Response to the comments of Reviewer 1

Thank you for reviewing our manuscript thoroughly and providing constructive comments and valuable suggestions. Please find our point-by-point responses listed below. The reviewer's comments are in black *Italics* followed by our responses (in blue). The red text within the quotation marks is the revisions in the manuscript, while the black text is the unmodified content.

This manuscript proposes to reduced storage size of weather and climate data without compromising scientific integrity, and investigates various precision truncation strategies (combined with lossless compression) with the data from the Weather Research and Forecasting (WRF) simulations. The authors choose 2016 for the WRF simulation period, with 4-D data assimilation. Results were compared with hourly 2m air temperature and humidity and 10m wind speed, and hourly precipitation.
Metrics of relative data compression are percentage of original data when further compressed using bzip2 or gzip.

Metrics on errors due to data compression consist of RMSE of the encoded values vs. reference values, Pearson Correlation R, and Normalized Mean Bias NMB. Additional metrics for assessing impacts on extreme precipitation include number of days exceeding the 95% or 99% percentile of wet days, the maximum 1-day or 5-day precipitation total, annual count of days with daily precipitation over 10mm, count and total precipitation in wet days over a year, simple daily intensity index derived from that.

The paper is generally well written. The results are encouraging but not new (see my point below on the literature review), and the authors are not providing final compression results for the optimal strategy; it is thus unclear why the reader should actually care about doing this extra work of data compression. The paper would strongly benefit from being improved for clarity.

Thank you for your constructive feedback. Following your suggestion, we carefully revised the manuscript to improve clarity and to better articulate the motivation, methodological framework, and practical implications of our work. Our detailed, point-by-point responses outlining the exact modifications made to address each of your specific comments and suggestions are provided below.

Overall, we substantially revised the manuscript in several aspects. First, the research background and literature context have been expanded to better position our work within existing studies on atmospheric data compression. Second, the scientific motivation and practical relevance of evaluating precision reduction within a WRF modeling workflow are now stated more explicitly in the Introduction section and Abstract. Third, the presentation of compression performance has been enhanced to more explicitly demonstrate the quantitative storage gains achieved across different precision-reduction configurations. Fourth, the methodological framework and evaluation analyses have been refined and expanded, including additional diagnostics of local deviation, structural fidelity and downstream scientific impacts. These revisions collectively improve the clarity of the manuscript.

Finally, to improve clarity and avoid potential confusion, we standardized the terminology throughout the manuscript. The term "precision truncation" used in the previous version has been replaced with “precision reduction” or “decimal significant-digit rounding,” and expressions such as “truncation strategies” have been replaced by “precision reduction configurations.”

The fundamental limitation of the paper is that it is not properly situated in the comprehensive literature of data truncation and data compression, beyond three references: Baker et al (2016), Poppick et al (2020, lossy), Walters and Wong (2023).
The following work extensively investigated truncation strategies:
M Klöwer, M Razinger, JJ Dominguez, PD D uben, TN Palmer, Compressing atmospheric data into its real information content. Nat. Comput. Sci. 1, 713–724 (2021).
Moreover, several works have explored neural lossy compression:
L Huang, T Hoefler, Compressing multidimensional weather and climate data into neural networks. ICLR (2023).
T Han, S Guo, W Xu, L Bai, et al., Cra5: Extreme compression of era5 for portable global climate and weather research via an efficient variational transformer. arXiv preprint arXiv:2405.03376 (2024).
P Mirowski, D Warde-Farley, M Rosca, et al., Neural compression of atmospheric states. arXiv preprint arXiv:2407.11666 (2024).

Thank you for pointing out that the original manuscript did not sufficiently situate our study within the broader literature. Following this suggestion, we conducted a substantially expanded literature review and revised the Introduction (Lines 54–80) to more comprehensively position our work within the existing research landscape.

First, we incorporated the reviewer-recommended studies and clarified their relevance to our work. In particular, Klöwer et al. (2021) reframed atmospheric data compression from an information-theoretic perspective, demonstrating that the intrinsic precision requirements of atmospheric variables vary substantially and that a large portion of stored floating-point precision is redundant. This perspective strongly supports the motivation of our study and highlights the importance of variable-dependent precision reduction strategy. In addition, recent work on neural-network-based compression methods (Huang and Hoefler, 2023; Han et al., 2024; Mirowski et al., 2024) was added to place our approach within the broader context of emerging compression technologies.

Second, we expanded the discussion of classical precision reduction techniques and related evaluation frameworks. The revised Introduction now also references additional foundational studies, including Zender (2016) on bit grooming and unbiased quantization, Delaunay et al. (2019) on the Digit Rounding algorithm bridging decimal and binary representation, Silver and Zender (2017) on compression–error trade-offs in geophysical datasets, Baker et al. (2019) on structural fidelity diagnostics, and technical investigations of netCDF/HDF compression performance (Delaunay et al., 2019; Underwood et al., 2022). These revisions substantially broaden the literature context and more clearly distinguish our focus on analyzing the impacts of precision reduction across the entire WRF model workflow, including both input and output components.

