

Geophysical Research Letters®

RESEARCH LETTER

10.1029/2025GL116048

Key Points:

- ECMWF Re-Analysis V5 (ERA5) environmental variables exhibit robust and distinct coevolution patterns with IMERG errors throughout precipitation events
- The normalized temporal gradients of environmental variables exhibit a clear probabilistic inference capability for event stages
- Adding environmental variables and their temporal gradients to a machine-learning-based IMERG bias-correction yields clear improvements

Supporting Information:

Supporting Information may be found in the online version of this article.

Correspondence to:

R. Li,
runzli1@uci.edu

Citation:

Li, R., Guilloteau, C., & Foufoula-Georgiou, E. (2025). Added value of environmental variables for satellite precipitation retrieval: A temporal coevolution perspective and a machine learning integration assessment. *Geophysical Research Letters*, 52, e2025GL116048. <https://doi.org/10.1029/2025GL116048>

Received 18 MAR 2025

Accepted 20 MAY 2025

Author Contributions:

Conceptualization: Runze Li, Efi Foufoula-Georgiou

Formal analysis: Runze Li

Funding acquisition: Efi Foufoula-Georgiou

Investigation: Runze Li

Methodology: Runze Li

Project administration: Efi Foufoula-Georgiou

Supervision: Efi Foufoula-Georgiou

Visualization: Runze Li

Writing – original draft: Runze Li

© 2025. The Author(s).

This is an open access article under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

Added Value of Environmental Variables for Satellite Precipitation Retrieval: A Temporal Coevolution Perspective and a Machine Learning Integration Assessment

Runze Li¹ , Clement Guilloteau¹, and Efi Foufoula-Georgiou^{1,2} 

¹Department of Civil and Environmental Engineering, University of California, Irvine, CA, USA, ²Department of Earth System Science, University of California, Irvine, CA, USA

Abstract Satellite precipitation retrieval is inherently an underdetermined inverse problem where additional physical constraints could substantially enhance accuracy. While previous studies have explored static (pixel-based/spatial-context-based) environmental variables at discrete satellite observation times, their temporal dynamic information remains underutilized. Building on our earlier finding that retrieval errors depend on storm progression (event stage), we propose a new, physically interpretable mechanism for improving retrievals, namely, leveraging environmental variables' temporal dynamics as proxies for event stages. Using IMERG satellite product and GV-MRMS as ground-truth over CONUS (2018–2020), we first demonstrate robust coevolution patterns of environmental variables and satellite errors throughout events, and show that these variables' temporal gradients reliably infer event stages. We then demonstrate that incorporating these variables and their gradients into a machine-learning post-processing framework improves retrieval accuracy. This work inspires and guides more thorough utilization of spatiotemporal atmospheric fields encoding rich physical information within advanced machine-learning frameworks for further algorithm improvement.

Plain Language Summary Satellite-based precipitation estimates are crucial for weather forecasting, hydrological monitoring, and climate research. However, as satellites estimate surface rainfall indirectly from cloud radiometric signals, the estimation inevitably carries significant uncertainty, underscoring the need for additional information sources. As precipitation initiates, develops, and dissipates within the context of cloud microphysical and large-scale environment, such information may provide valuable constraints for retrievals. Yet, most retrieval algorithms use only “snapshots” of environmental variables at isolated observation times, overlooking the rich information embedded in their temporal dynamics. Building on our earlier findings that satellite errors systematically vary with storm progression, we propose a novel, physically interpretable mechanism—harnessing environmental variables' temporal dynamics as proxies for event stages to improve precipitation retrievals. Statistical analyses confirm that environmental variables and satellite errors coevolve robustly and that these variables' temporal gradients can serve as reliable indicators of event stages. To empirically test the accuracy gains in retrievals, we conduct a machine-learning-based bias-correction of IMERG incorporating these variables and their gradients, and demonstrate pronounced improvements in both detection skill and intensity accuracy. This work is expected to guide future efforts to integrate spatiotemporal evolution patterns of environmental variables within advanced deep learning frameworks for further satellite retrieval improvement.

1. Introduction

Satellite-based precipitation products provide high-resolution, continuous observations crucial for hydrologic modeling, water resources applications, and climate studies (Fassoni-Andrade et al., 2021; Good et al., 2021; Nearing et al., 2024). Over recent decades, advances in sensor technology and retrieval algorithms have significantly improved product quality (Hou et al., 2014; Huffman et al., 2023), yet notable discrepancies persist when benchmarked against ground observations (Guilloteau et al., 2021; Li et al., 2018, 2022). A principal reason lies in the indirect estimation of surface precipitation from the radiometric signatures of hydrometeors, leading to an inherently ill-posed underconstrained inversion problem with significant uncertainty in the resulting estimates (Guilloteau et al., 2021, 2023).

A salient manifestation of this ill-posedness, as demonstrated in our previous work, is the varying performance of satellite products throughout precipitation lifecycles (Li et al., 2021, 2025). As cloud microphysics evolve with

Writing – review & editing:
Clement Guilloteau, Efi Foufoula-Georgiou

storm development, both the radiative response to total hydrometeor content aloft and the efficiency with which these hydrometeors convert into surface precipitation (i.e., precipitation efficiency) alter significantly (Bouniol et al., 2016; Gupta et al., 2023). These changes collectively result in marked variations in the radiometric-signal-to-rain-rate relationship during precipitation lifecycles (Guilloteau & Foufoula-Georgiou, 2024), while the current input parameters/algorithms do not account for such variations, lacking “stage-awareness” as a necessary constraint in the solution space, thereby substantially contributing to the retrieval uncertainty.

The fact that the “real” event stage cannot be accessible from satellite observations, drives us to seek alternative sources of information that encapsulate or infer stages. A physically intuitive approach leverages the fact that precipitation emerges from the synergy between cloud microphysics and the broader atmospheric environment like moisture flux, atmospheric instability, and wind fields—a feedback loop in which the evolving environment drives precipitation development while simultaneously receiving feedback from it (Gupta et al., 2023; Lee et al., 2016). This ongoing interplay, or “coevolution”, implies that environmental variables may inherently encode physical “signatures” of storm stages. For example, high upward vertical velocity, indicative of vigorous convective updrafts signal storm initiation, while its subsequent decline/reversal marks maturation and dissipation. By integrating these variables into satellite retrieval frameworks, we can theoretically bolster event-stage sensitivity, impose physically grounded constraints on the inversion process, and ultimately improve retrieval accuracy.

While both the operational Global Precipitation Measurement (GPM) Passive Microwave (PMW) retrieval algorithm, that is, the Goddard Profiling Algorithm (GPROF) (Passive Microwave Algorithm Team Facility, 2022), and other algorithmic research have incorporated a few environmental variables in their retrieval frameworks (Pfreundschuh et al., 2022; Upadhyaya et al., 2022), they predominantly utilize instantaneous pixel-based/spatial-context-based information at discrete satellite observation times, without exploring the role of their temporal dynamics in a physically interpretable context. Motivated by the theoretical coevolution of environmental variables and satellite performance during precipitation events, this study investigates whether the temporal context in these variables can indeed function as a proxy for event stages and thereby improve retrieval accuracy. Through statistical analysis and a Machine Learning (ML)-based postprocessing experiment, we address three key questions: (a) Do environmental variables from global reanalysis exhibit robust coevolution patterns with satellite retrieval errors over an event's lifecycle? (b) Can these variables serve as robust probabilistic indicators of event stages? (c) Does integrating these stage-informative signals via a ML bias-correction framework enhance satellite precipitation estimates? To this end, the Integrated Multi-satellite Retrievals for GPM (IMERG) and atmospheric parameters from European Centre for Medium-Range Weather Forecasts (ECMWF) Re-Analysis V5 (ERA5) are analyzed against the Ground Validation-Multi-Radar/Multi-Sensor (GV-MRMS) product over CONUS from 2018 to 2020.

