Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Distinct convective precipitation structures underlie similar surface rainfall and regional contrasts in rainfall extremes

Zelin Wang
National University of Defense Technology
Xiong Hu
huxiong18@nudt.edu.cn
National University of Defense Technology
Weihua Ai
National University of Defense Technology
Weimin Zhang
National University of Defense Technology
Shuo Ma
National University of Defense Technology
Heyuan peng
National University of Defense Technology
Zhonghui Tan
National University of Defense Technology

Article

Keywords:

Posted Date: November 13th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11565580/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Distinct convective precipitation structures underlie similar surface rainfall and regional contrasts in rainfall extremes

Zelin Wang[1], Xiong Hu2*, Weihua Ai[1]*, Weimin Zhang[1], Shuo Ma[1], Heyuan Peng[1], Zhonghui Tan[1]

[1]College of Meteorology and Oceanography, National University of Defense Technology, Changsha, China

[2]College of Basic Military and Political Education, National University of Defense Technology, Changsha, China

Corresponding authors: X. Hu (huxiong18@nudt.edu.cn), W. Ai (aiweihua@nudt.edu.cn).

Abstract

Conditioning radar profiles on near-surface rain rates and drop size distributions (DSDs) often obscures distinct convective structures. Instead, this study applies representation-learning-guided clustering to reflectivity-profile shapes from 1.65 million convective profiles observed by the Global Precipitation Measurement Dual-frequency Precipitation Radar, identifying six structural archetypes without near-surface inputs. They span shallow warm-rain to deep convection but overlap in near-surface precipitation properties. Despite this similarity, echo-top heights within the modal DSD bin differ threefold, reflecting contrasting ice-phase and warm-rain contributions. Across rainfall rates, these microphysical pathways diverge primarily at moderate intensities, where shallow archetypes show greater warm-layer rain-rate and drop-size increases, but converge at higher intensities as deeper archetypes show stronger drop-size growth. At the regional scale, this classification separates tropical land–ocean extreme-rainfall contrasts into opposing contributions from archetype prevalence and within-archetype extreme-rainfall probability. These findings provide structure-resolved benchmarks for evaluating convective precipitation in weather and climate models.

Introduction

Near-surface precipitation is the result of combined effects of microphysical processes aloft. In convective precipitation, processes including riming, aggregation, melting, collision–coalescence, breakup and sub-cloud evaporation shape the rainfall intensity and raindrop size distribution (DSD)[1–4]. Connecting these surface signatures to their vertical origins is essential for constraining satellite precipitation retrieval[5,6], evaluating convective and microphysical parameterizations[7], and understanding precipitation extremes under global warming[8–12].

Previous studies have investigated this linkage by conditioning on near-surface precipitation properties, grouping events by near-surface rain rate (*nsRR*) or DSD to examine their associated vertical and microphysical characteristics[13–16]. Regional studies using the Global Precipitation Measurement (GPM) Dual-frequency Precipitation Radar (DPR) have documented monsoon-related differences in precipitation efficiency and raindrop characteristics[16,17]. At the global scale, rain-rate-conditioned composites reveal a shift in the dominant precipitation growth mechanism from liquid-phase to mixed-phase processes as rainfall intensifies[18], while analyses in the DSD plane identify microphysical patterns and pathways across surface precipitation states[19–21]. These studies establish systematic relationships between near-surface precipitation states and their mean microphysical characteristics. However, such analyses do not establish whether these mean relationships reflect structurally similar precipitation events or conceal a mixture of distinct vertical structures.

Observational evidence suggests that this distinction matters. Near-surface rainfall intensity does not uniquely reflect the vertical development or strength of convection. The heaviest rainfall is only rarely associated with the most intense convection[22–24]; and relatively weak and intense convective systems contribute comparable numbers of extreme rainfall events over tropical and subtropical land[25]. Heavy rainfall can arise from shallow warm-rain convection dominated by collision–coalescence or from deep convection involving substantial ice-phase growth and melting[21,26–31]. Similar near-surface DSDs can also result from different combinations of microphysical processes [19,20,28,32]. Averaging reflectivity profiles over events with similar near-surface precipitation properties may therefore obscure distinct convective structures.

Advances in spaceborne radar observations and objective precipitation classification methods provide opportunities to resolve this hidden structural heterogeneity. Precipitation classification has progressed from heuristic threshold-based methods[33,34] to DSD-based clustering[20,28], while feature-based analyses of radar reflectivity profiles have demonstrated that vertical structural information beyond convective intensity can be extracted and used to distinguish precipitation regimes[35]. Recent global classifications based on GPM DPR data have revealed distinct vertical structures and microphysical characteristics across precipitation-system types[36]. However, it remains

poorly understood how these properties vary when precipitation events share comparable near-surface rain rates or DSDs.

A decade of GPM DPR observations is utilized to investigate the vertical structural heterogeneity of convection under similar near-surface precipitation states. We define convective precipitation events (CPEs) as heavy-precipitation cores embedded within larger precipitation systems and apply representation-learning-guided clustering to identify archetypal vertical structures from radar reflectivity-profile shapes. Based on this classification, we then quantify how these archetypes coexist within the same *nsRR* and near-surface D_m–$\log_{10}N_w$ regimes, examine their distinct vertical evolution of rain rate and drop size across the warm layer, and disentangle shifts in archetype composition from intrinsic microphysical changes within individual archetypes. Finally, we evaluate the extent to which regional differences in extreme rainfall arise from shifts in archetype prevalence versus changes in the conditional probability of extremes within each archetype. The overall study framework is illustrated in Fig. 1. By defining vertical structure archetypes independently of near-surface properties, this framework unravels the apparent near-surface homogeneity yet underlying structural diversity in convective precipitation, observational constraints and mechanistic insights to improve precipitation simulations.

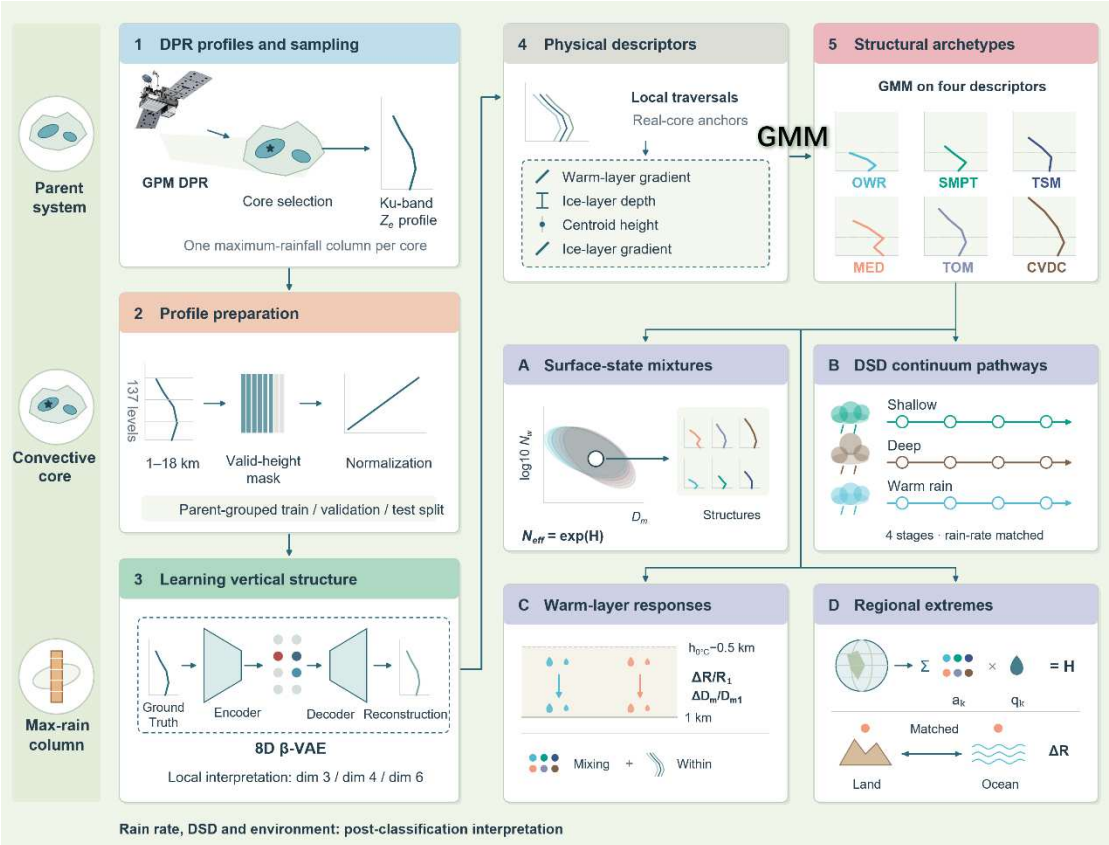


Figure 1. Overview of the study framework. Schematic of the Global Precipitation Measurement (GPM) Dual-frequency Precipitation Radar (DPR) analysis workflow,

from convective-core identification and β-variational autoencoder (β-VAE)-guided feature extraction to Gaussian mixture model (GMM) clustering and subsequent analyses of drop size distribution (DSD) mixtures and warm-layer microphysics.

Results

Decoupling vertical structure from near-surface states resolves six global convective archetypes

From approximately 1.64 million CPEs, we developed a representation-learning-guided framework to identify recurrent vertical structures. This framework first applied a β-variational autoencoder (β-VAE) to compress normalized Ku-band radar reflectivity (Z_e) profiles at peak rainfall and diagnose their dominant structural degrees of freedom via latent-space traversal. Gradient-aware reconstruction and valid-height constraints were incorporated to preserve the vertical geometry of the observed profiles. Guided by these latent modes, the framework then performed Gaussian mixture model (GMM) clustering on the corresponding physically interpretable profile descriptors, while deliberately withholding *nsRR*, DSD variables, and absolute reflectivity magnitude from the inputs. This design ensured that the classification was driven purely by vertical profile geometry.

The framework recovered six recurrent convective archetypes: oceanic warm rain (OWR; 6.95%), shallow mixed-phase transition (SMPT; 7.69%), tropical shallow mixed-phase (TSM; 24.64%), midlatitude embedded-deep (MED; 18.19%), tropical organized multicore (TOM; 28.01%) and continental very-deep convection (CVDC, c06; 14.53%) (Fig. 2; Table S2). Geographical descriptors in the archetype names refer to their preferential spatial occurrence rather than exclusive geographical domains. The archetypes span a structural spectrum mainly organized by vertical evolution and ice-phase involvement (Fig. 3a,b; Fig. S1c). OWR and CVDC occupy the shallow and deep ends, respectively, whereas SMPT, TSM, MED and TOM represent intermediate or alternative configurations of mixed-phase and deep convection. Spatial, environmental and parent-system characteristics vary systematically across the regimes, although none of these variables entered the primary classification (Figs. 2 and 4; Fig. S1d–f). These contrasts are integrated into a conceptual framework in Fig. 5.

The first part of the sequence describes increasing mixed-phase development. OWR has a low mean storm-top height (STH; 3.90 km), almost no effective ice-phase echo and a radar reflectivity centroid at 1.56 km (Fig. 3a; Table S2). It occurs preferentially over subtropical oceans and is associated with small, convectively dominated parent systems (Fig. 2a; Fig. S1d,f), consistent with the shallow, isolated oceanic precipitation documented by earlier radar observations[37]. SMPT extends the echo only slightly above the freezing level, with an STH of 5.23 km and an ice-cloud depth (ICD) of 0.66 km,

while the diagnosed warm-rain contribution remains dominant (Fig. 3a,b). Despite having the highest median relative humidity in the 900–800 hPa layer (Fig. 4d), it remains vertically confined, resembling congestus convection[38]. TSM represents a more developed mixed-phase configuration, with an STH of 6.84 km, an ICD of 1.93 km and a centroid height of 2.70 km (Table S2). The OWR–SMPT–TSM sequence is therefore characterized by progressively deeper ice-phase involvement, rather than by surface rain rate alone (Fig. S1a–c).

