – this is an optional change, but a shift in tone to emphasise the value added by your work
Reply: Thank you for the helpful suggestion. We have revised the sentence in the manuscript to adopt the proposed wording.

Line 544: ‘NASA/G-oddard’ → ‘NASA Goddard’
Reply: Revised.