2. Data and Methods

2.1. Data

IMERG V07B Final Run is GPM's level 3 multi-satellite merged global precipitation product (Huffman et al., 2023), which delivers 0.5-hourly, $0.1^\circ \times 0.1^\circ$ precipitation estimates by merging multiple PMW and Infrared (IR) observations, calibrated with measurements from GPM Core Observatory and adjusted through monthly gauge analyses. ECMWF Re-Analysis V5 offers a comprehensive suite of atmospheric/land/oceanic variables at 1 hr, $0.25^\circ \times 0.25^\circ$ resolution (Hersbach et al., 2023a, 2023b), which are linearly interpolated to 0.5 hr, $0.1^\circ \times 0.1^\circ$ to match IMERG. GV-MRMS is a radar gauge blended Quantitative Precipitation Estimation (QPE) based on National Oceanic Atmospheric Administration MRMS, post-processed specifically for satellite evaluation (Kirstetter et al., 2018). An accompanying Radar Quality Index (RQI; 0–100) is employed to retain relatively high-quality regions ($RQI > 60$) and is applied across all data types (Petersen et al., 2020). GV-MRMS is spatially aggregated from its native $0.01^\circ \times 0.01^\circ$, 0.5 hr resolution to IMERG's $0.1^\circ \times 0.1^\circ$ gridbox.

2.2. Methods

2.2.1. Statistical Analysis

We first define a precipitation event as an uninterrupted sequence of half-hourly timesteps with intensity ≥ 0.1 mm/hr at each $0.1^\circ \times 0.1^\circ$ pixel in GV-MRMS within an Eulerian framework and extract all events over

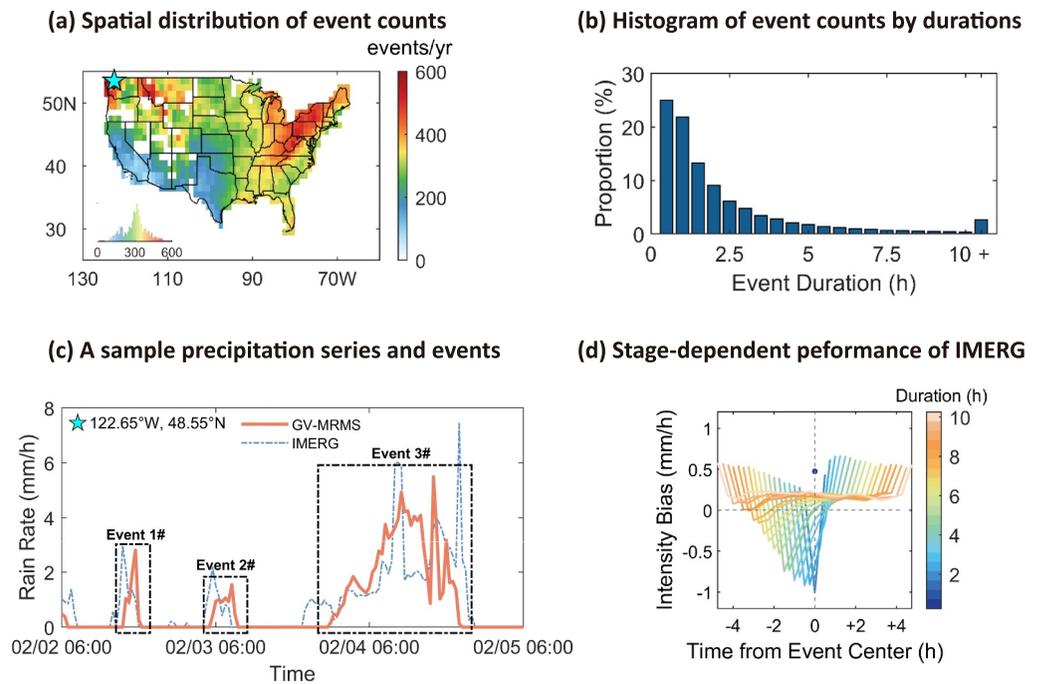


Figure 1. Overview of precipitation event characteristics. (a) Spatial distribution of annual GV-MRMS event counts, which has been aggregated from raw $0.1^\circ \times 0.1^\circ - 1^\circ \times 1^\circ$ for enhanced visualization, with a corresponding histogram in the panel's bottom-left corner, (b) histogram of GV-MRMS event counts by duration, with durations >10 hr grouped and labeled as “+”, (c) sample precipitation time series from GV-MRMS (red) and IMERG (blue) at 122.65°W , 48.55°N (asterisk in (a)) during February 2–5, 2018. Three major GV-MRMS events are outlined, and (d) temporal evolution patterns of mean Intensity Bias (IMERG—GV-MRMS) for rainy hours (when both detect precipitation, i.e., “hits”) at half-hour intervals during GV-MRMS events of varying durations (color-coded for durations up to 10 hr), aligned to event centers. Panels (a), (b), and (d) are calculated using all CONUS data (2018–2020).

CONUS during the study period of 2018–2020 (Li et al., 2023a). The basic information about events, that is, the annual event count map and the distribution of event durations across CONUS, is shown in Figures 1a and 1b, respectively. As the foundation and motivation for this study, Figure 1c presents a representative precipitation time series, highlighting the divergent evolution patterns between GV-MRMS and IMERG events, while Figure 1d statistically demonstrates IMERG's varying systematic intensity bias during events across various durations.

To analyze stage-based performance across events of varying durations (≥ 1 hr), we segment each event into five equal-duration intervals—defined as “event stages” (S)—and linearly interpolate the associated variables at the endpoints of these intervals. To place events in their environmental context, we extend the analysis time window associated to each event forward and backward by the duration of the event, normalizing these windows similarly into five stages each, forming a unified 15- S framework: 5 Before-event (B1–B5), 5 During-event (D1–D5), and 5 After-event (A1–A5) stages.

The statistical analysis of how environmental variables (V), mediated by event stage, are indicative of satellite errors (E) is framed in a forward (diagnostic) and inverse (inferential) perspective. In the forward analysis, S is assumed known from GV-MRMS and conditional expectations $E(V|S)$ and $E(E|S)$ are computed to reveal if there are any systematic coevolution patterns of environmental variables and satellite errors with event stage's progression. Seven environmental variables alongside two event-based satellite performance metrics are used here, as detailed in Figure 2. Specifically, the *Detection Proportion* quantifies rain/no-rain classification accuracy, representing false positive proportion outside events and true positive proportion inside events, respectively:

$$\text{Detection Proportion} (S) = \frac{N_{\text{IMERG}} (S)}{N_{\text{Event}}} \times 100\%, \quad (1)$$