The sequence branches beyond TSM. MED has an STH similar to that of TSM (6.79 km) but a much deeper ice-phase layer (ICD, 3.39 km) (Fig. 3a; Table S2). Its cores are embedded within large, stratiform-dominated parent systems and occur preferentially along midlatitude storm tracks (Fig. 2d; Fig. S1d,f). Its environmental signature combines relatively low convective available potential energy (CAPE) with strong shear, moisture convergence and mid-tropospheric ascent, consistent with dynamically supported convection (Fig. 4a–c). Together with its embedded morphology, this is consistent with convection occurring within larger-scale ascending precipitation systems[39]. TOM forms a different deep-convective configuration. Its STH and ICD increase to 9.41 and 4.71 km, respectively, and its parent systems contain multiple convective cores (Fig. 3a; Fig. S1e; Table S2). Despite its greater vertical extent, TOM has a lower mean *nsRR* than MED, although their distributions overlap broadly (Fig. S1a,b; Table S2). This contrast is consistent with previous observations that heavier surface rainfall need not accompany taller or more intense convection[22,25].

CVDC occupies the deepest end of the distribution, with an STH of 12.15 km, an ICD of 7.53 km and a centroid height of 4.41 km (Fig. 3a; Fig. S1c; Table S2). It is concentrated over continental convective regions, peaks in warm seasons and local afternoon hours, and has the highest occurrence of heavy ice precipitation (HIP) and hail among the archetypes (Fig. 2f; Fig. S2f; Table S2). Its continental preference is consistent with the land-dominated distribution of extremely deep and intense convection documented by the Tropical Rainfall Measuring Mission (TRMM)[40]. Its environment combines the highest CAPE with lower column moisture than TSM and TOM, distinguishing it from the tropical deep-convective configurations (Fig. 4c; Fig. S4a,b). After accounting for local and seasonal background differences, all CAPE and total column water vapour (TCWV) anomaly scores remain small (Fig. 4e), indicating that the pronounced structural differences among the archetypes are not primarily controlled by anomalous instability or moisture.

Microphysical diagnostics further distinguish these structures. The diagnosed warm-rain contribution decreases from nearly unity in OWR to 0.480 in MED before increasing to 0.714 in TOM, whereas CVDC has the largest liquid- and ice-phase precipitation contributions simultaneously (Fig. 3b). Between 3 km and 1 km, the

fraction of samples showing joint increases in radar reflectivity and mass-weighted diameter decreases from 90.88% in OWR to 51.32% in CVDC, indicating different net warm-layer adjustments (Fig. 3c).

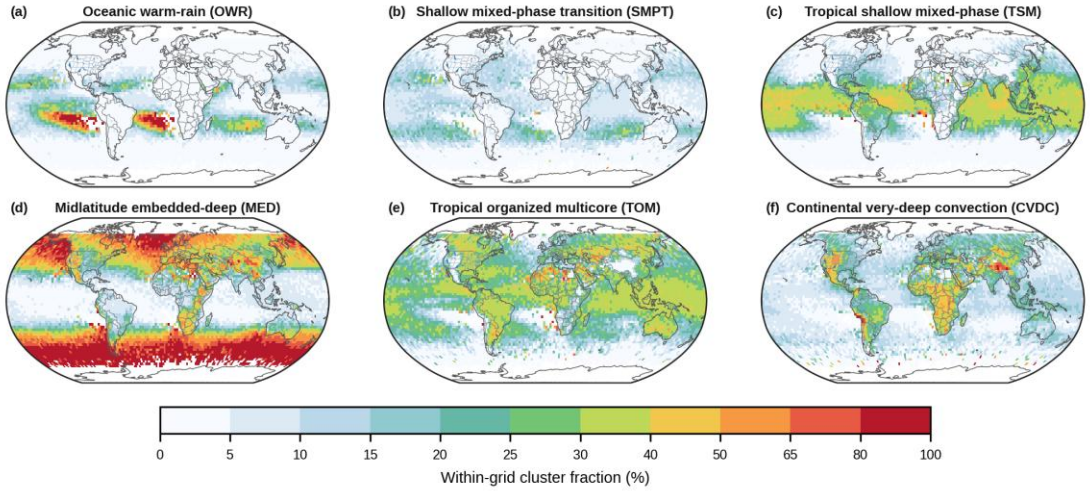


Figure 2. Fraction (%) of convective precipitation events (CPEs) assigned to each archetype within each 2.5° × 2.5° global grid cell. (a) Oceanic warm rain (OWR; n = 114,159; 6.9%). (b) Shallow mixed-phase transition (SMPT; n = 126,386; 7.7%). (c) Tropical shallow mixed-phase (TSM; n = 404,918; 24.6%). (d) Midlatitude embedded-deep (MED; n = 298,992; 18.2%). (e) Tropical organized multicore (TOM; n = 460,341; 28.0%). (f) Continental very-deep convection (CVDC; n = 238,859; 14.5%). The colour scale denotes the within-grid fraction of CPEs assigned to the indicated archetype.

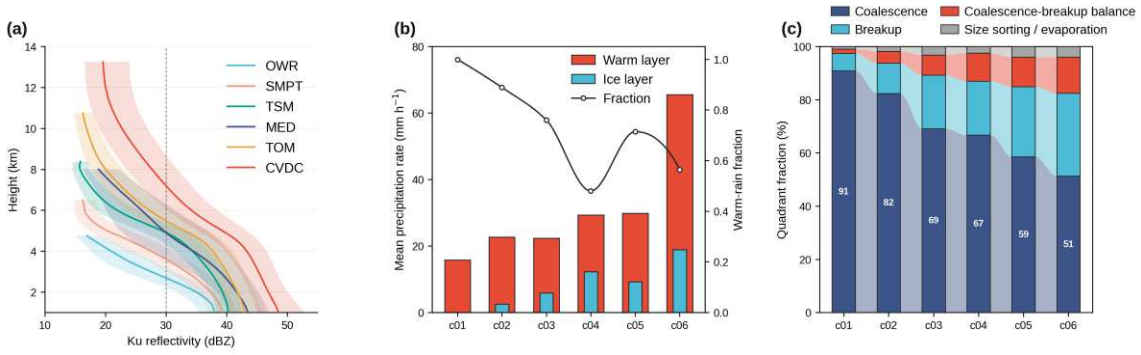


Figure 3. Structural and microphysical characterization of the six convective archetypes. (a) Median Ku-band radar reflectivity profile of the maximum-rain column versus height, truncated at each archetype's mean storm-top height (STH); shading shows the interquartile range (IQR; 25th–75th percentiles); the dashed line marks the convective threshold of 30 dBZ. (b) Warm-layer and ice-layer mean precipitation rates (bars) and the warm-rain fraction (line, right axis). (c) Fractions in the four quadrants of the joint ΔZ_e–ΔD_m distribution, with ΔZ_e and ΔD_m defined as the corresponding values at 1 km minus those at 3 km. The quadrant where both ΔZ_e and ΔD_m are positive corresponds to coalescence; where both are negative, to breakup; where $\Delta Z_e < 0$ and

$\Delta D_m > 0$, to size sorting or evaporation; and where $\Delta Z_e > 0$ and $\Delta D_m < 0$, to a coalescence–breakup balance.

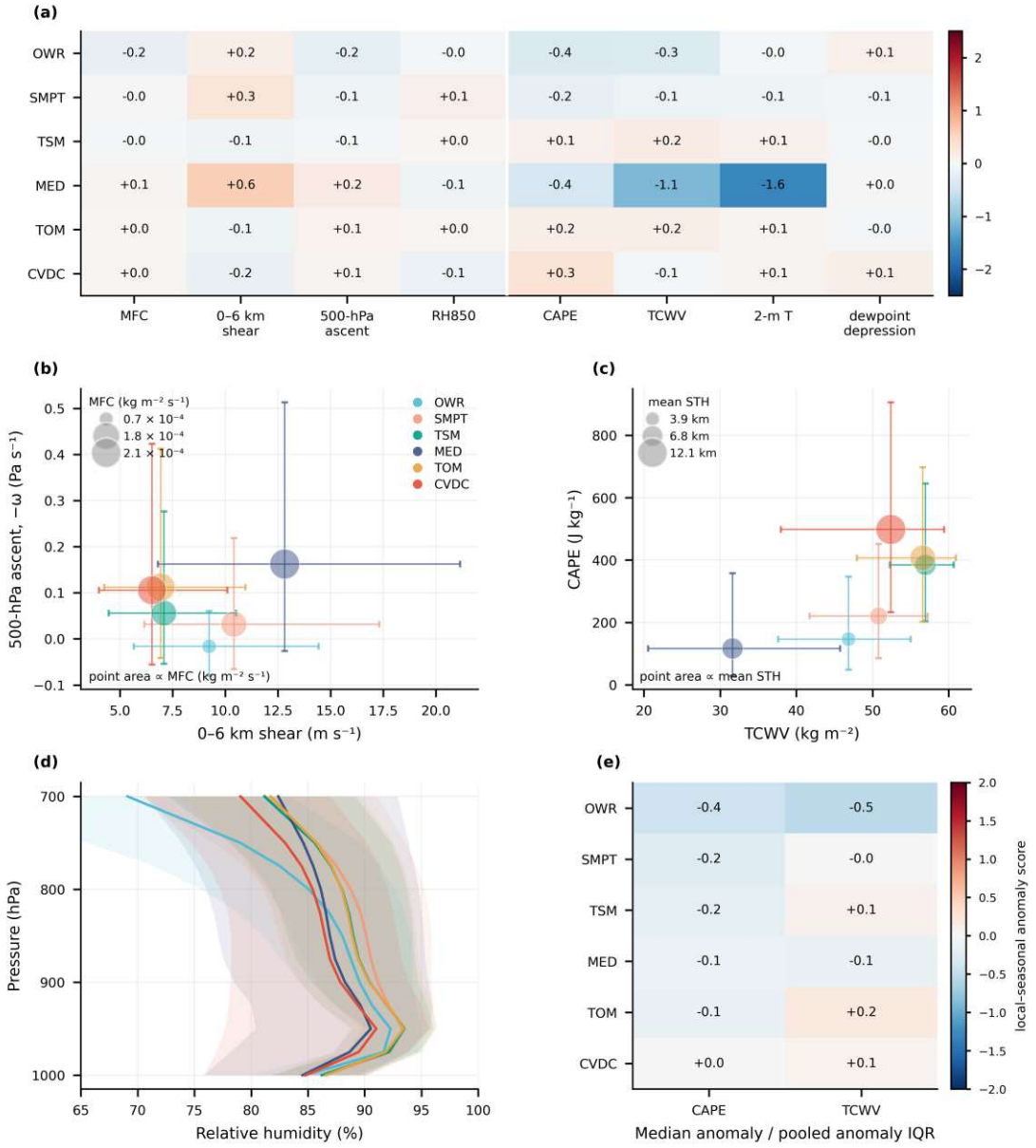


Figure 4. Environmental context of six convective archetypes from the fifth-generation reanalysis (ERA5) produced by the European Centre for Medium-Range Weather Forecasts. (a) Event-time environmental fingerprint expressed as the archetype median relative to the pooled median, scaled by the pooled interquartile range (IQR), for moisture-flux convergence (MFC; kg m^{-2} s^{-1}), 0–6-km shear (m s^{-1}), 500-hPa ascent represented by $-\omega 500$ (Pa s^{-1}), relative humidity at 850 hPa (RH850; %), convective available potential energy (CAPE; J kg^{-1}), total column water vapour (TCWV; kg m^{-2}), 2-m air temperature (K) and 2-m dewpoint depression (K). (b) Event-time shear–ascent configuration; point area is proportional to MFC. (c) Event-time TCWV–CAPE configuration; point area is proportional to mean STH (km). (d) Event-time relative-

humidity profiles, with lines showing archetype medians and shading showing the IQR. (e) Local–seasonal anomaly fingerprint for CAPE and TCWV, expressed as the archetype median anomaly divided by the pooled anomaly IQR. Error bars in (b,c) denote the IQR of the corresponding variables; colours identify the six archetypes. Larger positive robust scores indicate stronger, moister or warmer conditions, as appropriate for each variable.