Third, inspired by the evaluation framework used in Baker et al. (2019) and related studies, we strengthened the methodological analysis of structural fidelity in our manuscript. Specifically, we introduced the Structural Similarity Index Measure (SSIM) as a complementary diagnostic to grid-scale numerical deviations (Section 2.4: Lines 256–273), which is presented in the newly added Section 3.3 titled "Maximum Deviations and Structural Fidelity". This section systematically analyzes the relationship between maximum absolute deviations and spatial structural similarity across precision-reduction configurations, providing additional insight into how precision reduction, especially for the input, affects the spatial integrity of simulated meteorological fields. The conclusion is as follows (Lines 623–640, Lines 726–730): Precision reduction applied to time-varying inputs introduces perturbations that interact with nonlinear model dynamics during integration. Although the large-scale atmospheric structures remain morphologically coherent, the resulting spatial phase shifts can substantially affect localized diagnostics (e.g., precipitation and 2-m temperature). Consequently, input precision reduction should generally be avoided when exact spatiotemporal correspondence is required, while it remains a viable approach for prioritizing the preservation of overall statistical distributions.

Lines 54–80:

"A variety of compression techniques applicable to atmospheric model archives have been developed to alleviate the rapidly growing storage demands of numerical simulations. Lossless algorithms alone preserve bitwise reproducibility but generally achieve only a modest compression ratio for floating-point geophysical fields because the high entropy of mantissa bits limits compressibility (e.g., Poppick et al., 2020). To address this limitation, combining precision reduction with lossless compression has emerged as a widely explored strategy for improving storage efficiency while attempting to preserve scientific fidelity. Within this paradigm, many approaches focus on manipulating the least significant bits of floating-point data. Early approaches such as bit shaving (Zender, 2016) reduce precision by zeroing trailing mantissa bits, which can introduce systematic bias. Zender (2016) subsequently introduced bit grooming, a statistically robust quantization approach that alternates bit shaving and bit setting of trailing IEEE mantissa bits, thereby preserving the mean in expectation while improving compressibility. Building on this line of work, Delaunay et al. (2019) introduced the Digit Rounding algorithm, which bridges decimal precision control and binary representation. Compared to bit grooming, Digit Rounding optimizes the allocation of mantissa bits required, thereby improving compression efficiency while retaining controlled numerical precision. In contrast to hybrid methods like Digit Rounding, alternative approaches regulate precision purely within the decimal domain. For example, decimal significant-digit rounding (Walters and Wong, 2023) directly constrains the number of significant digits, yielding a highly transparent and interpretable approach to precision reduction.

Alongside algorithmic developments, some studies have expanded the evaluation of precision reduction beyond simple error statistics toward broader notions of statistical and structural fidelity. Baker et al. (2016) proposed an ensemble-based framework to assess whether compressed datasets remain statistically indistinguishable from internal climate variability, while Baker et al. (2019) emphasized the importance of evaluating structural integrity in compressed climate data. Silver and Zender (2017) quantified the compression–error trade-off for gridded datasets, demonstrating that carefully designed precision reduction can remove substantial false precision with limited impact on scientific conclusions. Complementary technical investigations have examined how data representation and codec behavior influence achievable compression ratios within netCDF and HDF data formats (Delaunay et al., 2019; Underwood et al., 2022). Recently, Klöwer et al. (2021) reframed atmospheric data compression from an information-theoretic perspective, demonstrating that the intrinsic precision requirements of atmospheric variables vary substantially and that a large fraction of stored floating-point precision is redundant. In parallel, machine-learning-based compression methods have emerged that exploit nonlinear data manifolds to achieve very high compression ratios (Huang and Hoefler, 2023; Han et al., 2024; Mirowski et al., 2024)."

Lines 256–273:

"Second, to evaluate the impacts of precision reduction on both local numerical accuracy and spatial structures, we analyzed point-wise deviations together with spatial similarity metrics. At the grid scale, we computed the absolute grid-scale deviation (AD), defined as the absolute difference between the precision-reduced configurations and the full-precision baseline simulation WRF_bl at each grid point and hourly time step. Because point-wise metrics alone cannot capture changes in the spatial organization of meteorological fields, we additionally employed the Structural Similarity Index Measure (SSIM) (Wang et al., 2004; Baker et al., 2019; Klöwer et al., 2021). This metric is particularly important for evaluating simulations driven by precision-reduced inputs, where small numerical perturbations may propagate through nonlinear dynamics and alter evolving mesoscale structures. Unlike AD, which measures local differences, SSIM quantifies the preservation of large-scale spatial patterns and structural textures. For a full-precision reference spatial window x and the corresponding window y from precision-reduced configurations, the SSIM is calculated as:

$$\mathrm{SSIM}(x,y)=\frac{(2\mu_x\mu_y+C_1)(2\sigma_{xy}+C_2)}{(\mu_x^2+\mu_y^2+C_1)(\sigma_x^2+\sigma_y^2+C_2)} \tag{3}$$

where μ_x and μ_y are the local means, σ_x^2 and σ_y^2 are the local variances, and σ_{xy} is the local covariance between the two fields. The stabilization constants C_1 and C_2 are scaled by the dynamic range of the specific meteorological variable being evaluated to accommodate diverse atmospheric fields. SSIM values range from 0 to 1, with unity indicating perfect structural similarity. Structural similarity was computed using an 11×11 Gaussian weighting window sliding across the domain. Given the 12 km horizontal grid spacing, this corresponds to a spatial footprint of approximately 130 km, enabling assessment of mesoscale structural consistency. To avoid artificial inflation of scores caused by large, structurally uniform dry regions, a standard wet-day mask with a threshold of 0.1 mm h^{-1} was applied prior to the calculation of precipitation SSIM."