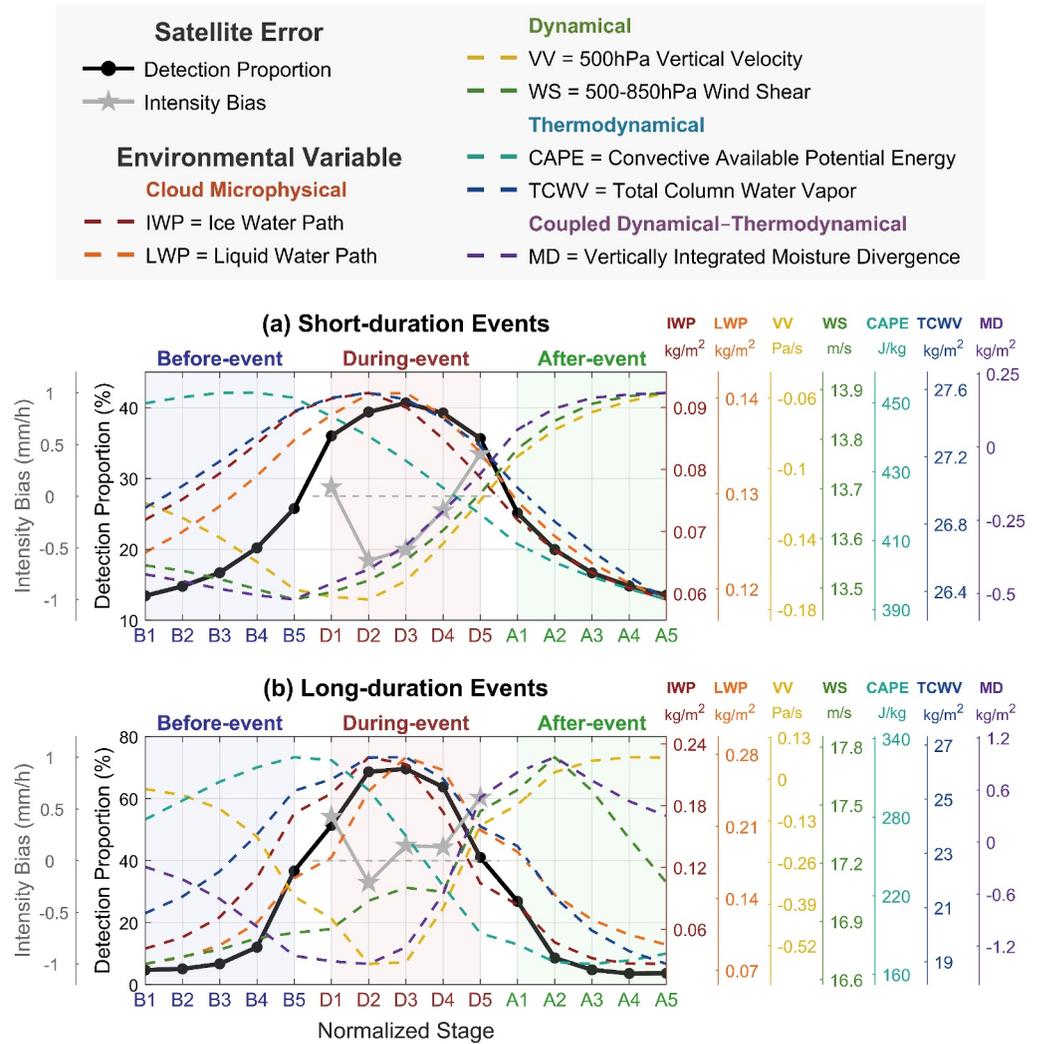


Figure 2. Evolution of the mean values of environmental variables (from ECMWF Re-Analysis V5) with event stage (extracted from MRMS), plotted together with the evolution of IMERG performance (detection proportion and intensity bias) to highlight their coevolution. A total of 15 stages, as defined in Methods, are presented: Before-event (B1–B5, blue shading), during-event (D1–D5, red shading), and after-event (A1–A5, green shading). Two satellite error metrics and seven representative precipitation-related variables are presented, as listed in the legend above. Color-coded axes denote corresponding variable/metric-specific units and ranges. Only events without overlapping events in their before-event/after-event periods are selected for presentation to avoid potential confounding effects. Results are separately calculated for (a) short-duration (≤ 5 hr), and (b) long-duration (> 5 hr) events, using all CONUS data (2018–2020), comprising a total of 10,321,875 and 433,242 events, respectively.

where $N_{IMERG}(S)$ is the number of IMERG rainy detections at GV-MRMS stage S and N_{Event} is the total GV-MRMS event count. The *Intensity Bias* represents intensity quantification accuracy during concurrent detections (i.e., “hit”):

$$Intensity\ Bias(S) = I_{IMERG}(S) - I_{GV-MRMS}(S), \quad (2)$$

where $I_{IMERG}(S)$ and $I_{MRMS}(S)$ are the mean intensities from IMERG and GV-MRMS, respectively, at stage S .

In the inverse analysis, we assess the probabilistic discriminatory skill of environmental variables on the stage, without reference to any satellite product. Drawing on the robust, quadratic-like patterns observed in the diagnostic step, we hypothesize that the temporal gradients of these variables may better capture stages than absolute

values, with these gradients further normalized by the variables' magnitudes to mitigate regional/seasonal background environment differences. We compute these gradients via central differences:

$$\Delta V_{n,t} = \frac{V_{t+1} - V_{t-1}}{2V_t}, \quad (3)$$

and quantify ΔV_n 's probabilistic inference power for each stage S . Specifically, all ΔV_n values over the study period are first stratified into 10-percentile bins. Within each ΔV_n bin, we compute the conditional probability $P(S|\Delta V_n)$ for all 15 defined stages, and identify and display the three stages with the highest probabilities. This ranking reveals whether different ΔV_n magnitudes exhibit distinct stage preferences (higher probability of co-occurrence), thereby reflecting their stage-discriminative skill.

2.2.2. Machine Learning Postprocessing

Based on statistical insights, we apply a ML-based postprocessing bias-correction test to IMERG to assess whether incorporating environmental variables and their temporal gradients practically improves IMERG's accuracy. To explicitly evaluate the added value of temporal gradients, we employ a tabular-data-based model that allows manual assessment of time-related features—offering greater transparency and interpretability than deep learning models, which typically require architectural modifications to isolate these effects. Specifically, we adopt CatBoost—a Gradient Boosting Decision Tree (GBDT) model that excels in modeling complex nonlinear relationships and feature interactions (Prokhorenkova et al., 2018). Its innovative mechanisms such as symmetric trees and ordered target encoding, further reduce prediction bias and overfitting, making it a leading tree-based model.

Given the prevalence of zero precipitation values and the highly skewed precipitation distribution of non-zero values, a two-step regression is employed: a classifier for rain/non-rain identification; then a regressor predicting intensity only for rainy cases, which has been demonstrated to outperform one-step regression in both prior studies and our experiments (Kossieris et al., 2024; Lei et al., 2022). Within this framework, our comparative tests indicate that CatBoost outperforms other tree-based models (e.g., Random Forest, XGBoost, LightGBM), particularly in the classification stage, which in turn improves the regression and overall accuracy, and is therefore selected.

To assess the added value of environmental variables (V) and their gradients (ΔV) in improving accuracy, we design three postprocessing tests with progressively expanded input configurations: Test 1: IMERG only, Test 2: IMERG + V , and Test 3: IMERG + V + ΔV , with GV-MRMS as the target. Here we incorporate raw ΔV rather than the normalized ΔV_n , as the variables' magnitude V is already explicitly added. Given the vast data volume and diverse environmental/precipitation conditions year-round across CONUS, which pose challenges for tree-based models lacking explicit spatiotemporal representations, we focus on the Northeastern US during winter—a representative region where IMERG performance is notably degraded by snow/ice and complex terrain (Li et al., 2023a). To further mitigate computational constraints and potential overfitting from dense spatiotemporal sampling, one-third of the samples are randomly selected and partitioned into 70%/30% training-test sets. The training set is further subdivided into a 56%/14% (4:1 ratio) for training/validation in hyperparameter tuning. The final sample sizes for the training, validation, and test datasets are 4,842,467, 1,210,617, and 2,594,179, respectively.