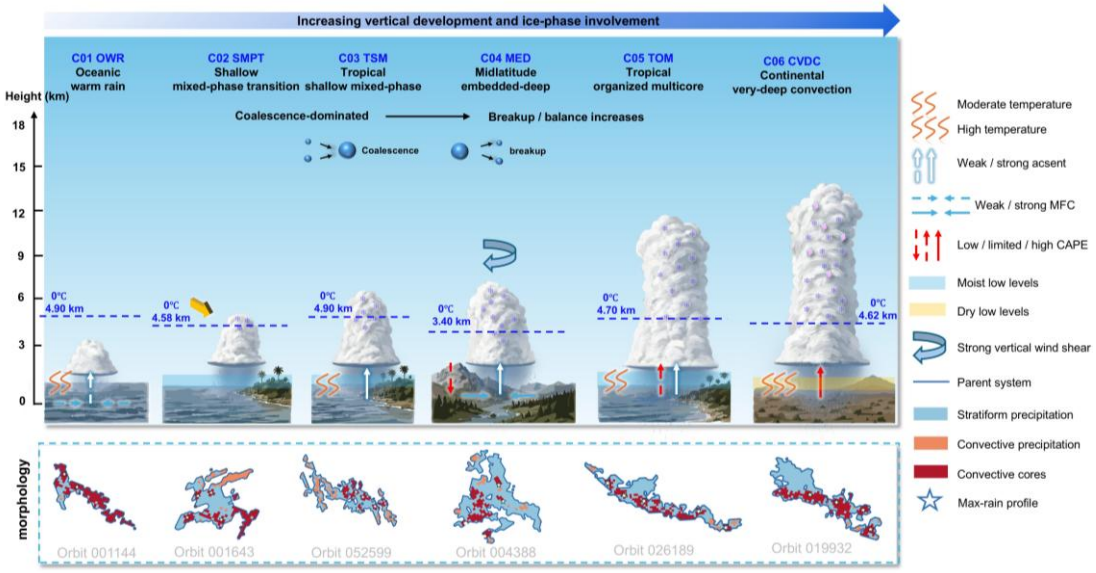


Figure 5. Structural archetypes of global convective precipitation events (CPEs). Conceptual representation of oceanic warm rain (OWR), shallow mixed-phase transition (SMPT), tropical shallow mixed-phase (TSM), midlatitude embedded-deep (MED), tropical organized multicore (TOM), and continental very-deep convection (CVDC), with vertical cloud extent scaled to their mean storm-top heights (STHs; 3.90, 5.23, 6.84, 6.79, 9.41, and 12.15 km, respectively). Dashed blue lines mark the corresponding mean 0 °C levels. Dark and light blue distinguish convective cores from their parent precipitation systems, respectively, while symbols indicate characteristic warm-rain, ice-phase, hail, multicore, and stratiform features. Horizontal dimensions and cloud morphology are schematic.

Near-surface DSD states conceal diverse convective structures

The six archetypes revealed distinct vertical structures and microphysical characteristics, but their near-surface DSDs largely overlapped in the D_m–$\log_{10}N_w$ space (Fig. 6a). This indicates that different precipitation structures may produce similar surface microphysical signatures. The modal bin (D_m = 1.6–1.7 mm, $\log_{10}N_w$ = 3.9–4.0; n = 24,701) included all archetypes, none accounting for more than 34% of events (Fig. 6a,c). At 1 km, radar reflectivity and rain rate were comparable and differences became apparent above 2 km. OWR had a mean 15-dBZ echo-top height (MAXHT15) of 3.8 km, while CVDC maintained 30-dBZ echoes up to 6.4 km and had a mean MAXHT15

of 11.7 km. MED had a lower echo top than CVDC, but about 3 km of its echo extended above the freezing level (Fig. 6b). Similar near-surface DSDs could arise from either a warm-rain archetype below the freezing level or a deep mixed-phase archetype with an echo top about three times as high. The vertical profiles within the modal DSD bin show clear differences in precipitation development across the melting level (Fig. S5). MED and CVDC have greater ice water content (IWC) above the freezing level, with relatively high rain rates and liquid water content (LWC) immediately below it. Their D_m is already close to its near-surface value and changes little farther downward. In contrast, OWR and SMPT have much weaker precipitation near the top of the warm layer, followed by rapid downward increases in rain rate, LWC and D_m. These features are consistent with a greater contribution from ice-phase growth and melting in deep convection, while shallow convection relies more on collision–coalescence within the warm layer[18,21,41]. The stronger warm-layer growth in shallow convection can compensate for its weaker precipitation supply from aloft, resulting in rain rates and DSDs comparable to those produced by deeper convection.

This overlap was not confined to the modal bin. Every DSD bin containing at least 100 CPEs included two or more archetypes, and 81% contained at least three archetypes, each contributing at least 10% of the events. We summarized this mixture using the entropy-based effective archetype number, N_{eff}, which equals 1 for a single-archetype bin and 6 for an equally mixed bin (Methods). The event-weighted mean N_{eff} was 4.11, rising to 4.83 in the modal bin. N_{eff} decreases towards higher *nsRR* contours, and, within a given rain-rate range, forms a band of elevated values at intermediate D_{m} before decreasing towards larger-D_{m}, lower-N_{w} states (Fig. 6d). *nsRR* shifted the archetype mixture without reducing it to a single archetype (Fig. 6e). OWR declined from 17% below 12.5 mm h^{-1} to near zero above 100 mm h^{-1}, while CVDC increased from 5% to 66% above 220 mm h^{-1}. The event-weighted N_{eff} fell only from 4.74 to 2.49, leaving even the most intense CPEs as mixtures dominated by TOM, MED and CVDC. Drop size showed the same pattern. Small drops favoured OWR (73% at $D_m = 0.85$ mm), whereas large drops favoured CVDC (57% at $D_m = 2.95$ mm), but both ends contained relatively few events. N_{eff} peaked at 4.42 near $D_m = 1.75$ mm, where TSM, MED, TOM and CVDC overlapped (Fig. 6f). The near-surface DSD represents a mixture of vertical structures, with *nsRR* and drop size modulating the archetype prevalence.

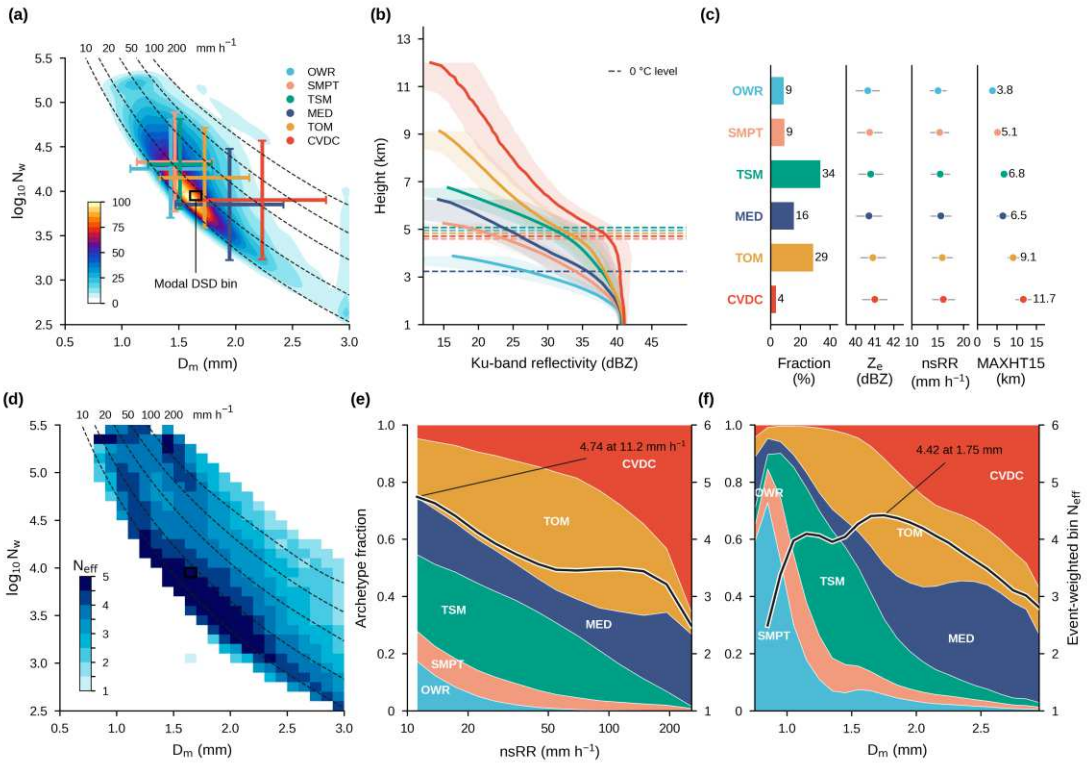


Figure 6. Characteristics and diversity of six precipitation archetypes. (a) Relative density of convective precipitation events (CPEs) in the 1-km drop size distribution (DSD) plane, normalized to its peak (%). Crosses show archetype means ±1 standard deviation (s.d.), the black square marks the modal DSD bin, and dashed curves indicate fitted rain-rate contours. (b) Median Ku-band radar reflectivity profiles of the six archetypes for 24,701 CPEs in the modal bin. Shading indicates the interquartile range, truncated at 12 dBZ, and colour-matched dashed lines mark the mean 0 °C levels. Statistics at each height include only CPEs with valid echoes. (c) Archetype fractions and mean 1-km equivalent radar reflectivity (Z_e), near-surface rain rate (nsRR) and 15-dBZ echo-top height (MAXHT15) in the modal bin. Grey lines indicate ±1 s.d. (d) Effective archetype number (N_{eff}, where H is Shannon entropy) across 0.1 × 0.1 DSD bins containing at least 100 CPEs. (e, f) Archetype fractions as functions of nsRR (e) and D_m (f), pooled across DSD bins containing at least 100 CPEs and across N_w, respectively. Black curves indicate the event-weighted mean N_{eff} (right axes), with annotations marking its peak values.