Lines 623–640:

"**4.1 Implications of Input Precision Reduction**

Precision reduction applied to time-varying model inputs introduces perturbations that interact with nonlinear atmospheric dynamics during model integration. These perturbations propagate through the simulation and may alter the precise spatial and temporal positioning of weather systems. As demonstrated in Section 3.3, such perturbations can generate large grid-scale maximum AD. However, the simultaneously high SSIM values indicate that the overall morphology of the simulated atmospheric systems remains largely preserved. Consequently, the resulting errors primarily manifest as pronounced maximum AD driven by local spatial or temporal phase shifts, accompanied by a background of widespread, disordered weak perturbations, rather than a fundamental structural collapse of the simulated fields. Therefore, input precision reduction may be acceptable for applications that prioritize large-scale statistics or aggregated diagnostics, where exact grid-scale correspondence is not required. However, this pathway introduces intrinsic nondeterministic perturbations into the simulation, meaning that the exact location and timing of individual weather systems may shift. For impact-oriented applications requiring strict spatiotemporal consistency, such as event attribution or local hydrological analyses, this loss of deterministic correspondence may be undesirable.

In addition, the chaotic amplification of these input perturbations exhibits strong seasonal dependence. Error propagation peaks during the summer when local land-atmosphere feedbacks and convective processes dominate. Hypothetically, aggressive precision reduction could be applied to input forcings during less sensitive, non-summer months, while retaining full precision during the dynamically volatile summer to help constrain the cascading growth of nonlinear errors. However, the extent to which this adaptive approach would restore summer fidelity remains an open question warranting further investigation."

Lines 726–730:

"Precision reduction applied to time-varying inputs introduces perturbations that interact with nonlinear model dynamics during integration Although the large-scale atmospheric structures remain morphologically coherent, the resulting spatial phase shifts can substantially affect localized diagnostics (e.g., precipitation). Consequently, input precision reduction should generally be avoided when exact spatiotemporal correspondence is required, while it remains a viable approach for preserving aggregate statistical distributions."

How does this work differ from the conclusions in all these previous works - is it by using the compressed data as inputs to WRF simulation? This should be made explicit.

Thank you for the suggestion. As you correctly noted, a key distinction of our work compared with previous studies is the explicit evaluation of precision reduction applied to time-varying input forcings used in WRF simulation. Most previous studies on atmospheric data compression primarily evaluated truncation strategies using static datasets, such as reanalysis products or model outputs, without propagating the compressed data through a dynamical model integration. In contrast, our study investigates precision reduction within the full WRF workflow, applying significant-digit rounding to both time-varying WRF input forcings and model outputs, and evaluating how these perturbations propagate through the nonlinear dynamics of the model.

In the revised manuscript, we have clarified this distinction more explicitly in the Abstract (Lines 20–24), the Introduction (Lines 81–94), and the Conclusions (Lines 726–730). We emphasize that precision reduction applied to time-varying inputs can manifest as spatial or temporal phase shifts in simulated weather systems. While the large-scale atmospheric structures generally remain morphologically coherent, the resulting phase shifts can substantially affect localized diagnostics. Consequently, our results show that input precision reduction may be acceptable for applications focusing on aggregate statistical properties but should generally be avoided when exact spatiotemporal correspondence is required.

In addition to evaluating input versus output precision reduction, our study also extends previous work by systematically examining how precision reduction affects downstream scientific diagnostics. The revised analysis framework now spans multiple levels of evaluation, including domain-scale statistical errors, regional extreme deviations, spatial structural fidelity, and downstream diagnostic metrics.

The prime focus of our study is precipitation, which represents one of the most challenging variables for compression because of its highly skewed distribution, intermittency, and cumulative representation in WRF (RAINNC + RAINC). To more comprehensively evaluate compression impacts across the full precipitation spectrum, we expanded our analysis beyond the extreme precipitation indices used in the original manuscript. In the revised Section 3.4, we also include diagnostics that explicitly assess the lower tail of the precipitation distribution, including precipitation occurrence metrics such as the Probability of Detection (POD), False Alarm Ratio (FAR), Critical Success Index (CSI), Frequency Bias (Bias), and a Zero Preservation Ratio (ZPR). These additions allow us to evaluate how precision reduction affects both light precipitation occurrence and extreme precipitation diagnostics. Together, these new analyses reveal that precision reduction applied to inputs and outputs produces fundamentally different impacts on downstream diagnostics, and they allow us to systematically identify safe precision thresholds for precipitation.

The methodological exposition and definition of the precipitation lower tail index have been incorporated into the revised manuscript (Section 2.4: Lines 274–283, Table 4). The extended analysis section is in Section 3.4 (Lines 511–555, Table 9).

Lines 20–24:

"Scientifically, precision-reduced inputs interact with nonlinear model dynamics and can induce spatial phase shifts in simulated meteorological systems. Although this process reduces deterministic grid-scale correspondence, the overall spatial morphology of the atmospheric structures remains largely preserved. Consequently, aggregate statistical distributions are weakly affected, especially during dynamically less active periods, rendering input precision reduction suitable for large-scale spatial aggregates and event-averaged statistical analyses."

Lines 81–94:

"Despite these advancements, most existing studies primarily consider precision reduction as a post-processing operation applied to static datasets or model outputs. Comparatively less attention has been paid to how precision reduction applied to time-varying model inputs propagates through nonlinear numerical integrations. A fundamental distinction exists between the precision reduction of input data and that of output data. Perturbations introduced into the input forcings may be amplified through nonlinear model physics, potentially influencing simulated trajectories and downstream diagnostics. Such responses are unlikely to be spatially or temporally uniform, as sensitivity to perturbations depends on the prevailing dynamical and thermodynamic regime.