We assess the overall improvement by the Heidke Skill Score (HSS) for classification (precipitation detection) and the distribution of Absolute Error (AE) for regression (precipitation magnitude). Heidke Skill Score measures detection skill relative to a random prediction, weighting misses/false alarms equally (Wang et al., 2025). Here we extend the original HSS to various intensity thresholds (T), defining a precipitation occurrence in labels/predictions only when its intensity exceeds T :

$$HSS = \frac{2(N_h(T) * N_{cn}(T) - N_f(T) * N_m(T))}{(N_h(T) + N_m(T)) * (N_m(T) + N_{cn}(T)) + (N_f(T) + N_{cn}(T)) * (N_h(T) + N_f(T))}, \quad (4)$$

where $N_h(T)$, $N_f(T)$, $N_m(T)$, and $N_{cn}(T)$ denote the numbers of hits, false alarms, misses, and correct negatives, respectively, at a given T . Heidke Skill Score ranges from $-\infty$ to 1, where one signifies a perfect forecast, zero implies no skill beyond random guessing, and negative values indicate worse-than-random performance.

Additionally, we leverage a widely used Explainable Artificial Intelligence (XAI) technique, SHapley Additive exPlanations (SHAP), to reveal each input's relative contribution (Lundberg & Lee, 2017). This game-theoretic framework computes a feature's marginal contribution for each instance by comparing the model's prediction against a baseline expectation across all feature coalitions. Here we aggregate these instance-level contributions into global SHAP values to elucidate overall feature importance, computing them separately for classification and regression in Test 3# that incorporates all variables (see discussion in the next section).

3. Results

Figure 2 illustrates the coevolution of IMERG satellite errors—detection fraction and mean intensity bias—and environmental variables across the 15 normalized event stages. Both short- and long-duration precipitation events exhibit quasi-parabolic life-cycle progression patterns of satellite errors and environmental variables, though with disparities in phase and magnitude. For short-duration events (≤ 5 hr, Figure 2a), which constitute the majority ($\sim 90\%$, Figure 1b), the detection fraction is relatively symmetric, rising steadily with proximity to the event center. This smooth transition of IMERG detection proportion at events' edges (start/end times) actually reflects a high number of false alarms immediately before/after events (B5/A1) and a high number of misses at their onsets/cessations (D1/D5). This might be attributed to IMERG's timing/drizzle/interpolation issues (Li et al., 2023b). The intensity bias exhibits a second-order-like variation as well, with positive biases at the beginning/end (D1/D5) and negative biases in the intermediate stages (D2/D3/D4), and an asymmetric peak in the early-to-mid stage (D2), which has been previously attributed to the combined effects of intensity and cloud microphysical changes (Li et al., 2025), highlighting again the clear stage-dependent satellite performance.

The environmental variables' behaviors are also physically explainable for short-duration events (Figure 2a). For example, both IWP and LWP follow a lifecycle pattern—gradually accumulating pre-event, peaking during precipitation, then dissipating—with IWP peaking earlier (D2) due to rapid updraft-driven ice nucleation, and LWP peaking slightly later (D3) owing to slower droplet coalescence (Braga & Vila, 2014). VV, closely linked to convective strength, exhibits a marked parabolic pattern as well, remaining negative (upward) and peaking in the early-to-mid stage (D2), aligning with the approximate intensity peak time (Figure S1a in Supporting Information S1). WS exhibits relatively modest variation but still follows a “decrease-then-increase” pattern, likely driven by initial momentum exchange from convective mixing, followed by precipitation-induced evaporative cooling that triggers downbursts and outflow boundaries at low levels, culminating in a WS peak at event onset (D1) (Romanic et al., 2022).

CAPE accumulates well before precipitation onset (B1/B2), peaks just before precipitation (B3), then gradually decreases as precipitation progresses due to unstable energy release. Driven by evapotranspiration and horizontal moisture transport, TCWV increases toward the precipitation peak (D2), then decreases as condensation/deposition remove water vapor. MD mirrors the evolution of WS, shifting from convergence to divergence after precipitation onset (D1) due to a similar mechanism: rainfall-induced outflow boundaries generating localized divergence. On larger scales, MD also inversely acts as a driver of precipitation cessation by reducing moisture supply (Tsuji et al., 2021). Conceptualizing MD partially as the time derivative of TCWV also helps explain why its peak (B5) precedes that of TCWV (D2).

Long-duration events (> 5 hr) exhibit similar but less smooth patterns in both satellite errors and environmental variables (Figure 2b). A more asymmetric detection profile and a less regular intensity-bias curve are observed, compared to short-duration events. Quantitatively, the detection proportion is generally higher within/around long-duration events, with one important reason being their larger extent both in space and time, making them more readily detectable by satellites and less vulnerable to spatiotemporal mismatches. Regarding intensity bias, while for short-duration events the bias is mostly negative (D1–D3, Figure 2a), for long-duration events it is mostly positive across the event duration (D1, D3–D5, Figure 2b), likely due to satellites' tendency to underestimate intense, short-lived convective precipitation and to slightly overestimate weaker, longer-lasting stratiform precipitation (Li et al., 2025). Overall, despite differences in magnitude, environmental variables evolve similarly to those in Figure 2a, with some distinctions attributable to extended timescales. For instance, well

before event onsets (B1/B2), VV undergoes a more pronounced transition from near-zero to negative, marking the initiation of upward motion, while CAPE exhibits a clearer energy buildup phase. Conversely, well after an event cessation (A3–A5), WS and MD exhibit secondary reductions once initial outflow boundary effects subside after precipitation. Especially, wind shear exhibits greater variability over longer timescales than other variables, likely due to its sensitivity to complex multi-scale influences (Yang et al., 2023). Figure S1 in Supporting Information S1 additionally shows the evolution of GV-MRMS event intensity and environmental variables at their original 30 min intervals for events of varying durations, revealing covariation patterns consistent with Figure 1.

Overall, across both short- and long-duration GV-MRMS events, environmental variables exhibit relatively stable, physically interpretable temporal evolution patterns, to some extent highlighting the reliability of ERA5 variables. However, their evolution does not closely mirror that of satellite errors, suggesting that rather than serving directly as error proxies, these environmental variables are better suited for identifying event stages, which can then act as intermediaries for indicating satellite errors. Indeed, the roughly quadratic evolution observed in most variables implies that their gradients may function as quasi-monotonic “stage” indicators. An additional test that classifies events based on their mean intensity on top of event duration reveals that the relative evolution patterns of environmental variables remain largely consistent regardless of intensity level (Figure S2 in Supporting Information S1), further supporting the potential of these variables to serve as indicators of event stages.

Building on this insight, Figure 3 evaluates environmental variables' predictive capacity for event stages by marking the top 3 highest conditional probabilities of stage $P(S|\Delta V_n)$ across ΔV_n quantiles. Notably, IWP, VV, and MD exhibit near-monotonic, stable ΔV_n -stage relationships (Figures 3a–3g). For example, IWP's high/middle/low quantiles—roughly corresponding to positive/near-zero/negative gradients—sequentially align with the pre-/during-/post-event stages (Figure 3a). This pattern is fully consistent with the rise-plateau-decline behavior of IWP in Figure 2, with similar trends observed for VV and MD (Figures 3c and 3g). Although WS and TCWV lack strict monotonicity, their distinct quantiles still show clear preferences for particular event stages (Figures 3d and 3f). LWP and CAPE appear much noisier but likewise follow discernible patterns, with lower/middle/higher quantiles generally mapping to before-/mid-/after-event periods, respectively (Figures 3b and 3e). A similar trend persists when isolating short-duration events, as they comprise the majority (Figure S3 in Supporting Information S1). However, with the extended timespan enhancing inter-stage differentiation for long-duration events, all variables except WS exhibit even more monotonic ΔV_n -stage relationships (Figure S4 in Supporting Information S1). Overall, these findings confirm that the normalized temporal gradients of environmental variables encode valuable “event stage” information for potential algorithmic improvements, though further empirical validation is warranted.