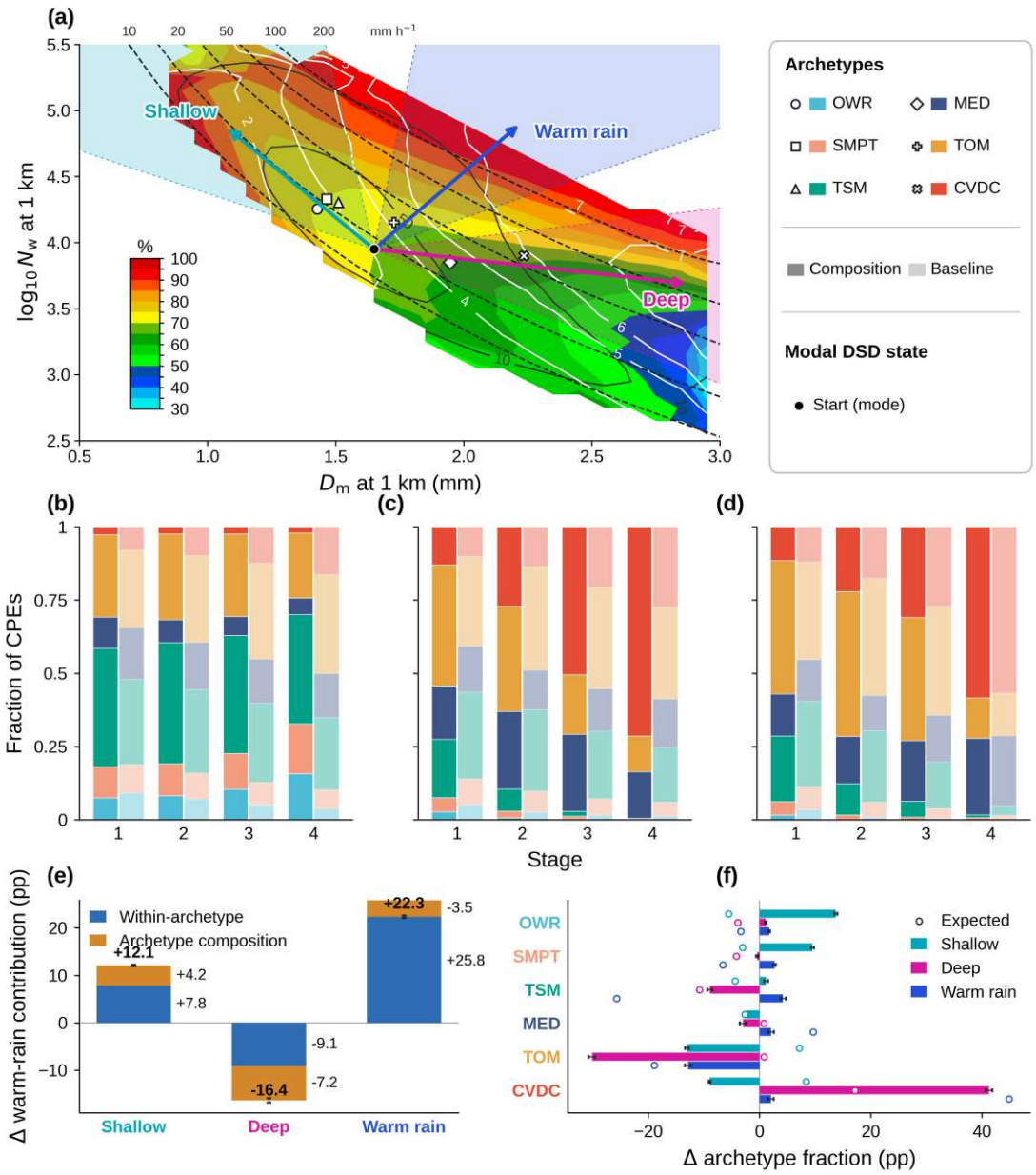
Movement through DSD space can arise from two different mechanisms: a redistribution among archetypes with different vertical structures, or enhanced precipitation production within existing archetypes. To distinguish these contributions, we examined three directions of DSD evolution: shallow convection, deep convection and warm-rain intensification[21]. These pathways were originally defined from the global summer GPM population based on changes in ICD and warm-rain contribution. Because our convective cores represent a different DSD population, we re-derived the

three pathways for our samples. The shallow-convection pathway extends toward decreasing ICD, the deep-convection pathway follows increasing ICD, and the warm-rain pathway follows the gradient of warm-rain contribution and approximately aligns with increasing *nsRR* in DSD space.

To quantify whether pathway evolution reflects changes in archetype composition beyond those expected from increasing rain intensity, we constructed a rain-rate-matched baseline. Each pathway was sampled within an angular sector drawn around it from the modal state (half-width 25°), and CPEs were separated into four stages according to their projected distance along the pathway. The observed archetype composition of each stage was compared with the composition expected from the full CPE population with the same *nsRR* distribution. Deviations from this baseline indicate pathway-specific changes in archetype prevalence that cannot be explained by increasing rain rate alone (Fig. 7b-d).

Both the shallow and deep pathways exhibit substantial archetype redistribution beyond that expected from rain-rate changes alone (Fig. 7b,c). Along the deep-convection pathway, CVDC increases from 13% to 71% of CPEs, exceeding the rain-rate-matched baseline by 41 percentage points, while TOM decreases by 30 percentage points relative to the baseline (Fig. 7c,f). This movement is associated with a transition toward deeper convective structures as expected. Similarly, along the shallow-convection pathway, OWR and SMPT become increasingly prevalent beyond the rain-rate expectation, increasing by 14 and 10 percentage points, respectively (Fig. 7b,f). Despite their opposite directions in DSD space, both pathways therefore represent systematic changes in convective archetype composition. In contrast, the warm-rain pathway shows a different behavior. Although CVDC increases from 12% to 58% along this pathway, this increase is almost entirely reproduced by the rain-rate-matched baseline (12% to 57%) (Fig. 7d,f). The apparent increase in CVDC along the warm-rain pathway mainly reflects its association with higher rain rates, rather than pathway-specific enrichment.

A further distinction emerges when considering the increase in warm-rain contribution along the warm-rain pathway. Decomposing its stage-1-to-stage-4 increase into archetype-mixing and within-archetype contributions (Fig. 7e; Methods), we find that the total increase of 22.3 percentage points is accounted for by a positive within-archetype term (+25.8 percentage points), partly offset by a negative composition term (−3.5 percentage points). Changes in archetype prevalence alone would tend to reduce the overall warm-rain contribution. The observed increase in warm-rain contribution therefore reflects changes within archetypes, rather than a shift toward convective structures that inherently favor warm rain.



316

317 **Figure 7.** Archetype composition along three drop size distribution (DSD) pathways.
318 (a) Warm-rain contribution (%) in the 1-km DSD plane for convective precipitation
319 events (CPEs) with near-surface rain rate (nsRR) ≥ 10 mm h^{-1}. Black contours indicate
320 CPE frequency, white contours indicate ice-cloud depth (ICD; km), and dashed curves
321 show fitted rain-rate contours. Symbols mark archetype means, and arrows indicate the
322 shallow-convection, deep-convection and warm-rain-intensification pathways
323 originating from the modal DSD state (black dot; D_m = 1.65 mm, $\log_{10}N_w$ = 3.95). (b–
324 d) Archetype fractions across four equal-count stages along the shallow-convection (b),
325 deep-convection (c) and warm-rain-intensification (d) pathways, compared with rain-
326 rate-matched baselines derived from the full CPE population. Observed fractions (left
327 bars) and baseline fractions (right bars) are shown for each stage, with median nsRR
328 (mm h^{-1}) and mean ICD (km) indicated above. (e) Change in warm-rain contribution

between stages 1 and 4, partitioned into compositional and within-archetype contributions. (f) Changes in archetype fractions between stages 1 and 4 beyond those expected from rain rate alone (bars), compared with the rain-rate-driven changes (open circles). Error bars indicate 95% confidence intervals from 500 day-block bootstrap resamples.

Archetypes reach heavy rain through distinct warm-layer routes that converge at extreme rain rates

To understand the processes driving the within-archetype changes, we examined how precipitation evolves through the warm layer across archetypes. We quantified this using the fraction of surface rain rate added within the warm layer ($\Delta R/R_1$) and the fractional growth of mass-weighted drop diameter ($\Delta D_m/D_{m1}$), evaluated between 0.5 km below the 0 °C level and 1 km above the surface. Simultaneous increases in both quantities indicate accretion-like behaviour, whereas simultaneous decreases indicate breakup-like behaviour (Fig. 8a).

At moderate rain rates, the archetypes follow two distinct warm-layer routes, distinguished by where precipitation mass is established and how the drop population evolves (Fig. 8a,b). OWR, SMPT and TSM follow a warm-rain route. Already at 10–15 mm h^{-1}, they gain most of their surface precipitation within the warm layer ($\Delta R/R_1$ = 0.94, 0.79 and 0.59), while D_m increases by 39%, 24% and 11%, respectively. Every OWR event lies in the accretion-like quadrant. This coupled increase is consistent with substantial precipitation growth below the melting level, through the collection of cloud droplets and collision–coalescence among raindrops. SMPT and TSM follow the same pathway but with a progressively greater contribution from precipitation formed aloft. MED and CVDC instead follow an ice-seeded route. At 10–15 mm h^{-1}, they gain only 9–10% of their surface precipitation within the warm layer, and D_m changes little ($\Delta D_m/D_{m1}$ = −0.01 and −0.02). Together with their relatively large drops entering the warm layer (Fig. 8c), these changes indicate that their rainfall is largely established before entering the warm layer, consistent with ice-phase growth followed by melting that supplies relatively large liquid drops. The weak net growth does not imply that coalescence is absent, but may reflect its effects being offset by breakup, evaporation or size sorting. TOM occupies an intermediate position ($\Delta R/R_1$ = 0.47, $\Delta D_m/D_{m1}$ = 0.06), suggesting a transition between warm-layer-dominated and ice-seeded precipitation growth rather than two mutually exclusive regimes. The two routes converge as rain intensifies. Above 130 mm h^{-1}, more than 95% of MED and CVDC events become accretion-like. Above 170 mm h^{-1}, all five archetypes with at least 500 events show 22–33% drop growth through the warm layer.

In the deep archetypes, the stronger warm-layer growth arises largely from changes in the entry state. From 10–15 to above 170 mm h^{-1}, the exit D_m remains nearly unchanged

in CVDC (2.28 to 2.27 mm) and increases only modestly in MED (1.96 to 2.08 mm). By contrast, the entering DSD shifts markedly. In CVDC, D_m decreases from 2.29 to 1.66 mm while $\log_{10}N_w$ rises from 3.09 to 4.23 (Fig. 8c). This shift is consistent with a greater supply of smaller drops available for collection by larger, ice-derived raindrops. Differences in fall speed allow the larger drops to sweep up these smaller drops, potentially favouring coalescence over breakup when liquid water is abundant[21,41]. This explains how deep-convective archetypes move from weak net drop growth at moderate rain rates to growth comparable with the other archetypes under extreme rainfall.

The population-average curve partly conceals these differences because the archetype mixture changes with rain rate (Fig. 6e). From 10–15 to above 170 mm h^{-1}, the CVDC share rises from 6% to 64%, whereas SMPT and TSM together decline from 44% to below 3%. We separated the population change into within-archetype and composition effects using a symmetric Kitagawa decomposition (Fig. 8d; Methods). The population-mean $\Delta D_m/D_{m1}$ increases by 0.160. Within-archetype intensification contributes +0.209, whereas the shift towards MED and CVDC contributes −0.049 because these archetypes show weaker warm-layer drop growth at a given rain rate. Below approximately 60 mm h^{-1}, the two effects nearly cancel, leaving the population curve almost flat even though every archetype increases its warm-layer drop growth. Thus, the population-average curve understates the strengthening of warm-layer drop growth within individual archetypes, particularly at moderate rain rates.