Crucially, the extent to which precision reduction applied to model inputs and outputs propagates into downstream diagnostic analyses remains largely unquantified. While the fidelity of fundamental thermodynamic and kinematic fields (e.g., temperature and wind) must be systematically verified, this uncertainty becomes particularly critical in downstream diagnostic studies of precipitation, which is output as a cumulative quantity and exhibits intermittent and highly skewed characteristics. Given the disproportionate societal and economic impacts of extreme precipitation (Seneviratne et al., 2021; Davenport et al., 2021), it is therefore essential to assess whether precision reduction introduces artificial biases in extreme-event diagnostics, such as inflated dry-area coverage or altered peak precipitation intensities, that exceed expected analysis uncertainty."

Lines 274–283:

"Finally, the evaluation of precipitation for both the lower-tail thresholds that control the occurrence of light rainfall and the behavior of upper-tail extreme events is necessary given the highly skewed distribution of precipitation fields. Moreover, because total precipitation in WRF is represented as the cumulative sum of grid-resolved (RAINNC) and parameterized convective (RAINC) components, it exhibits a compounded sensitivity to numerical precision loss.

To assess structural fidelity at the lower bound of the precipitation spectrum, we first performed a dichotomous verification for hourly precipitation, using a wet–dry threshold of 0.1 mm h^{-1}. For each precision-reduced configuration, grid cells were classified relative to WRF_bl. Based on this comparison, several categorical verification metrics were calculated (Roebber, 2009), including the Probability of Detection (POD), False Alarm Ratio (FAR), Critical Success Index (CSI), and Frequency Bias (Bias). In addition, a Zero Preservation Ratio (ZPR) was introduced to quantify the fraction of dry grid cells that remain correctly classified after precision reduction."

Lines 511–555:

"3.4 Impacts on Precipitation Diagnostics: Zero-Value Preservation and Extreme Precipitation Indices

While the SSIM analysis reveals how these numerical artifacts alter the spatial structure of precipitation fields, an equally important question is how distortions introduced by precision reduction in cumulative precipitation fields propagate into downstream scientific diagnostics. Precipitation exhibits a highly skewed spatial and temporal distribution, characterized by two distinct and sensitive regimes: a widespread lower tail dominated by zero or near-zero values, and a rare but high-impact upper tail driven by extreme rainfall events. Both regimes are particularly vulnerable to numerical artifacts introduced by precision reduction. To quantify these impacts, the following section evaluates how precision-reduced configurations affect precipitation diagnostics, including zero-value preservation and extreme precipitation indices.

We first evaluate the preservation of the precipitation occurrence spectrum, specifically focusing on the zero-value boundary (using the 0.1 mm h^{-1} wet-day threshold). Standard contingency metrics, including ZPR, POD, FAR, CSI, and Bias, are employed to diagnose how different compression configurations alter the fundamental wet/dry morphology (Table 9).

For input-only precision reduction configurations, the metrics indicate that the overall precipitation spectrum is largely preserved. The Frequency Bias remains close to unity (e.g., 0.99942 for WRF_3), confirming that the total precipitation area is conserved. Furthermore, the nearly symmetric error rates, where the moderate FAR (7.84%–8.72%) closely balances the corresponding miss rates (1 - POD), together with a stable CSI (~0.84) indicate that input precision reduction induces spatial phase shifts in precipitation systems rather than systematically disrupting their structure. In contrast, aggressive output precision reduction introduces substantial numerical distortions that directly erode the bottom end of the precipitation spectrum, most prominently in the WRF_fx3 configuration. Under this configuration, the POD drops sharply to 71.94%, indicating that approximately 28% of valid light precipitation events are artificially truncated to zero due to the loss of significance in the arithmetic of cumulative precipitation variables. As a result, a pronounced systematic dry bias emerges (Bias ≈ 0.77).

Interestingly, WRF_fx3 exhibits an apparently higher ZPR (99.12%) and a lower FAR (6.90%) compared with WRF_fx4 (98.5% and 8.98%, respectively). However, this does not indicate improved structural fidelity. Instead, it reveals that WRF_fx3 suppresses both numerical noise (reducing FAR) and genuine light stratiform precipitation (reducing POD), effectively converting them into widespread non-physical dry regions.

The structural degradation previously indicated by the late-stage decline in SSIM is further corroborated by the temporal evolution of precipitation occurrence metrics under the WRF_fx4 configuration (Fig. S4). At the domain scale, the ZPR remains high (> 97 % throughout the year), largely reflecting the overwhelming dominance of climatologically dry background grids. However, metrics that evaluate performance specifically within precipitation events (CSI, POD, and FAR) reveal a pronounced late-stage deterioration. Consistent with the SSIM decline observed in November, the CSI decreases from its summer peak (~0.89) to 0.769 by December. This degradation reflects a dual numerical distortion associated with the expanding accumulation baseline. As the quantization step increases during November and December, the FAR rises substantially (reaching 11.4%), indicating the growing occurrence of spurious drizzle signals introduced by rounding artifacts. At the same time, the POD declines to 85.4%, implying that approximately 15% of genuine winter stratiform precipitation events are artificially truncated to zero. The Frequency Bias further highlights this shift in behavior. While the system exhibits a slight wet bias in spring, it gradually transitions to a systematic dry bias by December (Bias = 0.964). The simultaneous increase in false alarms and the loss of weak precipitation events progressively disrupt the spatial consistency of precipitation structures, definitively explaining the late-stage SSIM collapse.