For this purpose, a ML-based IMERG postprocessing bias-correction experiment is conducted, with its workflow, validation results, and model interpretation presented in Figure 4. As detailed in Methods, three sets of input configurations were established by progressively incorporating environmental variables and their temporal gradients (Figure 4a-①), applied to the northeastern U.S (Figure 4a-②). The two-step CatBoost algorithm is employed (Figure 4a-③), using a random 56%/14%/30% split for training, validation, and testing, respectively (Figure 4a-④). For the validation using HSS and AE (Figures 4b and 4c), Test 1# (IMERG-only) lowers HSS relative to the raw IMERG (Figure 4b), indicating that complex nonlinear fitting alone cannot overcome IMERG's intrinsic information limitations. In contrast, adding environmental variables (Test 2#) markedly improves both HSS and AE, likely due to their role in indicating storm regime and partially storm stage. Further including variables' temporal gradients (Test 3#) yields even stronger gains, as it provides more insight into storm internal evolution. Quantitatively, for example, at the 0% threshold (≥ 0.1 mm/hr), HSS falls from 0.41 (raw IMERG) to 0.35 (−14%) using only IMERG (Test 1#), but climbs to 0.51 (+24%) with V added (Test 2#) and 0.60 (+46%) with ΔV further included (Test 3#) (Figure 4b). Meanwhile, the median AE drops from 0.66 (raw IMERG) to 0.60 (−9%, Test 1#), 0.40 (−39%, Test 2#), and 0.33 (−49%, Test 3#), respectively (Figure 4c), clearly illustrating the benefits of leveraging both environmental variables and their gradients. An additional correlation coefficient analysis in Figure S5 in Supporting Information S1 yields a similar conclusion. Additionally, we re-evaluate the detection rate and intensity bias of different post-processed results relative to raw IMERG (Figure S6 in Supporting Information S1). Overall, Test3# exhibits both reduced absolute errors and asymmetry across events, indicating the benefit of stage-awareness.

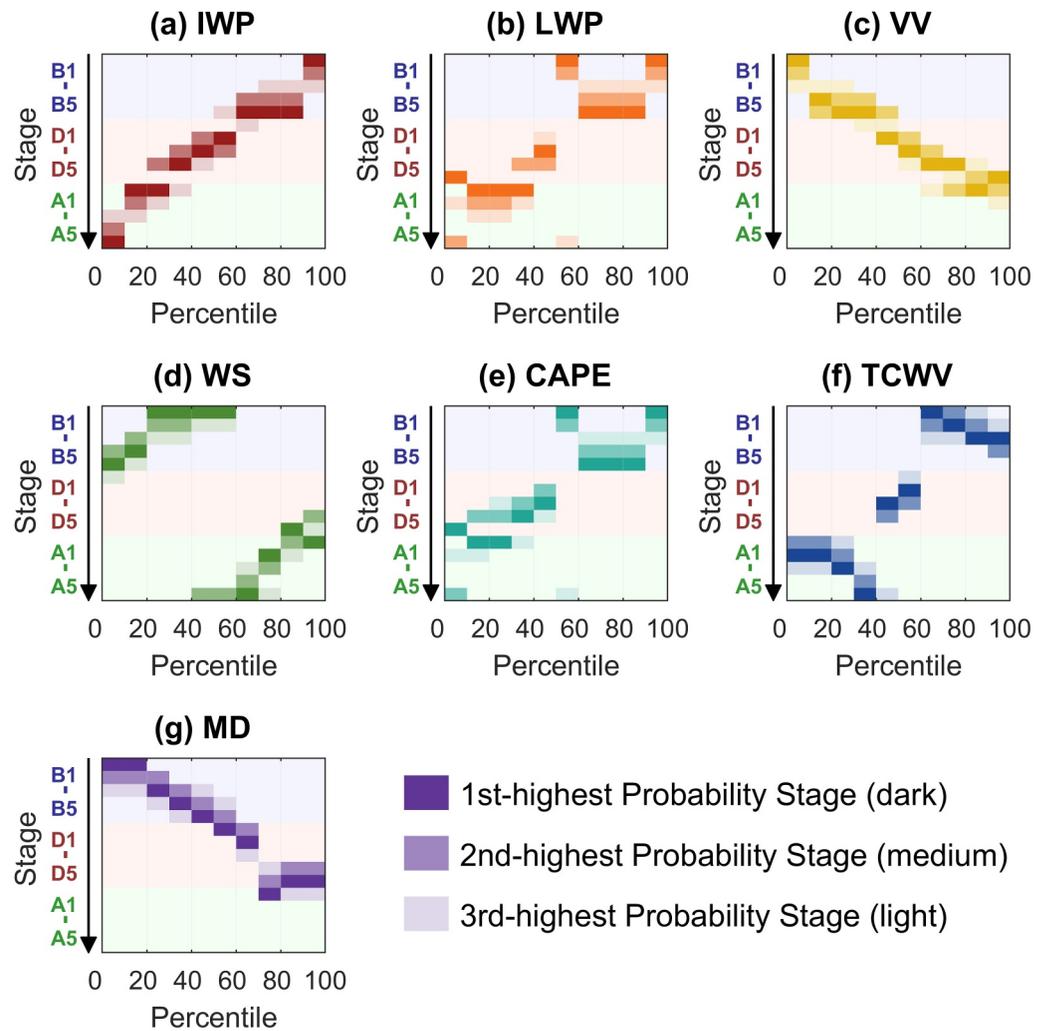


Figure 3. Probabilistic inference power of normalized temporal gradients of environmental variables (ΔV_n) for event stage (S). We only mark the stages with the three highest conditional probabilities, $P(S|\Delta V_n)$, independently computed for each percentile interval of the ΔV_n distribution from the entire study period (see an example legend for panel (g)). Arrows to the left of each panel indicate events' temporal progression through the 15 stages delineated in Figure 2. Results are calculated using all CONUS data (2018–2020).

Finally, Figure 4d illustrates the global SHAP feature importance for both the classification (top) and regression (bottom) in Test 3# that includes all variables. Notably, IMERG contributes minimally to classification, reflecting its large uncertainties over ice-/snow-covered surfaces (You et al., 2017). In contrast, IWP emerges as a dominant detection driver—even though it partially overlaps with IMERG's IR/PMW information and is affected by uncertainties in ERA5's cloud parameterizations (Hersbach et al., 2020)—underscoring the value of physics-based representation. As expected, CAPE is also a major factor for rain detection. For regression, while IMERG remains influential, LWP—which PMW struggles to detect—contributes significantly, underscoring its additional value. MD also proves influential for intensity estimation. Though already included in GPROF (Passive Microwave Algorithm Team Facility, 2022), TCWV remains contributory to both classification and regression, underscoring ML's deeper variable exploitation over traditional statistical models, and the role of environmental variables in the interpolated components of IMERG between the MW overpasses. The gradients generally contribute less, which is unsurprising since absolute values establish the baseline, with gradients enhancing finer details. Even though, ΔTCWV and ΔIWP still account for $\sim 10\%$ and $\sim 5\%$, respectively, for both classification and regression, while ΔCAPE contributes to another $\sim 5\%$ in regression, collectively, these gradients account for 26%/20% for classification/regression predictions, respectively, indicating a substantial overall contribution.