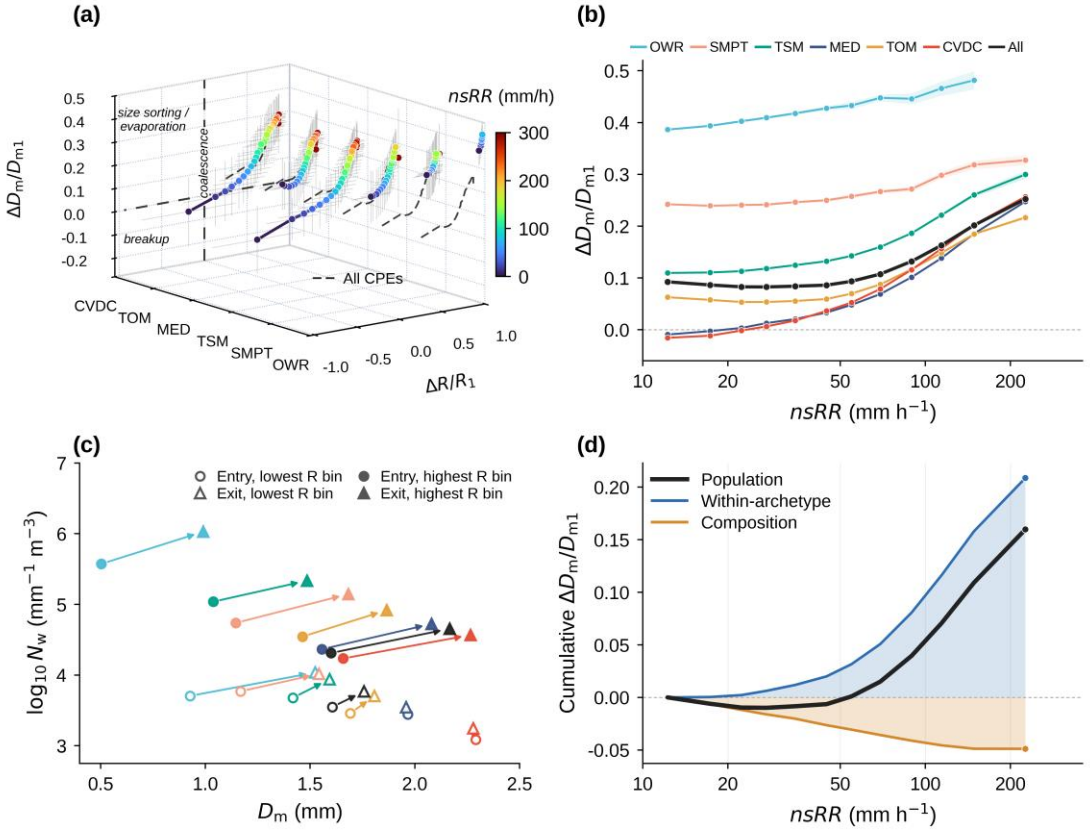


Figure 8. Archetype-dependent warm-layer pathways to heavy rain. (a) Relative changes in rain rate and mass-weighted mean drop diameter across the warm layer, stacked by archetype on a categorical axis. Points and connecting lines are coloured by near-surface rain rate (nsRR) from blue (low) to red (high); crosses indicate the within-bin means of $\Delta R/R_1$ and $\Delta D_m/D_{m,1}$. The black dashed curve shows the mean for all convective precipitation events (CPEs), repeated on every archetype layer. Dashed lines distinguish accretion-like (right), size sorting/evaporation (lower right), and breakup-like (lower left) regimes. (b) Relative change in drop diameter ($\Delta D_m/D_{m,1}$) against nsRR for each archetype and all CPEs. Archetype–bin combinations with fewer than 30 CPEs use the corresponding archetype's pooled mean. (c) Mean DSD states (D_m, $\log_{10}N_w$) at warm-layer entry (0.5 km below the 0 °C level; dashed lines) and exit (1 km above the surface; solid lines) across rain-rate bins ($R \geq 10$ mm h^{-1}) for the six archetypes (colours) and all CPEs (black). Arrows connect entry and exit states in the lowest and highest bins containing at least 30 CPEs. (d) Cumulative change in population-mean $\Delta D_m/D_{m,1}$ relative to the 10–15 mm h^{-1} bin (black), decomposed into within-archetype (blue) and compositional (orange) contributions. Open circles indicate the 10–15 mm h^{-1} bin, and filled circles mark the highest bin containing at least 30 CPEs. Shading indicates 95% confidence intervals from 500 day-block bootstrap resamples. N_w is expressed in mm^{-1} m^{-3}.

Regional rainfall extremes reflect competing archetype occurrence and within-archetype intensification

The preceding analyses revealed distinct pathways to heavy rainfall across convective archetypes and showed that changes in archetype composition can partly offset intensification within individual archetypes (Fig. 8). Previous studies have demonstrated that changes in the occurrence of organized deep convection contribute substantially to observed variations in tropical mean rainfall[42]. We therefore examined whether regional differences in the occurrence of extreme precipitation similarly reflect differences in which convective structures occur or in how frequently each structure produces extreme rainfall. The global domain was divided into tropical, subtropical and midlatitude bands, each separated into land and ocean. Using a common threshold of 72.5 mm h^{-1}, corresponding to the 99th percentile of event-maximum *nsRR* among all identified precipitation events during 2014–2023 (Methods), the fraction of selected CPEs exceeding this threshold ranged from 4.6% over midlatitude land to 15.7% over tropical oceans (global mean, 12.9%; Fig. 9a).

Despite higher extreme-rainfall fractions over oceans, the decomposition revealed archetype composition contributions favoring land across all latitude bands (Fig. 9b–d). The clearest reversal occurred in the tropics, where composition favored land by 7.1 percentage points, whereas within-archetype extreme-rainfall probability favored ocean

by 11.6 percentage points. At the archetype level, CVDC contributed positively through composition but negatively through within-archetype probability over tropical land, with the opposite pattern over tropical oceans. MED exhibited a similar pattern over tropical ocean. CVDC and MED accounted for most of the positive within-archetype contribution over tropical oceans (+5.1 of +6.6 percentage points). These results demonstrate that regional extreme-rainfall occurrence depends not only on the prevalence of convective structures but also on their region-dependent probabilities of producing extreme rainfall.

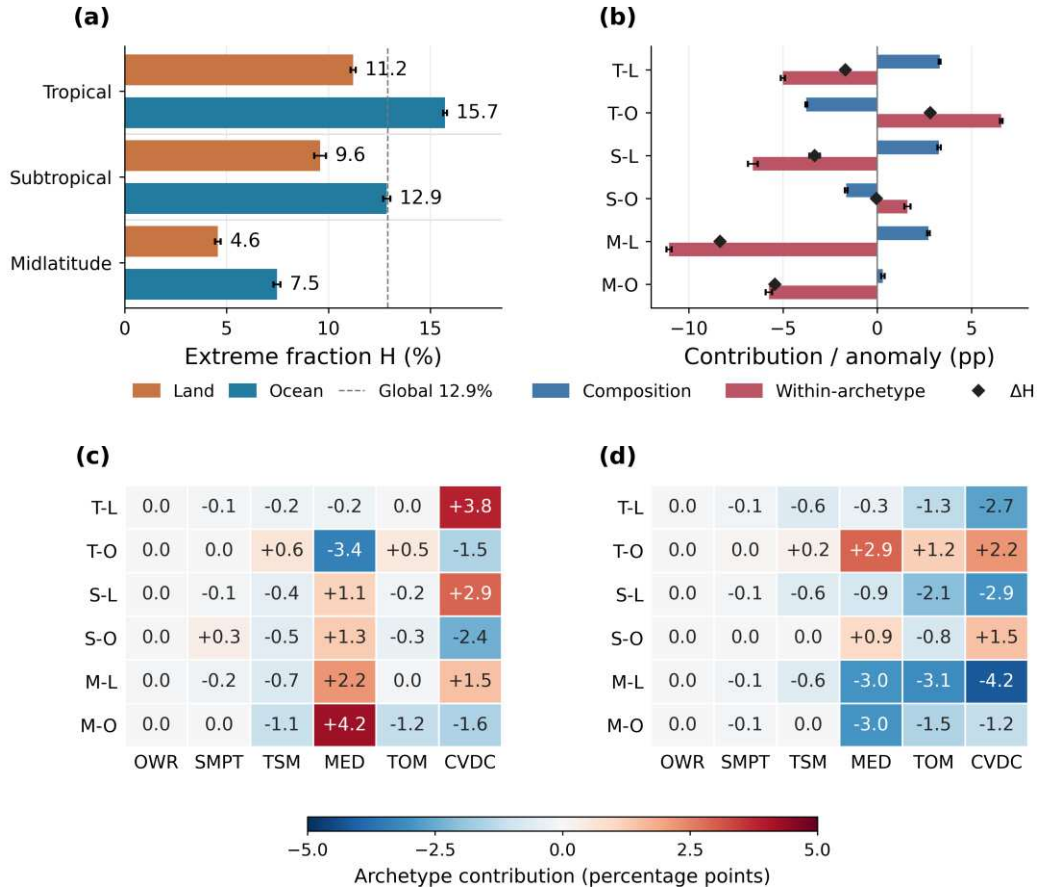


Fig. 9. Regional extreme-precipitation fractions and archetype contributions. (a) Extreme-precipitation fraction (H; %) across six latitude–surface regions; the dashed line marks the global reference H_0 = 12.91%. (b) Composition (blue) and within-archetype (rose) contributions to ΔH = H − H_0; black diamonds mark ΔH. (c,d) Corresponding archetype-specific contributions, expressed in percentage points with a shared zero-centred colour scale. Intervals in (a,b) indicate 95% parent-system block-bootstrap confidence intervals. Extremes have maximum core near-surface rain rate (nsRR) ≥ 72.5 mm h^{-1}. Latitude bands are tropical (<23.5°), subtropical (23.5–35°) and midlatitude (≥35°), using absolute latitude. Region abbreviations in (b–d): T, tropical; S, subtropical; M, midlatitude; L, land; O, ocean.

To investigate whether within-archetype contrasts in extreme rainfall are accompanied by differences in warm-layer precipitation evolution, we compared land and ocean CPEs at comparable warm-layer entry rain rates and within common latitude–month support (Fig. 10). Oceanic CPEs exhibited greater net rain-rate growth, particularly within MED, TOM and CVDC, with a difference of approximately 56 mm h^{-1} in CVDC at entry rates of 20–30 mm h^{-1}. Consistent with earlier TRMM observations of downward-increasing radar reflectivity over ocean[43], this contrast persisted within broadly defined vertical-structure archetypes. Within the same 20–30 mm h^{-1} entry-rate bin, extreme-rainfall probabilities were also substantially higher over ocean than land for CVDC (58.81% versus 26.00%; difference: 32.82 percentage points, 95% parent-block confidence interval: 31.79–33.86) and MED (38.92% versus 25.17%; difference: 13.75 percentage points, 95% parent-block confidence interval: 12.35–15.09). Nevertheless, supplementary radar reflectivity profiles reveal residual structural differences between continental and oceanic CVDC (Fig. S6), indicating that the observed contrasts cannot be attributed exclusively to warm-layer processes. Beyond these residual structural differences, environmental conditions may also modulate precipitation microphysics. Recent GPM-based studies have highlighted the roles of large-scale circulation, aerosol loading and meteorological conditions in shaping regional variations in raindrop characteristics and vertical precipitation profiles[44,45].