The analyses above focus on the lower boundary of the precipitation spectrum, where precision reduction primarily affects the occurrence of light rainfall and the preservation of dry conditions. However, precipitation diagnostics are also strongly influenced by the opposite end of the distribution."

Lines 726–730:

"Precision reduction applied to time-varying inputs introduces perturbations that interact with nonlinear model dynamics during integration. Although the large-scale atmospheric structures remain morphologically coherent, the resulting spatial phase shifts can substantially affect localized diagnostics (e.g., precipitation). Consequently, input precision reduction should generally be avoided when exact spatiotemporal correspondence is required, while it remains a viable approach for preserving aggregate statistical distributions."

Table 4. Definitions of categorical verification metrics for hourly precipitation and daily extreme precipitation indices.

Name	Definition	Units
ZPR	Percentage of dry grids (< 0.1 mm h^{-1}) in WRF_bl that remain below the threshold after precision reduction.	%
POD	Percentage of wet grids (≥ 0.1 mm h^{-1}) in WRF_bl that are correctly retained after precision reduction.	%
FAR	Percentage of wet grids after precision reduction that are false alarms relative to WRF_bl.	%

CSI	A comprehensive metric for the spatial fidelity of the precipitation field, penalizing both the artificial elimination (missed events) and generation (false events) of wet events.	-
Bias	Ratio of the total number of wet grids in the precision-reduced configurations to that in WRF_bl. Values > 1 indicate an artificial inflation of precipitation spatial extent.	-
R95p_days	Number of days per year with daily precipitation exceeding the 95th percentile of wet-day amounts (≥ 1 mm), thresholds derived from the 2001–2015 baseline period.	days
R99p_days	Same as R95p_days, but for the 99th percentile threshold.	days
Rx1_day	Maximum 1-day precipitation total in a year.	mm
Rx5_day	Maximum total precipitation accumulated over any consecutive 5-day period.	mm
R10mm_days	Annual count of days with daily precipitation ≥ 10 mm.	days
PRCPTOT	Total annual precipitation from wet days.	mm
wet_days	Annual count of wet days (≥ 1 mm).	days
SDII	Simple Daily Intensity Index, calculated as PRCPTOT divided by wet_days.	mm day^{-1}

Table 9. Categorical verification metrics for hourly precipitation across different precision reduction configurations relative to the WRF_bl.

Configuration	ZPR (%)	POD (%)	FAR (%)	CSI	Bias
WRF_5	98.7086	92.1787	7.8366	0.8548	1.00017
WRF_4	98.7078	92.1526	7.8431	0.8545	0.99995
WRF_3	98.5641	91.2276	8.7198	0.8391	0.99942
WRF_fx5	99.8289	99.2410	1.0357	0.9822	1.00280
WRF_fx4	98.4663	94.3254	8.9820	0.8629	1.03634
WRF_fx3	99.1221	71.9411	6.8957	0.6830	0.77269
WRF_5fx5	98.6073	91.8428	8.4280	0.8468	1.00296
WRF_5fx4	97.4814	88.3594	14.7486	0.7664	1.03646
WRF_5fx3	98.5405	68.4151	11.4639	0.6285	0.77274
WRF_4fx5	98.6069	91.8190	8.4321	0.8466	1.00274
WRF_4fx4	97.4812	88.3404	14.7522	0.7663	1.03628
WRF_4fx3	98.5400	68.4076	11.4680	0.6284	0.77269
WRF_3fx5	98.4697	90.9292	9.2680	0.8320	1.00217
WRF_3fx4	97.3682	87.6111	15.4209	0.7554	1.03585
WRF_3fx3	98.4717	67.9713	12.0081	0.6220	0.77247

*Note: A threshold of 0.1 mm h^{-1} is applied to distinguish between dry and wet grids. ZPR is the Zero Preservation Ratio, POD denotes the Probability of Detection, FAR is the False Alarm Ratio, CSI is the Critical Success Index, and Bias represents the Frequency Bias.

The article would benefit from an illustration of what are the input and output variables for the Weather Research Forecasting models, and a schematic of how the data interact.

Thank you for this helpful suggestion. The initial atmospheric state is provided via either an initial condition file, typically for the very first day of the study, or a restart file (wrfrst) which is generated after each day simulation, to serve as an initial condition for a subsequent day of simulation. Either initial condition file or wrfrst file, it provides full three-dimensional atmospheric and land-surface prognostic states required for seamless temporal continuity. For consistency purposes, the annual run started with a 2-day spin-up on 12/30/2015 so every day in 2016 the model started with a restart file and four prescribed time-varying external forcing files, which include atmospheric and surface analysis nudging fields (wrffdda, wrfsfdda), lateral boundary condition tendencies (wrfbdy), and lower boundary inputs such as sea surface temperature (wrflowinp). The outputs are written to wrfout files and contain near-surface and surface variables, three-dimensional atmospheric state fields, cloud microphysical quantities, radiation and energy fluxes, and other diagnostics. We have systematically detailed the exact data structures of the WRF model's inputs and outputs in the revised manuscript (Section 2.1: Lines 136–143, 145–147).