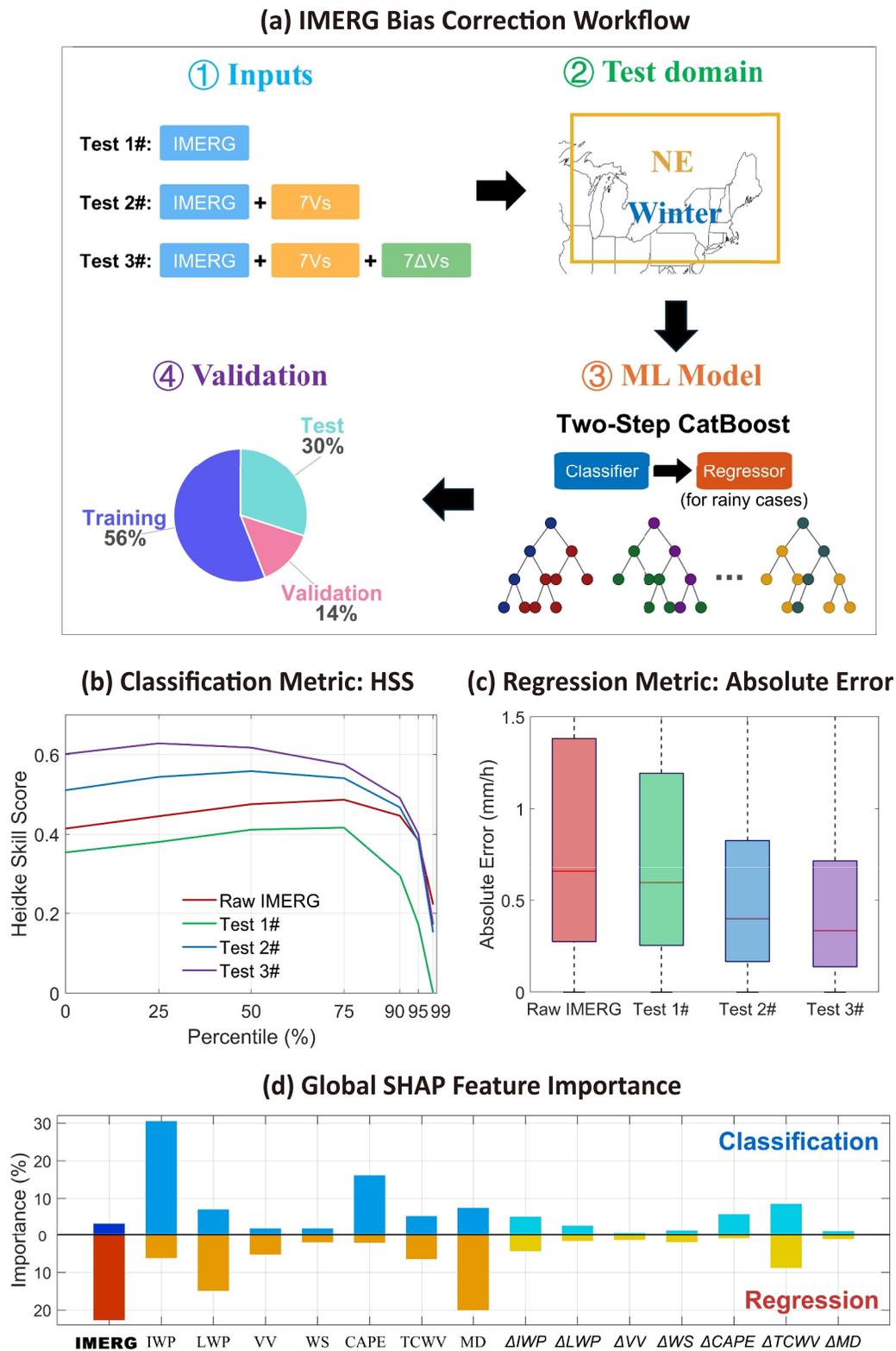


Figure 4.

4. Conclusions and Discussion

In this study, we explore the potential of leveraging the dynamic evolution of environmental variables to improve satellite precipitation retrieval accuracy in a physically interpretable manner. Building on evidence that satellite retrieval errors depend on precipitation-event stages and that environmental conditions evolve significantly throughout storm lifecycles, we hypothesize that the temporal dynamics of these variables could serve as proxies for storm progression, thereby indicating stage-dependent errors and boosting retrievals' overall accuracy. Our statistical analysis shows that ERA5 environmental variables and IMERG errors coevolve robustly during events, with the variables' temporal gradients reliably inferring event stages. A subsequent machine-learning post-processing test incorporating these variables and their gradients into IMERG effectively demonstrates accuracy gains.

While our statistical analysis and machine-learning experiment are based on IMERG, we note that the insights derived are not product-specific. This generality stems not only from the physical reality that environmental fields intrinsically coevolve with the precipitation lifecycle (Figure 2), but also from our analysis in Figure 3, which shows that the normalized temporal gradients of environmental variables exhibit robust inferential power for identifying the “real” event stages, independent of any specific retrieval. In other words, these variables could serve as effective proxies for event stages and potentially improve various satellite products' accuracy through this mechanism, provided their errors exhibit stage dependence. In fact, Li et al. (2025) have demonstrated that the stage-dependent error is intrinsic to the single-sensor microwave precipitation retrievals—the backbone of most global satellite precipitation estimates. Therefore, the insights presented here are transferable to improving satellite products beyond IMERG.

Beyond reducing errors arising from the unresolved evolution of retrieval relationships during storm progression, environmental variables' stage-inference capability is also promising to help mitigate spatiotemporal mismatch errors—particularly the temporal lag between cloud signatures and surface rainfall caused by hydrometeor fall time (Guiloteau et al., 2018; You et al., 2019). For example, satellite-derived event stage from temporally continuous satellite products can be inferred and aligned with the “true” stage proxy from environmental cues to estimate their phase offsets automatically. The phase-offset-specific correction magnitudes of satellite products can then be learned directly, without resorting to ad hoc time-shifting, which warrants future investigation. By contrast, physically based environmental variables are likely less effective in addressing another non-physical, time-dependent major error source in multi-satellite merged products, namely abrupt transitions among data sources of differing quality (e.g., PMW, IR, or interpolated data). Mitigating these artifacts will require alternative strategies, such as embedding sensor-specific characteristics within ML frameworks, which can be explored in future work.

Although this study implements the insights derived from our statistical analysis in a postprocessing framework for IMERG, the overarching insight—fully leveraging spatiotemporal atmospheric fields rich in physical information—applies to a range of satellite precipitation tasks, including instantaneous retrieval, observation-gap filling, bias-correction, downscaling, nowcasting, and multi-source data fusion (Dai & Ushijima-Mwesigwa, 2025; Guiloteau et al., 2025; Rahimi et al., 2024). On the other hand, however, to generalize our methods/insights to broader spatial domains, extensive testing across diverse regions remains essential, given the substantial variability in environmental conditions and satellite-derived errors. This could ultimately lead either to region-specific models trained within comparatively uniform environmental settings, or to a single, large-scale global model that adapts to local precipitation regimes through geolocation-/climate-aware embeddings combined with a pre-training/fine-tuning framework (e.g., Immorlano et al., 2025). Moreover, although the gradient-based contributions are moderate (~20%–25%; Figure 4d), this aligns with expectations given our relatively simple test model and limited temporal dynamics employed. Looking ahead, advanced deep-learning approaches,

Figure 4. ML-based bias-correction tests for IMERG incorporating environmental variables (V) and their temporal gradients (ΔV). (a) Workflow: ① three tests with progressively expanded input configurations ② over Northeastern U.S. (40°N–50°N, 90°W–65°W) in winter (December–January–February, DJF) using ③ the two-step Catboost algorithm: (i) A binary classifier for rainy/non-rainy discrimination; (ii) A regressor for identified rainy cases. ④ Validation follows a random 70%:30% training/test split, with the 70% training set further divided 2:8 (14% validation, 56% training) for hyperparameter tuning, (b) classification metric: Heidke Skill Score for raw IMERG and each test under varying detection thresholds, defined as percentiles of the GV-MRMS intensity distribution, (c) regression Metric: Boxplots of *Absolute Error* for raw IMERG and each test, only showing the 25th, 50th, and 75th percentiles due to the highly skewed distributions, and (d) global SHAP feature importances (%) of inputs for classification (top) and regression (bottom), separately in Test 3#, with SHAP values normalized for direct comparison between the two modules.

especially video-based architectures (e.g., Srivastava et al., 2024), are poised to deliver larger gains by automatically extracting more comprehensive spatiotemporal features, while our physically grounded analyses can, in turn, guide model design (e.g., incorporating lag-aware attention modules that capture delayed correlations among environmental variables and precipitation intensities (e.g., Yang et al., 2024)).

