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?

- 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. 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.

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.'

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?

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?

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.

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

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.

Lines 112–114: ‘The parameters used in this machine learning model include DSD parameters (D_m and N_w), near-surface precipitation rate (mm h^{-1}), attenuation-corrected radar reflectivity (dBZ), reflectivity near the surface (Z_{surf}), 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	Long name	Unit
`precipRateNearSurface`	near-surface precipitation rate	mm h^{-1}

Lines 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?

Lines 126–128: ‘The average D_m and N_w profiles were used for each PS, and if the profiles of the maximum D_m and N_w values in each layer were involved, MAX-D_m and MAX-N_w, respectively, were used’ – what does ‘involved’ mean here?

Line 128: what is Z_e? This is never defined. (Reflectivity by analogy with Z_{surf}?)

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 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 (Z_e, D_m, or N_w). 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.

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.

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.

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.

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.

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.

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.

Lines 374–375: ‘more concentrated distributions’ – what does ‘concentrated’ mean? Narrower distributions?

Line 384: ‘showed a significant abnormal value of N_w’ – abnormal how? High or low, or both?

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

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.

Line 426: ‘were significantly distinct’ – do not use ‘significantly’ qualitatively; add quantitative measure of significance, or rephrase.

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’.

Line 469: ‘The diurnal cycles of N_w were basically different with those of D_m and occurrence’ – what does this mean? Rephrase.

Lines 504–506: ‘Additionally, continental rainfall is associated with lower N_w values due to the cold rain mechanism, whereas oceanic rainfall is associated with higher N_w 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.

Lines 507–509: ‘PS with a higher land proportion exhibit more concentrated N_w values, whereas those with a greater ocean proportion exhibit larger N_w values. However, the distribution of D_m is the opposite: PS with a higher ocean proportion exhibit more concentrated D_m values than land-dominated PSs do’ – as in lines 374–375, it’s not clear what ‘concentrated’ means here. Smaller? More tightly clustered?

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 system<u>s</u> are less than’ and ‘Subtropical shallow precipitation system<u>s</u> 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.

Line 5: ‘Mou ntain’ → ‘Mountain’

Line 18: ‘mainly locate over tropical continent’ → ‘are mainly located over tropical continents’

Lines 21–22: ‘In contrary’ → ‘In contrast’

Lines 69 and 70: ‘Precipitation Rate’ → ‘precipitation rate’ (no need to capitalise)

Line 72: ‘demonstrate obvious seasonal variations’ → ‘demonstrated obvious seasonal variations’

Line 126: ‘N_w profiles’ → ‘N_w profiles’ (space between words has been accidentally subscripted)

Line 166: ‘the maximum precipitation rate near the surface, H_{top}, the precipitation area’ → ‘the maximum precipitation rate near the surface, the echo top height of the PS, the precipitation area’

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

Line 239: ‘There regions’ → ‘These regions’
Lines 250–251: ‘which isa hallmark’ → ‘which is a hallmark’

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

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

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 km^2, 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 km^2, significantly larger than continental extreme PS (strong PS and extreme strong PS)’

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

Line 345: ‘the routinely retrieval algorithm’ → ‘the routine retrieval algorithm’

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’?

Line 367: ‘stratiform and convection samples’ → ‘stratiform and convective samples’

Figure 7: unnecessary parentheses: ‘(the blue and orange rectangles denote the maritime and continental convective clusters, respectively, in D_m and log10(N_w) space from Bringi et

al. (2003))' → 'The blue and orange rectangles denote the maritime and continental convective clusters, respectively, in D_m and log10(N_w) space from Bringi et al. (2003)'

Line 388: 'Ni et al. (2019) revealed' → 'Ni et al. (2019) revealed'

Line 392: 'Note that although the mean D_m and N_w values do not capture the variety of DSDs in each PS' → 'Note that the mean D_m and N_w values do not capture the variety of DSDs in each PS'; 'although' doesn't make sense here

Line 424: 'due to predominate evaporation' → 'due to predominantly evaporation'

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'

Line 455: 'but shown a peak' → 'but showed a peak'

Line 464: 'have peaks in the around 15 pm' → 'have peaks around 15:00 local time'

Line 473–474: 'The extreme strong PS shown low values of N_w in the afternoon and little variations at night' → 'The extreme strong PS showed low values of N_w in the afternoon and little variation at night'

Line 492: 'unique microphysical and convection properties' → 'unique microphysical and convective properties'

Lines 519–520: 'Continental convection Cluster peak in the afternoon and summer' → 'Continental convection clusters peak in the afternoon and summer'

Line 529: 'treated as integrated entity' → 'treated as an integrated entity'

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

Line 544: 'NASA/G-oddard' → 'NASA Goddard'