The entire precision reduction workflow is visually anchored by the newly introduced Figure 1. As detailed in the revised manuscript (Section 2.2: Lines 185–194), we delineate the workflow into preprocessing, model integration, and post-processing stages. Specifically, we clarify that input precision reduction exclusively targets time-varying forcing files (wrffdda, wrfsfdda, wrfbdy, wrflowinp) during the preprocessing stage. Evaluating each input precision level strictly necessitates a fully independent WRF simulation to physically capture the non-linear propagation of rounding errors. Conversely, output precision reduction is explicitly defined as a static, offline post-processing operation applied to the final wrfout files without the need to rerun the model.

Lines 136–143:

"To establish dynamical consistency prior to the evaluation period, the annual simulation for 2016 was preceded by a 2-day spin-up initialized on 30 December 2015. The annual integration was conducted in consecutive daily segments, with each day initialized from a model restart file (wrfrst) generated at the conclusion of the previous day. These restart files provide the three-dimensional atmospheric and land-surface state required for seamless temporal continuity. In addition to daily initialization, the model evolution was continuously constrained by prescribed time-varying external forcing fields, including atmospheric and surface analysis nudging (wrffdda, wrfsfdda), lateral boundary condition tendencies (wrfbdy), and lower-boundary updates such as sea surface temperature (wrflowinp)."

Lines 145–147:

"In contrast, model outputs (wrfout) contain near-surface variables, three-dimensional atmospheric states, cloud microphysical quantities, and energy fluxes, representing the prognostic results of the completed simulation."

Lines 185–194:

"The systematic integration of this rounding framework into the overall experimental design is summarized in Fig. 1. The schematic depicts the three core workflow stages: preprocessing on input with the precision-reduction tool, model integration with full-precision input and precision-reduction input, and post-processing on model integration result, wrfout file, with the precision-reduction tool. Lossless compression will be applied to all input and output. For input precision reduction configurations, the rounding procedure was applied during the preprocessing stage, exclusively targeting the time-varying forcing files (wrffdda, wrfsfdda, wrfbdy, and wrflowinp). Because these files define the dynamic external constraints on the model, evaluating each input precision level necessitated a fully independent WRF simulation to capture the non-linear propagation of rounding errors. Conversely, for output precision reduction, rounding was applied to the final wrfout files after simulation completion. As a purely static post-processing operation, output precision can be flexibly adjusted without requiring computationally expensive model re-runs."

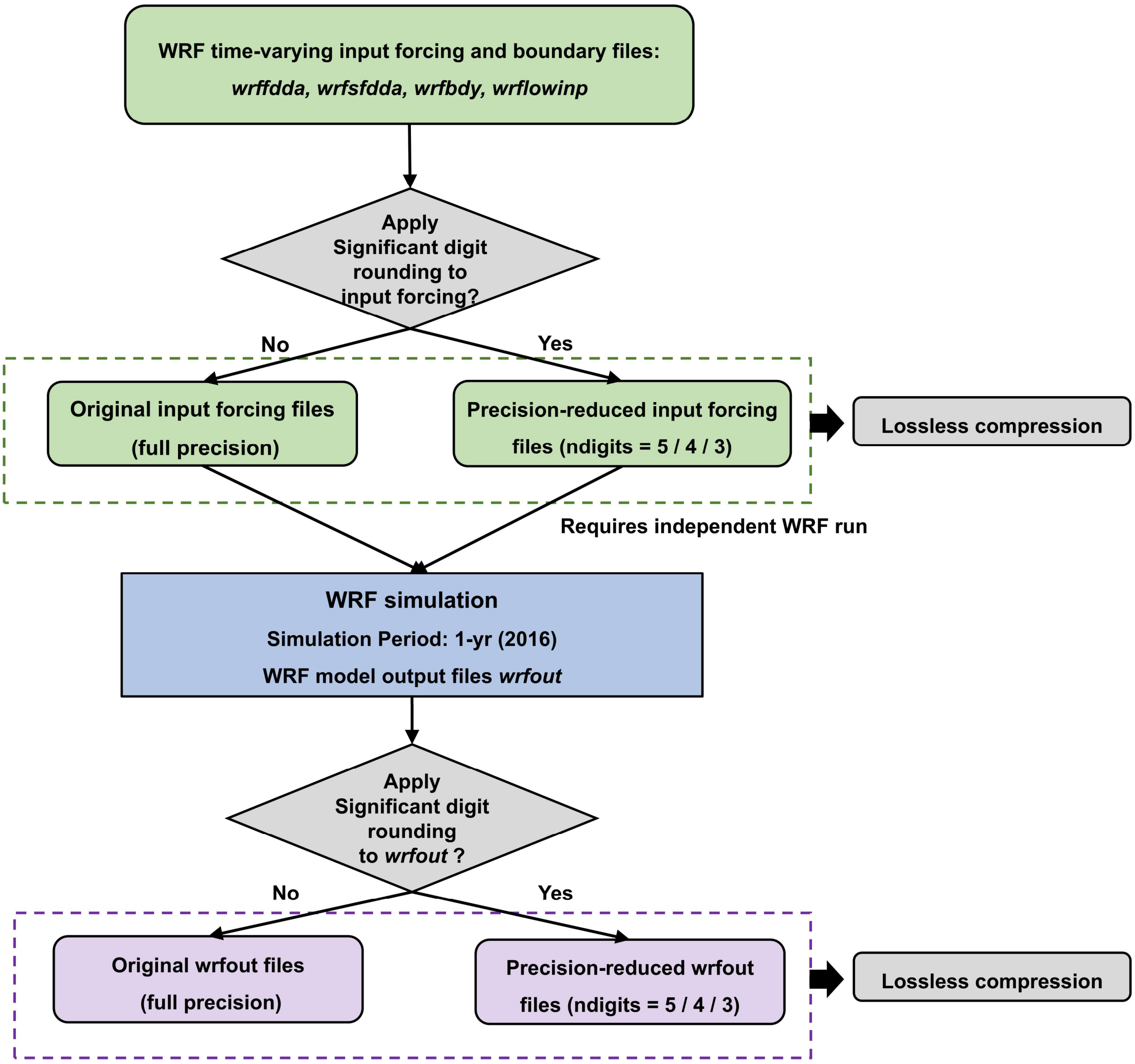

Figure 1: Schematic illustration of the experimental workflow for WRF data precision reduction and lossless compression.