Data Availability Statement

GV-MRMS is available at NASA Global Hydrology Resource Center (GHRC) (Kirstetter et al., 2018); IMERG is available at NASA Goddard Earth Sciences (GES) Data and Information Services Center (DISC) (Huffman et al., 2023); ERA5 is available at ECMWF Copernicus Climate Change (C3S) Service Climate Data Store (CDS) (Hersbach et al., 2023a, 2023b).

Acknowledgments

This research was supported by the NASA Global Precipitation Measurement Mission (Grant 80NSSC22K0597), NASA Weather and Atmospheric Dynamics (Grant 80NSSC23K1304), NSF Division of Information and Intelligent Systems (Expand AI2ES, Grant IIS 2324008), and the Samuelli endowed chair. We thank Pierre-Emmanuel Kirstetter for his data support and Jackson Tan for his invaluable feedback on the manuscript. We also thank the three anonymous reviewers for their constructive comments which enhanced the clarity of our presentation.

References

- Bouniol, D., Roca, R., Fiolleau, T., & Poan, E. (2016). Macrophysical, microphysical, and radiative properties of tropical mesoscale convective systems over their life cycle. *Journal of Climate*, 29(9), 3353–3371. <https://doi.org/10.1175/JCLI-D-15-0551.1>
- Braga, R. C., & Vila, D. A. (2014). Investigating the ice water path in convective cloud life cycles to improve passive microwave rainfall retrievals. *Journal of Hydrometeorology*, 15(4), 1486–1497. <https://doi.org/10.1175/JHM-D-13-0206.1>
- Dai, T.-Y., & Ushijima-Mwesigwa, H. (2025). PrecipDiff: Leveraging image diffusion models to enhance satellite-based precipitation observations. *Proceedings of the AAAI Conference on Artificial Intelligence*, 39(27), 27932–27939. <https://doi.org/10.1609/aaai.v39i27.35010>
- Fassoni-Andrade, A. C., Fleischmann, A. S., Papa, F., Paiva, R. C. D. d., Wongchuig, S., Melack, J. M., et al. (2021). Amazon Hydrology from space: Scientific advances and future challenges. *Reviews of Geophysics*, 59(4). <https://doi.org/10.1029/2020RG000728>
- Good, P., Chadwick, R., Holloway, C. E., Kennedy, J., Lowe, J. A., Roehrig, R., & Rushley, S. S. (2021). High sensitivity of tropical precipitation to local sea surface temperature. *Nature*, 589(7842), 408–414. <https://doi.org/10.1038/s41586-020-2887-3>
- Guilloteau, C., & Foufoula-Georgiou, E. (2024). Life cycle of precipitating cloud systems from synergistic satellite observations: Evolution of macrophysical properties and precipitation statistics from geostationary cloud tracking and GPM active and passive microwave measurements. *Journal of Hydrometeorology*, 25(5), 789–805. <https://doi.org/10.1175/JHM-D-23-0185.1>
- Guilloteau, C., Foufoula-Georgiou, E., Kirstetter, P. E., Tan, J., & Huffman, G. J. (2021). How well do multi-satellite products capture the space-time dynamics of precipitation? Part I: Five products assessed via a wavenumber-frequency decomposition. *Journal of Hydrometeorology*, 22(11), 2805–2823. <https://doi.org/10.1175/jhm-d-21-0075.1>
- Guilloteau, C., Foufoula-Georgiou, E., Kummerow, C. D., & Petkovic, V. (2018). Resolving surface rain from GMI high-frequency channels: Limits imposed by the three-dimensional structure of precipitation. *Journal of Atmospheric and Oceanic Technology*, 35(9), 1835–1847. <https://doi.org/10.1175/JTECH-D-18-0011.1>
- Guilloteau, C., Kerrigan, G., Nelson, K., Migliorini, G., Smyth, P., Li, R. Z., & Foufoula-Georgiou, E. (2025). A generative diffusion model for probabilistic ensembles of precipitation maps conditioned on multisensor satellite observations. *IEEE Transactions on Geoscience and Remote Sensing*, 63, 1–15. <https://doi.org/10.1109/Tgrs.2025.3548518>
- Guilloteau, C., Le, P. V. V., & Foufoula-Georgiou, E. (2023). Constraining the multiscale structure of geophysical fields in machine learning: The case of precipitation. *IEEE Geoscience and Remote Sensing Letters*, 20, 1–5. <https://doi.org/10.1109/Lgrs.2023.3284278>
- Gupta, A. K., Deshmukh, A., Waman, D., Patade, S., Jadav, A., Phillips, V. T. J., et al. (2023). The microphysics of the warm-rain and ice crystal processes of precipitation in simulated continental convective storms. *Communications Earth and Environment*, 4(1), 226. <https://doi.org/10.1038/s43247-023-00884-5>
- Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., et al. (2020). The ERA5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society*, 146(730), 1999–2049. <https://doi.org/10.1002/qj.3803>
- Hersbach, H., Bell, B., Berrisford, P., Biavati, G., Horányi, A., Muñoz Sabater, J., et al. (2023a). ERA5 hourly data on single levels from 1940 to present [Dataset]. *Copernicus Climate Change Service (C3S) Climate Data Store (CDS)*. <https://doi.org/10.24381/cds.adbb2d47>
- Hersbach, H., Bell, B., Berrisford, P., Biavati, G., Horányi, A., Muñoz Sabater, J., et al. (2023b). ERA5 hourly data on pressure levels from 1940 to present [Dataset]. *Copernicus Climate Change Service (C3S) Climate Data Store (CDS)*. <https://doi.org/10.24381/cds.bd0915c6>
- Hou, A. Y., Kakar, R. K., Neeck, S., Azarbarzin, A. A., Kummerow, C. D., Kojima, M., et al. (2014). The global precipitation measurement mission. *Bulletin of the American Meteorological Society*, 95(5), 701–722. <https://doi.org/10.1175/Bams-D-13-00164.1>
- Huffman, G. J., Stocker, E. F., Bolvin, D. T., Nelkin, E. J., & Tan, J. (2023). GPM IMERG final precipitation L3 half hourly 0.1 degree x 0.1 degree V07 [Dataset]. *Goddard Earth Sciences Data and Information Services Center (GES DISC), Greenbelt, MD*. <https://doi.org/10.5067/GPM/IMERG/3B-HH/07>
- Immorlano, F., Eyring, V., le Monnier de Gouville, T., Accarino, G., Elia, D., Mandt, S., et al. (2025). Transferring climate change physical knowledge. *Proceedings of the National Academy of Sciences of the U S A*, 122(15), e2413503122. <https://doi.org/10.1073/pnas.2413503122>
- Kirstetter, P., Petersen, W. A., & Gourley, J. J. (2018). GPM ground validation Multi-Radar/Multi-Sensor (MRMS) precipitation reanalysis for satellite validation product [Dataset]. *NASA Global Hydrometeorology Resource Center DAAC, Huntsville, Alabama, U.S.A*. <https://doi.org/10.5067/GPMGV/MRMS/DATA101>
- Kossieris, P., Tsoukalas, I., Brocca, L., Mosaffa, H., Makropoulos, C., & Angehele, A. (2024). Precipitation data merging via machine learning: Revisiting conceptual and technical aspects. *Journal of Hydrology*, 637, 131424. <https://doi.org/10.1016/j.jhydrol.2024.131424>
- Lee, K. O., Flamant, C., Ducrocq, V., Duffourg, F., Fourrié, N., & Davolio, S. (2016). Convective initiation and maintenance processes of two back-building mesoscale convective systems leading to heavy precipitation events in Southern Italy during HyMeX IOP 13. *Quarterly Journal of the Royal Meteorological Society*, 142(700), 2623–2635. <https://doi.org/10.1002/qj.2851>
- Lei, H. J., Zhao, H. Y., & Ao, T. Q. (2022). A two-step merging strategy for incorporating multi-source precipitation products and gauge observations using machine learning classification and regression over China. *Hydrology and Earth System Sciences*, 26(11), 2969–2995. <https://doi.org/10.5194/hess-26-2969-2022>
- Li, R. Z., Guilloteau, C., Kirstetter, P. E., & Foufoula-Georgiou, E. (2023a). How well does the IMERG satellite precipitation product capture the timing of precipitation events? *Journal of Hydrology*, 620, 129563. <https://doi.org/10.1016/j.jhydrol.2023.129563>