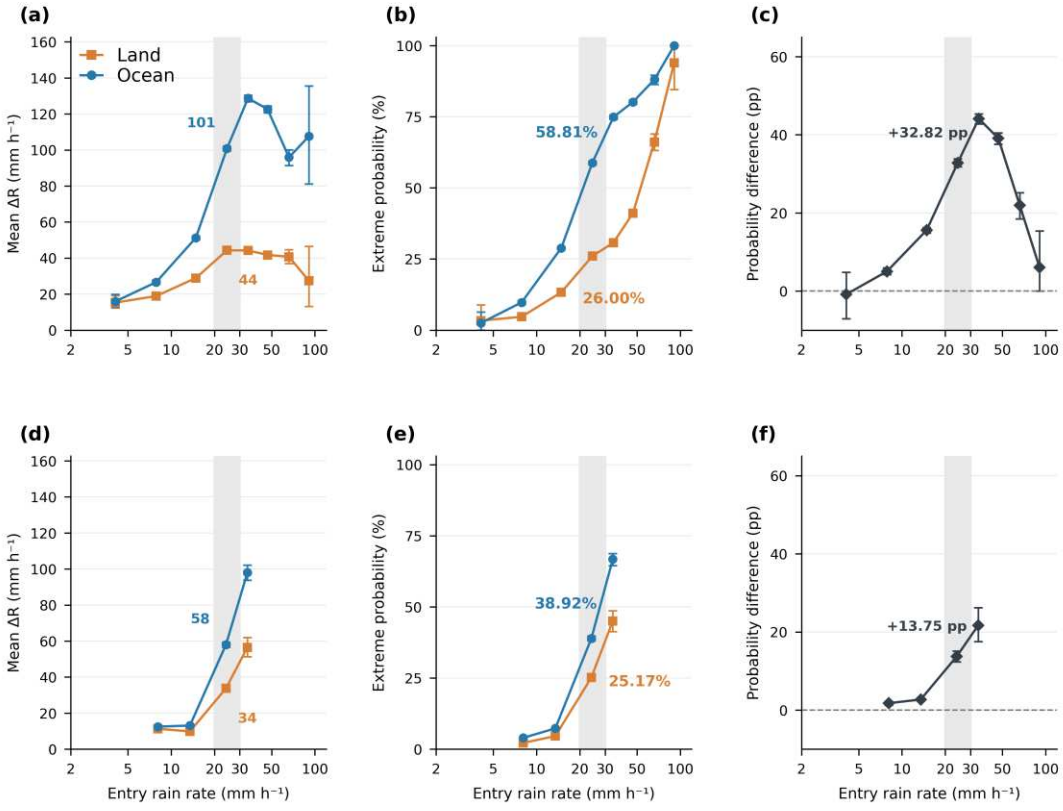


Fig. 10. Conditional land–ocean contrasts in warm-layer rain-rate growth and extreme-

rainfall probability within continental very-deep convection (CVDC) and midlatitude embedded-deep convection (MED). (a,d) Mean net rain-rate growth ($\Delta R = R_1 - R_entry$; mm h^{-1}), from 0.5 km below the 0 °C level to 1 km above the surface in the maximum-rainfall column, shown as a function of the entry rain rate, over land (orange squares) and ocean (blue circles), for CVDC and MED, respectively. (b,e) Probability of maximum core near-surface rain rate (nsRR) exceeding 72.5 mm h^{-1}, conditioned on the same entry-rate bins, over land and ocean as in (a,d). (c,f) Ocean-minus-land probability differences in percentage points (pp). Each point summarises one entry-rate bin (plotted at the pooled median entry rate), so the curves trace how growth and extreme probability evolve with increasing entry rain rate; estimates use common latitude–month weights, and error bars indicate pointwise 95% parent-system block-bootstrap intervals. Bold numbers give the values for the highlighted 20–30 mm h^{-1} entry-rate bin (grey bands). Dashed lines mark zero.

Discussion

This study reveals substantial vertical structural and microphysical heterogeneity concealed within similar near-surface precipitation states. Using a decade of GPM DPR observations encompassing over 1.6 million CPEs, we identify six recurrent vertical-structure archetypes through a representation-learning-guided clustering driven purely by profile geometry, excluding near-surface rain rate, DSD variables, and reflectivity magnitude. Structurally distinct archetypes coexist within the same rain-rate and DSD states, spanning shallow warm-rain convection to very deep convection with extensive ice-phase development. Near-surface DSD states therefore cannot be read as a unique fingerprint of precipitation formation aloft. This structural ambiguity has clear microphysical significance. At comparable rain rates, shallow archetypes exhibit substantial warm-layer rainfall and drop-size growth, whereas deep archetypes show weak net drop-size growth at moderate intensities before strengthening as rainfall intensifies. These contrasting vertical precipitation-growth signatures reveal information that surface-conditioned averages alone cannot resolve. This structural heterogeneity also has implications for interpreting precipitation variability and extremes. Changes in archetype composition can obscure intrinsic microphysical variations in population-mean relationships, while regional extreme-rainfall occurrence depends not only on the prevalence of convective structures but also on region-dependent variations in extreme-rainfall probability within each archetype. The stronger warm-layer rainfall growth observed over oceans within several deep archetypes further highlights the importance of processes operating below the melting level.

Our findings can provide an observational basis for evaluating whether satellite retrievals and numerical models reproduce the diversity of vertical structures and warm-

layer responses underlying similar surface precipitation states. Nevertheless, the archetypes depend on classification resolution, while DPR retrieval uncertainties, particularly at higher intensities as indicated by regional gauge-based evaluations[46], and the indirect nature of vertical microphysical diagnostics limit process-level attribution. Future integration of doppler radar observations, storm life-cycle tracking and convection-permitting simulations could help clarify the underlying microphysical mechanisms and assess how structural diversity influences precipitation extremes in a changing climate.

Methods

Data

The GPM DPR provides state-of-the-art observations of global precipitation microphysics across 65°S–65°N[47]. The version 07 GPM DPR Level-2A (2ADPR) precipitation products[48,49], spanning from April 2014 to November 2023, is used in this study. We utilize the profiles of attenuation-corrected Ku-band radar reflectivity factor Z_e (dBZ), rain rate R (mm h^{-1}), D_m (mm), and N_w ($\log_{10}$ mm^{-1} m^{-3}). Additional variables include *nsRR*, rain-type classification, hail and HIP flags, and the 0 °C isotherm height.

Environmental conditions were obtained from the ERA5 reanalysis, comprising single-level fields (2-m temperature T_{2m} (K), 2-m dewpoint temperature T_d (K), and CAPE) alongside pressure-level variables (specific humidity q and horizontal wind components u, v). These environmental parameters were paired with the spatial centroid and nearest time of each CPE through spatiotemporal nearest-neighbor matching.

ERA5 environmental diagnostics

RH850 was taken from ERA5 relative humidity at 850 hPa, and lower-tropospheric humidity profiles used available levels between 1000 and 700 hPa. Mid-tropospheric ascent was expressed as $-\omega_{500}$ (Pa s^{-1}), so positive values denote upward motion. Dewpoint depression was calculated as 2-m air temperature minus 2-m dewpoint temperature.

MFC (kg m^{-2} s^{-1}) represents horizontal convergence of vertically integrated specific-humidity transport between 1000 and 700 hPa. Its continuous definition is

$$MFC = -\frac{1}{g}\int_{700\,\mathrm{hPa}}^{1000\,\mathrm{hPa}} \nabla_h \cdot (q\mathbf{V}_h)\,dp \tag{1}$$

where q is specific humidity, $V_h = u, v$ is horizontal wind, g is gravitational acceleration and pressure is expressed in pascals. Positive MFC denotes net moisture convergence. Deep-layer shear was represented by the magnitude of the vector wind difference across

the nominal 0–6-km layer, rather than the difference in scalar wind speed.

$$S_{0-6} = \left[\left(u_6 - u_0 \right)^2 + \left(v_6 - v_0 \right)^2 \right]^{1/2} \quad (2)$$

For the environmental fingerprint in Fig. 4a, each archetype median was centred on the pooled CPE median and divided by the pooled IQR. Humidity profiles and two-variable diagrams report medians and IQRs.

For CAPE and TCWV in Fig. 4e, local–seasonal anomalies were calculated relative to event-time means pooled across archetypes and years. Reference groups used latitude rounded to 5°, longitude rounded to 10° and calendar month. Groups required at least 30 valid events. Otherwise, the reference fell back successively to latitude–month, month-only and finally the pooled global event mean. Each event value was reduced by its reference mean; the archetype median anomaly was then divided by the pooled anomaly IQR. These anomalies remove background differences within the sampled event population.

Nested CPE dataset (N-CPE)

Valuable progress has been made using pixel-level connected-component analyses [17,40,50], gridded spatial composites[28,51], and feature-based event identification[52,53]. Recent precipitation-feature analyses have further highlighted the relationship between convective-core organization and extreme instantaneous rainfall[54]. Building on these perspectives, the analyses in this study are event-based, using a nested dataset built from GPM DPR observations (2014–2023). Each heavy precipitation core, embedded within a parent system, is identified as a discrete CPE.

A parent precipitation system is an 8-connected region of at least four pixels with a *nsRR* more than 0.5 mm h^{-1}. Within a parent system, an 8-connected region of at least four convective pixels with a maximum *nsRR* above 10 mm h^{-1} is extracted as a heavy precipitation core. Multiple convective precipitation cores can be embedded within a single parent system. Cores touching the across- or along-track swath edge are removed, while truncated parent systems are kept. Cores are compact, so edge filtering changes core statistics little, whereas discarding truncated parents would remove most large systems. Parent-system metrics thus describe the storm within the swath rather than its full extent. 1,645,892 heavy precipitation events (cores) are retained for subsequent analysis (Table S2).

Three analytical tiers are constructed for each event: the peak-precipitation column, core-averaged diagnostics, and parent-system morphology. The tiers differ in sampling rather than in the variables they use. The peak column follows the established practice of representing a storm by its most intense element [55] and is the input to the

classification. Core-averaged profiles suppress single-pixel noise and characterize the core as a whole, whereas parent-system metrics locate the core within the surrounding precipitation area [40,56]. Because only the peak column enters the classification, the linkages among convective intensity, core microphysics and storm morphology are measured rather than prescribed.

Extreme convective precipitation is defined by a percentile threshold rather than by a fixed rain rate, following established practice in spaceborne radar studies[55,57]. We use the 99th percentile of the maximum *nsRR* over all DPR-detectable events from 2014 to 2023, which corresponds to 72.50 mm h^{-1}. Cores above this value are classified as extreme. The threshold is computed globally, so the same intensity defines an extreme everywhere. Regional differences in extreme frequency are preserved. We do not use the 99.9th percentile (299.74 mm h^{-1}), because the 2ADPR *nsRR* saturates near 300 mm h^{-1}. A class defined there would reflect the retrieval ceiling, not rainfall intensity.

Representation learning and structural clustering

Each Ku-band radar reflectivity profile is normalized by its maximum value to emphasize vertical structure rather than absolute radar reflectivity. Adjacent range bins are highly correlated, so the normalized profiles should be compressed before clustering[35,36]. Because the degrees of freedom are unknown, we first determine them from the data with an autoencoder and then build the feature set from physical metrics spanning them, which are inputs to the clustering.

The β-VAE[58] encodes each profile into a low-dimensional Gaussian latent space and reconstructs the profile from a sample of it. Training minimizes reconstruction error together with the Kullback–Leibler divergence between the posterior and a standard normal prior, weighted by β. We set β to 10^{-5}; without this term, the latent space collapses towards isolated points and traversal yields no coherent variation, whereas larger values degrade reconstruction. An eight-dimensional latent space was chosen at the elbow of the reconstruction root-mean-square error (RMSE; 0.62 dBZ) in Fig. S7. The full model setup is given in Text S1.

Anchor profiles were encoded into the latent space, each dimension was traversed with the others held fixed, and the physical meaning of a dimension was inferred from the resulting reconstructions (Text S2). Shape variation concentrates on three dimensions. These control vertical development and the partitioning of precipitation growth between ice-phase and warm-rain processes (Text S3, Table S4, Figs. S8 and S9).