Am I right that the truncation of input data to WRF has an impact on the output results coming from the WRF, and that this is the reason why, given the same output truncation, different input truncations can reduce the relative compression size of the outputs?

Thank you for this question. You are correct from a modeling perspective: input precision reduction effectively acts as a small perturbation. Due to the nonlinear dynamics of the WRF model, these perturbations definitely propagate and produce quantitatively different final output fields (as we evaluate extensively in our scientific impact analysis). However, regarding the compressibility (storage footprint) of these outputs, our results indicate that the impact of these input perturbations is actually negligible. Modifying the input precision does not meaningfully change the information entropy or the overall numerical distribution of the final outputs in a way that affects the performance of the backend lossless codecs.

We realize that our original visualization in Figure 2 caused this exact misunderstanding. In the original manuscript, Figure 2 included three subplots: (a) input relative compression ratio, (b) output relative compression ratio, and (c) total (combined) relative compression ratio. You reasonably observed in subplot (c) that under the same output precision reduction, applying different input precision reduction improved the combined compression ratio. However, because the calculation for subplot (c) pooled both datasets (inputs + outputs) together, the apparent improvement in the total relative compression ratio was driven entirely by the size reduction of the input dataset itself, not by any increased compressibility of the output dataset.

Recognizing that the combined metric in the original subplot (c) was misleading, we have removed it. We believe that presenting the input and output compression ratios separately is much clearer and accurately conveys our results. This updated visualization is now presented as Figure 3 in the revised manuscript. We have also added an explicit statement in the revised manuscript (Section 3.1: Lines 309–311) to clarify that while input precision reduction perturbs the physical output values, its impact on the ultimate compressibility of the model outputs is negligible.

<u>Lines 309–311:</u>

"As physically expected, the final compressed volume of the output is predominantly dictated by the precision applied during the output post-processing stage; varying the input precision exerts no immediate influence on the final compressed wrfout file size."

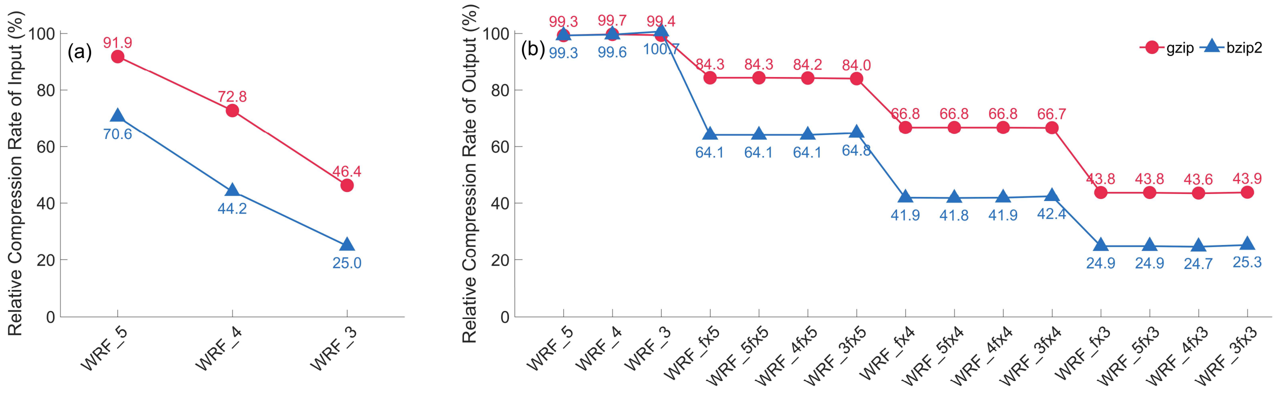

Figure 3: Relative compression ratios of (a) input data and (b) output data across diverse precision-reduction configurations.

Are input variables forcing variables, or are they also weather data? It is only on line 325 that we can infer that output variables are not recursively fed back into the model, since output-only truncation can happen after the model is run.

Thank you for the question. In this study, the WRF input data encompasses both internal weather states and external forcing variables. In our workflow, precision reduction was strictly applied only to these time-varying external forcing fields, leaving the internal weather states within the restart files intact. We have explicitly defined the targeted time-varying input forcings in the revised Methodology (Section 2.1: Lines 129–131; Section 2.2: Lines 188–190).

Regarding the model outputs, you are correct that the output variables are not recursively fed back into the model integration. The output precision reduction evaluated in this study functions strictly as an offline, post-processing filter. We have explicitly stated the non-recursive nature of the output in the revised manuscript (Section 2.2: Lines 192–194). The newly added Figure 1 in the revised manuscript also illustrates this information.

Lines 129–131:

"The sheer volume of these high-resolution input arrays significantly exacerbates the overall storage and data management burden. Consequently, these massive time-varying input files represent a critical target for our proposed precision reduction."