- Li, R. Z., Guilloteau, C., Kirstetter, P. E., & Foufoula-Georgiou, E. (2023b). A new event-based error decomposition scheme for satellite precipitation products. *Geophysical Research Letters*, *50*(22), e2023GL105343. <https://doi.org/10.1029/2023GL105343>
- Li, R. Z., Guilloteau, C., Kirstetter, P. E., & Foufoula-Georgiou, E. (2025). Understanding the error patterns of multi-satellite precipitation products during the lifecycle of precipitation events for diagnostics and algorithm improvement. *Journal of Hydrology*, *651*, 132610. <https://doi.org/10.1016/j.jhydrol.2024.132610>
- Li, R. Z., Qi, D., Zhang, Y., & Wang, K. C. (2022). A new pixel-to-object method for evaluating the capability of the GPM IMERG product to quantify precipitation systems. *Journal of Hydrology*, *613*, 128476. <https://doi.org/10.1016/j.jhydrol.2022.128476>
- Li, R. Z., Wang, K. C., & Qi, D. (2018). Validating the integrated multisatellite retrievals for global precipitation measurement in terms of diurnal variability with hourly gauge observations collected at 50,000 stations in China. *Journal of Geophysical Research-Atmospheres*, *123*(18), 10423–10442. <https://doi.org/10.1029/2018jd028991>
- Li, R. Z., Wang, K. C., & Qi, D. (2021). Event-based evaluation of the GPM multisatellite merged precipitation product from 2014 to 2018 over China: Methods and results. *Journal of Geophysical Research-Atmospheres*, *126*(1), e2020JD033692. <https://doi.org/10.1029/2020JD033692>
- Lundberg, S. M., & Lee, S. I. (2017). A unified approach to interpreting model predictions. paper presented at. In *Advances in Neural information processing Systems 30 (NeurIPS 2017)*.
- Nearing, G., Cohen, D., Dube, V., Gauch, M., Gilon, O., Harrigan, S., et al. (2024). Global prediction of extreme floods in ungauged watersheds. *Nature*, *627*(8004), 559–563. <https://doi.org/10.1038/s41586-024-07145-1>
- Passive Microwave Algorithm Team Facility. (2022). GLOBAL PRECIPITATION MEASUREMENT (GPM) MISSION algorithm theoretical basis document GPROF2021 version 1 (used in GPM V7 processing) (p. 14). Retrieved from https://gpm.nasa.gov/sites/default/files/2022-06/ATBD_GPM_V7_GPROF.pdf
- Petersen, W. A., Kirstetter, P.-E., Wang, J., Wolff, D. B., & Tokay, A. (2020). The GPM ground validation program. In *Satellite precipitation measurement* (pp. 471–502). Springer International Publishing. https://doi.org/10.1007/978-3-030-35798-6_2
- Pfreundschuh, S., Brown, P. J., Kummerow, C. D., Eriksson, P., & Norrestad, T. (2022). GPROF-NN: A neural-network-based implementation of the Goddard profiling algorithm. *Atmospheric Measurement Techniques*, *15*(17), 5033–5060. <https://doi.org/10.5194/amt-15-5033-2022>
- Prokhorenkova, L., Gusev, G., Vorobev, A., Dorogush, A. V., & Gulin, A. (2018). CatBoost: Unbiased boosting with categorical features. paper presented at. In *Advances in Neural Information Processing Systems 31 (NeurIPS 2018)*.
- Rahimi, R., Ravirathinam, P., Ebtehaj, A., Behrangi, A., Tan, J., & Kumar, V. (2024). Global precipitation nowcasting of integrated multi-satellite retrievals for GPM: A U-Net convolutional LSTM architecture. *Journal of Hydrometeorology*, *25*(6), 947–963. <https://doi.org/10.1175/Jhm-D-23-0119.1>
- Romanic, D., Taszarek, M., & Brooks, H. (2022). Convective environments leading to microburst, macroburst and downburst events across the United States. *Weather and Climate Extremes*, *37*, 100474. <https://doi.org/10.1016/j.wace.2022.100474>
- Srivastava, P., Yang, R., Kerrigan, G., Dresdner, G., McGibbon, J., Bretherton, C. S., & Mandt, S. (2024). Precipitation downscaling with spatiotemporal video diffusion. paper presented at. In *Advances in Neural Information Processing Systems 37 (NeurIPS 2024)*.
- Tsuji, H., Takayabu, Y. N., Shibuya, R., Kamahori, H., & Yokoyama, C. (2021). The role of free-tropospheric moisture convergence for summertime heavy rainfall in western Japan. *Geophysical Research Letters*, *48*(18). <https://doi.org/10.1029/2021GL095030>
- Upadhyaya, S. A., Kirstetter, P. E., Kuligowski, R. J., & Searls, M. (2022). Towards improved precipitation estimation with the GOES-16 advanced baseline imager: Algorithm and evaluation. *Quarterly Journal of the Royal Meteorological Society*, *148*(748), 3406–3427. <https://doi.org/10.1002/qj.4368>
- Wang, L., Chen, H., & Li, Z. (2025). Multiscale performance of global blended satellite precipitation products over Taiwan. *Ieee Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *18*, 2108–2125. <https://doi.org/10.1109/JSTARS.2024.3499910>
- Yang, B., Finn, D., Rich, J., Gao, Z. M., & Liu, H. P. (2023). Effects of low-level jets on near-surface turbulence and wind direction changes in the nocturnal boundary layer. *Journal of Geophysical Research-Atmospheres*, *128*(11). <https://doi.org/10.1029/2022JD037657>
- Yang, Y., Zhu, Q., & Chen, J. (2024). VCformer: Variable correlation transformer with inherent lagged correlation for multivariate time series forecasting. In *Proceedings of the thirty-third international joint conference on artificial intelligence*. Jeju. <https://doi.org/10.24963/ijcai.2024/590.590>
- You, Y., Meng, H., Dong, J., & Rudlosky, S. (2019). Time-lag correlation between passive microwave measurements and surface precipitation and its impact on precipitation retrieval evaluation. *Geophysical Research Letters*, *46*(14), 8415–8423. <https://doi.org/10.1029/2019GL083426>
- You, Y. L., Wang, N. Y., Ferraro, R., & Rudlosky, S. (2017). Quantifying the snowfall detection performance of the GPM microwave imager channels over land. *Journal of Hydrometeorology*, *18*(3), 729–751. <https://doi.org/10.1175/Jhm-D-16-0190.1>