These three degrees motivate the feature set used for the primary classification: ICD, ice-layer radar reflectivity gradient (G_{ice}), warm-layer radar reflectivity gradient (G_{warm}) and radar reflectivity centroid height (H_c) (Table S3). They are computed from the original profile. Unlike latent coordinates, whose response depends on the anchor at

which the traversal is performed (Text S3, Figs. S8 and S9, Table S4), these metrics have the same meaning throughout the sample and are directly comparable with ground-based and airborne observations.

Classification of surface DSD has moved from discrete types towards a continuum[19–21]. Vertical structure grades in the same way, but its classification has not followed. We therefore classify profiles with data-driven probabilistic clustering, using a GMM[59], which represents the sample as a mixture of Gaussian components and returns a posterior membership probability for each profile, so that intermediate structures are retained. The number of components k is chosen from the elbow of the Akaike and Bayesian information criteria[60,61], together with bootstrap stability measured by the adjusted Rand index (ARI). Because the solution depends on initialization, initial centroids are set by k-means++[62]. It gives $k = 6$ for the physical features, with a median posterior probability of 0.86 and a mean pairwise bootstrap ARI of 0.86, and $k = 5$ for both latent representations (Fig. S10). The full model setup is given in Text S4. Physical evaluation of the six archetypes is presented in the Results section.

DSD-space mixture and information diagnostics

Each CPE contributed one max-profile DSD coordinate and one archetype label. We used D_m and $\log_{10}N_w$ at 1 km within the maximum-rain profile. The DSD plane was divided into 0.1 × 0.1 bins over $0.5 \le D_m < 3.0$ mm and $2.5 \le \log_{10}N_w < 5.5$. The six structural archetypes were derived independently from the vertical profile shape, and neither DSD variables nor rain rate entered the clustering. For DSD bin b, let n_{bk} denote the number of CPEs assigned to archetype k, and let $n_b = \Sigma_k\, n_{bk}$. The conditional archetype fraction was calculated as $p_{bk} = n_{bk}/n_b$. We quantified the local archetype diversity using the Shannon entropy:

$$H_b = -\sum_{k=1}^{6} p_{bk} \ln p_{bk} \tag{3}$$

where terms with $p_{bk} = 0$ were set to zero. The entropy-based effective archetype number was then defined as

$$N_{\mathrm{eff},b} = \exp(H_b) \tag{4}$$

Thus, $N_{eff,b} = 1$ indicates complete dominance by one archetype, whereas $N_{eff,b} = 6$ corresponds to six equally represented archetypes. Importantly, N_{eff} is an entropy-equivalent diversity measure, rather than a literal integer count of archetypes. To summarize the diversity experienced by the observed CPE population, we calculated the CPE-weighted mean using all non-empty DSD bins:

642 $$\bar{N}_{\mathrm{eff}}^{\mathrm{CPE}} = \frac{\sum_b n_b N_{\mathrm{eff},b}}{\sum_b n_b} \quad (5)$$

643 **Decomposing compositional and within-archetype contributions**

644 Changes in population-level precipitation characteristics may arise from shifts in
645 archetype prevalence or changes in the response within individual archetypes.
646 Following regime-based decompositions of rainfall errors and cloud feedbacks[63,64], we
647 separated these contributions using the symmetric Kitagawa decomposition[65]. For a
648 group g, let p_{gk} denote the fraction of CPEs belonging to archetype k, and x_{gk} its
649 conditional mean response. The population mean is $Y_{\mathrm{g}} = \Sigma_{\mathrm{k}}\, p_{\mathrm{gk}}\, x_{\mathrm{gk}}$.

650 For two groups A and B, their difference can be decomposed exactly as

651 $$Y_A - Y_B = \underbrace{\sum_k \frac{p_{Ak} + p_{Bk}}{2}(x_{Ak} - x_{Bk})}_{\text{Within-archetype contribution}} + \underbrace{\sum_k \frac{x_{Ak} + x_{Bk}}{2}(p_{Ak} - p_{Bk})}_{\text{Composition contribution}}. \quad (6)$$

652 For warm-layer adjustment (Fig. 8d), groups A and B represent adjacent *nsRR* bins,
653 with A having the higher rain rate. Here, p_{jk} is the CPE fraction of archetype k in rain-
654 rate bin j, and x_{jk} is its mean fractional change in mass-weighted drop diameter $\Delta D_{\mathrm{m}}/D_{\mathrm{m1}}$
655 or rain rate $\Delta R/R_1$ across the warm layer. Contributions were accumulated from the 10–
656 15 mm h^{-1} bin through successive bins with boundaries at 15, 20, 25, 30, 40, 50, 60, 80,
657 100, 130, 170 and 300 mm h^{-1}. Archetype–bin combinations containing fewer than 30
658 CPEs were assigned the nearest valid mean of the same archetype, based on bin-index
659 distance.

660 For regional extreme-precipitation frequency (Fig. 9), groups A and B represent a region
661 r and the global reference population (denoted by 0), respectively. CPEs were divided
662 into six land–ocean regions across tropical |latitude| < 23.5°, subtropical 23.5° ≤
663 |latitude| < 35° and midlatitude |latitude| ≥ 35°bands. Extreme precipitation was defined
664 by a maximum core *nsRR* ≥ 72.5mm h^{-1}. For region r and archetype k, let N_{r} be the
665 total CPE count, n_{rk} the archetype count and e_{rk} the extreme-event count. We set $p_{\mathrm{rk}} =$
666 $n_{\mathrm{rk}}/N_{\mathrm{r}}$ and $x_{\mathrm{rk}} = e_{\mathrm{rk}}/n_{\mathrm{rk}}$, so that the regional extreme fraction is $H_{\mathrm{r}} = \Sigma_{\mathrm{k}}\, p_{\mathrm{rk}}\, x_{\mathrm{rk}}$. The global
667 reference included all shape-classified CPEs, including those in region r. Applying the
668 same decomposition to $H_{\mathrm{r}} - H_0$ yielded archetype-specific composition and within-
669 archetype contributions (Fig. 9c,d), expressed in percentage points.

670 **Clarification of convection-related terms.**

671 We specify the meaning of convection-related terms used in the text as follows.

672 a) Convection. Precipitation systems driven by atmospheric instability with significant

vertical motion, ranging from isolated cells to organized convective systems.
b) CPE. The analysis unit of this study: a heavy precipitation core together with the parent precipitation system in which it is embedded, identified as defined above.
c) Convective structure. The architecture of a CPE in the vertical and in the horizontal.

Vertical structure is the organization of a precipitation column with height, expressed through the vertical profiles of radar variables, of which the radar reflectivity profile is the basis. A radar reflectivity profile carries two kinds of information: its shape, including the layering, the radar reflectivity gradients, and the centroid height; and its magnitude, the radar reflectivity level reached within each layer. Horizontal structure is the morphology of the core and its parent system. d) Convective intensity. The strength of convective updrafts, proxied in this study by ICD (the height of the 15 dBZ level above the freezing level in the maximum-rain profile). e) Convective archetype. A class of CPEs defined from vertical shape, and characterized by its spatiotemporal distribution, microphysical processes, the morphology of its core and parent system, and its DSD characteristics. Membership is probabilistic, so that an individual CPE can lie between archetypes rather than belonging exclusively to one.

Data availability

The source observations are the version 07A GPM DPR Level-2A product, available from the National Aeronautics and Space Administration (NASA) GPM data-access page: https://gpm.nasa.gov/data-access/downloads/gpm. The environmental fields are from the ERA5 reanalysis, available through the Copernicus Climate Data Store for single levels at https://cds.climate.copernicus.eu/datasets/reanalysis-era5-single-levels and pressure levels at https://cds.climate.copernicus.eu/datasets/reanalysis-era5-pressure-levels. The derived N-CPE, together with the event-level summary data generated in this study, is available from the corresponding author upon reasonable request.

Code availability

The code for constructing the N-CPE, extracting the maximum-rainfall profile descriptors, fitting the GMMs, calculating the archetype-composition statistics and reproducing the figures is available from the corresponding author upon reasonable request. The analysis workflow will be deposited in a public repository with a persistent identifier upon publication.

Acknowledgements

This work was supported by the National Natural Science Foundation of China (Grant No. 42305150). The funder had no role in the study design, data collection, data analysis, interpretation of results, or preparation of the manuscript. We thank the NASA

GPM mission team and the Copernicus Climate Change Service for providing the data used in this study.

Author contributions

X.H. and Z.W. contributed equally to this work and should be considered co-first authors. X.H. and Z.W. designed the study. Z.W. processed the data, performed the analyses and wrote the manuscript. X.H. and Z.W. generated and discussed the results. X.H., W.A. and W.Z. supervised the study. S.M., H.P., and Z.T. contributed to discussions and manuscript revision. All authors reviewed and approved the final manuscript.

Competing interests

The authors declare no competing interests.

References

1. Matveev, L. T. *Cloud Dynamics*. (Springer Netherlands, Dordrecht, 1984). doi:10.1007/978-94-009-6360-3.
2. Beard, K. V., Ochs, H. T. & Kubesh, R. J. Natural oscillations of small raindrops. *Nature* **342**, 408–410 (1989).
3. Rosenfeld, D. & Ulbrich, C. W. Cloud Microphysical Properties, Processes, and Rainfall Estimation Opportunities. *Meteorol. Monogr.* **30**, 237–237 (2003).
4. Villermaux, E. & Bossa, B. Single-drop fragmentation determines size distribution of raindrops. *Nat. Phys.* **5**, 697–702 (2009).
5. Kummerow, C. D. *et al.* The Evolution of the Goddard Profiling Algorithm to a Fully Parametric Scheme. *J. Atmos. Ocean. Technol.* **32**, 2265–2280 (2015).
6. Liao, L. & Meneghini, R. GPM DPR Retrievals: Algorithm, Evaluation, and Validation. *Remote Sens.* **14**, 843 (2022).
7. Morrison, H., Thompson, G. & Tatarskii, V. Impact of Cloud Microphysics on the Development of Trailing Stratiform Precipitation in a Simulated Squall Line: Comparison of One- and Two-Moment Schemes. *Mon. Weather Rev.* **137**, 991–1007 (2009).

8. O'Gorman, P. A. Precipitation extremes under climate change. *Curr. Clim. Change Rep.* **1**, 49–59 (2015).
9. Donat, M. G., Lowry, A. L., Alexander, L. V., O'Gorman, P. A. & Maher, N. More extreme precipitation in the world's dry and wet regions. *Nat. Clim. Change* **6**, 508–513 (2016).
10. Prein, A. F. *et al.* The future intensification of hourly precipitation extremes. *Nat. Clim. Change* **7**, 48–52 (2017).
11. Papalexiou, S. M. & Montanari, A. Global and Regional Increase of Precipitation Extremes Under Global Warming. *Water Resour. Res.* **55**, 4901–4914 (2019).
12. Abdelmoaty, H. M., Papalexiou, S. M., Rajulapati, C. R. & AghaKouchak, A. Biases Beyond the Mean in CMIP6 Extreme Precipitation: A Global Investigation. *Earths Future* **9**, e2021EF002196 (2021).
13. Chen, B., Hu, Z., Liu, L. & Zhang, G. Raindrop Size Distribution Measurements at 4,500 m on the Tibetan Plateau During TIPEX-III. *J. Geophys. Res. Atmos.* **122**, (2017).
14. Wen, L., Chen, G., Yang, C., Zhang, H. & Fu, Z. Seasonal variations in precipitation microphysics over East China based on GPM DPR observations. *Atmos. Res.* **293**, 106933 (2023).
15. Sun, N., Lu, G. & Fu, Y. Microphysical characteristics of precipitation within convective overshooting over east China observed by GPM DPR and ERA5. *Atmos. Chem. Phys.* **24**, 7123–7135 (2024).
16. Wang, Z., Hu, X., Ai, W., Qiao, J. & Zhao, X. Microphysical Characteristics of Monsoon Precipitation over Yangtze-and-Huai River Basin and South China: A Comparative Study from GPM DPR Observation. *REMOTE Sens.* **16**, 3433 (2024).

17. Hu, X. *et al.* Microphysics of Summer Precipitation Over Yangtze-Huai River Valley Region in China Revealed by GPM DPR Observation. *Earth Space Sci.* (2022).
18. Shi, R., Lu, C., Xu, W. & Luo, Y. A global view on microphysical discriminations between heavier and lighter convective rainfall. *Commun. Earth Environ.* **6**, 511 (2025).
19. Bringi, V. N. *et al.* Raindrop Size Distribution in Different Climatic Regimes from Disdrometer and Dual-Polarized Radar Analysis. *J. Atmos. Sci.* **60**, 354–365 (2003).
20. Dolan, B., Fuchs, B., Rutledge, S. A., Barnes, E. A. & Thompson, E. J. Primary Modes of Global Drop Size Distributions. *J. Atmos. Sci.* **75**, 1453–1476 (2018).
21. Wen, L., Chen, G., Wang, S., Nie, J. & Zhao, K. Observed universal continuum morphology of raindrops reveals a concise diagram of heavy precipitation microphysics. *Proc. Natl. Acad. Sci.* **123**, e2525260123 (2026).
22. Hamada, A., Takayabu, Y. N., Liu, C. & Zipser, E. J. Weak linkage between the heaviest rainfall and tallest storms. *Nat. Commun.* **6**, 6213 (2015).
23. Gingrey, A., Varble, A. & Zipser, E. Relationships between Extreme Rain Rates and Convective Intensities from the Perspectives of TRMM and WSR-88D Radars. *J. Appl. Meteorol. Climatol.* **57**, 1353–1369 (2018).
24. Wang, T. & Tang, G. Spatial variability and linkage between extreme convections and extreme precipitation revealed by 22-year space-borne precipitation radar data. *Geophys. Res. Lett.* **47**, e2020GL088437 (2020).
25. Xu, W., Chen, H., Wei, H., Luo, Y. & Zhao, T. Extreme Precipitation Produced by Relatively Weak Convective Systems in the Tropics and Subtropics. *Geophys. Res. Lett.* **49**, e2022GL098048 (2022).

787 26. Sohn, B. J., Ryu, G.-H., Song, H.-J. & Ou, M.-L. Characteristic Features of
788 Warm-Type Rain Producing Heavy Rainfall over the Korean Peninsula Inferred
789 from TRMM Measurements. *Mon. Weather Rev.* **141**, 3873–3888 (2013).
790 27. Song, H., Sohn, B., Hong, S. & Hashino, T. Idealized numerical experiments on
791 the microphysical evolution of warm-type heavy rainfall. *J. Geophys. Res. Atmos.*
792 **122**, 1685–1699 (2017).
793 28. Ryu, J., Song, H.-J., Sohn, B.-J. & Liu, C. Global Distribution of Three Types of
794 Drop Size Distribution Representing Heavy Rainfall From GPM/DPR
795 Measurements. *Geophys. Res. Lett.* (2020).
796 29. Yu, S., Luo, Y., Wu, C. & Xu, W. Subseasonal Variations of Convective and
797 Microphysical Characteristics of Extreme Precipitation Over the Pearl River Delta
798 at Monsoon Coast. *J. Geophys. Res. Atmos.* **128**, e2022JD037804 (2023).
799 30. Chen, G. *et al.* Contrasts in Microphysical Characteristics Between the Meiyu
800 Warm-Sector and Frontal Heavy Rainfall in East China. *Geophys. Res. Lett.* **52**,
801 e2025GL117902 (2025).
802 31. Wang, Z. *et al.* Contrasting microphysics and environmental drivers of weak and
803 intense convection-induced extreme precipitation: Insights from GPM DPR
804 observations. *Atmos. Res.* **334**, 108762 (2026).
805 32. Ryu, J., Lee, J. & You, Y. A distinct type of heavy rainfall with large raindrops
806 over extratropical regions revealed by 10 years of GPM spaceborne radar
807 measurements. *Int. J. Appl. Earth Obs. Geoinformation* **144**, 104879 (2025).
808 33. Awaka, J. *et al.* Rain Type Classification Algorithm Module for GPM Dual-
809 Frequency Precipitation Radar. *J. Atmos. Ocean. Technol.* **33**, 1887–1898 (2016).
810 34. Awaka, J. *et al.* Development of Precipitation Type Classification Algorithms for
811 a Full Scan Mode of GPM Dual-frequency Precipitation Radar. *J. Meteorol. Soc.*

812 *Jpn. Ser II* **99**, 1253–1270 (2021).
813 35. Arulraj, M., Petkovic, V., Ferraro, R. R. & Meng, H. Precipitation Vertical
814 Structure Characterization: A Feature-Based Approach. *J. Hydrometeorol.* **24**,
815 2281–2297 (2023).
816 36. Zhang, Y., Zhang, X. & Ni, X. Microphysical properties of various precipitation
817 systems worldwide classified via objective methods based on dual-frequency
818 precipitation radar observations. *Atmos. Chem. Phys.* **26**, 4727–4747 (2026).
819 37. Schumacher, C. & Houze, R. A. The TRMM Precipitation Radar's View of
820 Shallow, Isolated Rain. *J. Appl. Meteorol.* **42**, 1519–1524 (2003).
821 38. Johnson, R. H., Rickenbach, T. M., Rutledge, S. A., Ciesielski, P. E. & Schubert,
822 W. H. Trimodal Characteristics of Tropical Convection. *J. Clim.* **12**, 2397–2418
823 (1999).
824 39. Oertel, A., Boettcher, M., Joos, H., Sprenger, M. & Wernli, H. Potential vorticity
825 structure of embedded convection in a warm conveyor belt and its relevance for
826 large-scale dynamics. https://doi.org/10.3929/ETHZ-B-000456726 (2020)
827 doi:10.3929/ETHZ-B-000456726.
828 40. Houze, R. A., Rasmussen, K. L., Zuluaga, M. D. & Brodzik, S. R. The variable
829 nature of convection in the tropics and subtropics: a legacy of 16 years of the
830 tropical rainfall measuring mission satellite. *Rev. Geophys.* **53**, 994–1021 (2015).
831 41. Raut, B. A. *et al.* Microphysical Origin of Raindrop Size Distributions During the
832 Indian Monsoon. *Geophys. Res. Lett.* **48**, e2021GL093581 (2021).
833 42. Tan, J., Jakob, C., Rossow, W. B. & Tselioudis, G. Increases in tropical rainfall
834 driven by changes in frequency of organized deep convection. *Nature* **519**, 451–
835 454 (2015).
836 43. Liu, C. & Zipser, E. J. Why does radar reflectivity tend to increase downward

toward the ocean surface, but decrease downward toward the land surface? *J. Geophys. Res. Atmos.* **118**, 135–148 (2013).

44. Luo, Y., Zheng, H., Zhang, Y. & Xie, Y. From Macro to Micro: How Large-Scale Atmospheric Circulation Governs Raindrop Size Variations during Warm-Rain Processes? *J. Clim.* **39**, 5885–5902 (2026).
45. Peng, H. *et al.* Aerosol-driven precipitation modification: spatiotemporal heterogeneity in precipitation microphysics and vertical profiles over China's megacity clusters. *Atmos. Chem. Phys.* **26**, 3299–3319 (2026).
46. Wang, Z. *et al.* Quantitative Evaluation and Error Analysis of GPM DPR Rainfall Products (Version 07) Over Yangtze-and-Huai River Basin in China. *Earth Space Sci.* **13**, e2025EA004778 (2026).
47. Hou, A. Y. *et al.* The Global Precipitation Measurement Mission. *Bull. Am. Meteorol. Soc.* **95**, 701–722 (2014).
48. Iguchi, T. *et al.* GPM/DPR Level-2 Algorithm Theoretical Basis Document. (2021).
49. Seto, S. *et al.* The Precipitation Rate Retrieval Algorithms for the GPM Dual-frequency Precipitation Radar. *J. Meteorol. Soc. Jpn. Ser II* **99**, 205–237 (2021).
50. Houze, R. A., Wilton, D. C. & Smull, B. F. Monsoon convection in the Himalayan region as seen by the TRMM Precipitation Radar. *Q. J. R. Meteorol. Soc.* **133**, 1389–1411 (2007).
51. Sohn, B.-J. *et al.* Revealing the hidden link of the Walker circulation on heavy rainfall patterns in the Eastern Pacific. *Commun. Earth Environ.* **6**, 144 (2025).
52. Nesbitt, S. W., Zipser, E. J. & Cecil, D. J. A Census of Precipitation Features in the Tropics Using TRMM: Radar, Ice Scattering, and Lightning Observations. *J. Clim.* **13**, 4087–4106 (2000).

53. Liu, C., Zipser, E. J., Cecil, D. J., Nesbitt, S. W. & Sherwood, S. A Cloud and Precipitation Feature Database from Nine Years of TRMM Observations. *J. Appl. Meteorol. Climatol.* **47**, 2712–2728 (2008).
54. Angulo-Umana, P., Kim, D., Blossey, P. N. & Kim, D. Morphological Characteristics of Tropical Precipitation Features Producing Extreme Instantaneous Rain Rates in GPM and DYAMOND Global Storm Resolving Models. *Geophys. Res. Lett.* **53**, e2026GL124306 (2026).
55. Zipser, E. J., Cecil, D. J., Liu, C., Nesbitt, S. W. & Yorty, D. P. Where are the most intense thunderstorms on Earth? *Bull. Am. Meteorol. Soc.* **87**, 1057–1072 (2006).
56. Nesbitt, S. W., Cifelli, R. & Rutledge, S. A. Storm Morphology and Rainfall Characteristics of TRMM Precipitation Features. *Mon. Weather Rev.* **134**, 2702–2721 (2006).
57. Hamada, A., Murayama, Y. & Takayabu, Y. N. Regional characteristics of extreme rainfall extracted from TRMM PR measurements. *J. Clim.* **27**, 8151–8169 (2014).
58. Higgins, I. *et al.* B-VAE: Learning basic visual concepts with a constrained variational framework. in (Toulon, France, 2017).
59. Bishop, C. M. *Pattern Recognition and Machine Learning*.
60. Akaike, H. A new look at the statistical model identification. *IEEE Trans. Autom. Control* **19**, 716–723 (1974).
61. Schwarz, G. Estimating the Dimension of a Model. *Ann. Stat.* **6**, (1978).
62. Arthur, D. & Vassilvitskii, S. K-means++: The advantages of careful seeding. in vols 07-09-January-2007 1027–1035 (New Orleans, LA, United states, 2007).
63. Tan, J., Oreopoulos, L., Jakob, C. & Jin, D. Evaluating rainfall errors in global

887 climate models through cloud regimes. *Clim. Dyn.* **50**, 3301–3314 (2018).
888 64. Zelinka, M. D., Tan, I., Oreopoulos, L. & Tselioudis, G. Detailing cloud property
889 feedbacks with a regime-based decomposition. *Clim. Dyn.* **60**, 2983–3003 (2023).
890 65. Kitagawa, E. M. Components of a Difference Between Two Rates*. *J. Am. Stat.*
891 *Assoc.* **50**, 1168–1194 (1955).
892

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- npjCASSI.docx