Lines 188–190:

"For input precision reduction configurations, the rounding procedure was applied during the preprocessing stage, exclusively targeting the time-varying forcing files (wrffdda, wrfsfdda, wrfbdy, and wrflowinp)."

Lines 192–194:

"Conversely, for output precision reduction, rounding was applied to the final wrfout files after simulation completion. As a purely static post-processing operation, output precision can be flexibly adjusted without requiring computationally expensive model re-runs."

The relationship between the 1622 stations and the surface data is unclear. Do the authors have access to simulated and data-assimilated dense surface data?

Thank you for the question. The 1622 stations represent the independent, point-based observational dataset used for evaluation. To establish a clear spatial relationship between these two distinct data collections, we applied a nearest-neighbor interpolation method to align the point-based observations with the simulated model grid. Specifically, each station was spatially matched to the WRF grid point with the smallest geographical distance. We have explicitly detailed this spatial matching procedure in the revised manuscript (Section 2.3: Lines 219–223).

Considering that you might still be concerned about whether the data used for evaluation and the nudging method we employed in the simulation process are in conflict, we emphasize that we used weak nudging coefficients for the free troposphere, which acts as a large-scale constraint to anchor the synoptic circulation. Atmospheric nudging was explicitly disabled within the planetary boundary layer and at the surface to prevent the suppression of near-surface thermodynamic variability. We have supplemented the details regarding the application of the nudging method in our simulation (Section 2.1: Lines 116–129).

Lines 116–129:

"To prevent systematic model drift during the long-term integration, the four-dimensional data assimilation (FDDA) via grid-nudging was employed within the WRF framework, utilizing the North American Mesoscale (NAM) analysis, produced operationally by the National Centers for Environmental Prediction (NCEP). In model settings, weak nudging coefficients (1.0×10^{-4} s^{-1} for wind and temperature, and 1.0×10^{-5} s^{-1} for moisture) were deliberately selected for the free troposphere. This configuration acts as a large-scale constraint to anchor the synoptic circulation while granting the model sufficient degrees of freedom to generate free-evolving, high-resolution mesoscale features. Furthermore, atmospheric nudging was explicitly disabled within the planetary boundary layer and at the surface to avoid suppressing near-surface thermodynamic variability. Because the boundary-layer and surface atmospheric fields evolved entirely according to model physics, subsequent evaluations of the WRF outputs against independent observational station data remain scientifically valid and uncontaminated by the direct assimilation process.

To mitigate long-term model drift in deep soil temperature and moisture, we employed the Pleim-Xiu land surface model combined with indirect soil nudging, which specifically requires the continuous ingestion of two-dimensional surface analysis fields (wrfsfdda). Coupled with the three-dimensional atmospheric fields (wrffdda) utilized for upper-air nudging, the model is forced to ingest massive datasets every 3 hours."

Lines 219–223:

"To spatially align the point-based measurements with the gridded model output, a nearest-neighbor interpolation scheme was applied, assigning each station to the WRF grid point with the smallest Euclidean distance in the latitude–longitude space. Despite the inherent representativeness limitations typical of point-to-grid regional climate validation, this extensive 1622-station network provides a robust benchmark for assessing the model's meteorological fidelity at both regional and CONUS-wide scales."

What are the inputs and outputs to the WRF? Do the authors re-run the WRF at different input data truncation strategies?

Thank you for the question. Regarding the specific definitions of WRF inputs and outputs within our workflow, we have described in our previous response (please refer to that reply).

Concerning the experimental execution, we reiterate that a completely separate, independent WRF simulation was conducted with distinct input (full precision input and precision-reduced input). This independent integration strategy is strictly necessary to physically capture how errors non-linearly propagate through the model dynamics. To make this parallel experimental design unambiguously clear, we have added an explicit declaration in the revised manuscript (Section 2.2: Lines 190–192). The newly added Figure 1 in the revised manuscript also illustrates this scheme.

Lines 190–192:

"Because these files define the dynamic external constraints on the model, evaluating each input precision level necessitated a fully independent WRF simulation to capture the non-linear propagation of rounding errors."

What is the N in the error formulas: is it the number of input/output datapoints in 2016? Or is it the number of discrete observation station measurements? How is the distribution of observation points compared to that of the data used in the WRF?

Thank you for the question. The term N dynamically represents the total number of valid, spatiotemporally matched data pairs used in each specific comparison. Specifically, for validation against discrete ground-based observations, N corresponds to the total number of valid station-hour pairs across the NCDC network (where each station measurement is spatially matched to its nearest WRF grid point). Conversely, for comparisons against gridded satellite products, which were spatially interpolated directly onto the WRF model grid, N refers to the total number of grid-time pairs within the WRF domain. We have incorporated this explicit definition of N in the revised manuscript (Section 2.4: Lines 253–255).

Regarding the spatial relationship between the datasets, the WRF model generates a continuous, high-resolution gridded surface over the contiguous United States (CONUS). In contrast, the 1622 NCDC surface stations provide a discrete, albeit dense, spatial sampling network. The geographical distribution of these observation stations explicitly visualized in the manuscript (presented as Figure 1a in the original submission, and now as Figure 2a in the revised manuscript).

Lines 253–255:

"For station-based validation against NCDC observations, N corresponds to the number of valid station–hour pairs across all stations. For comparisons against gridded satellite products, N corresponds to the total number of grid–time pairs within the evaluated domain."

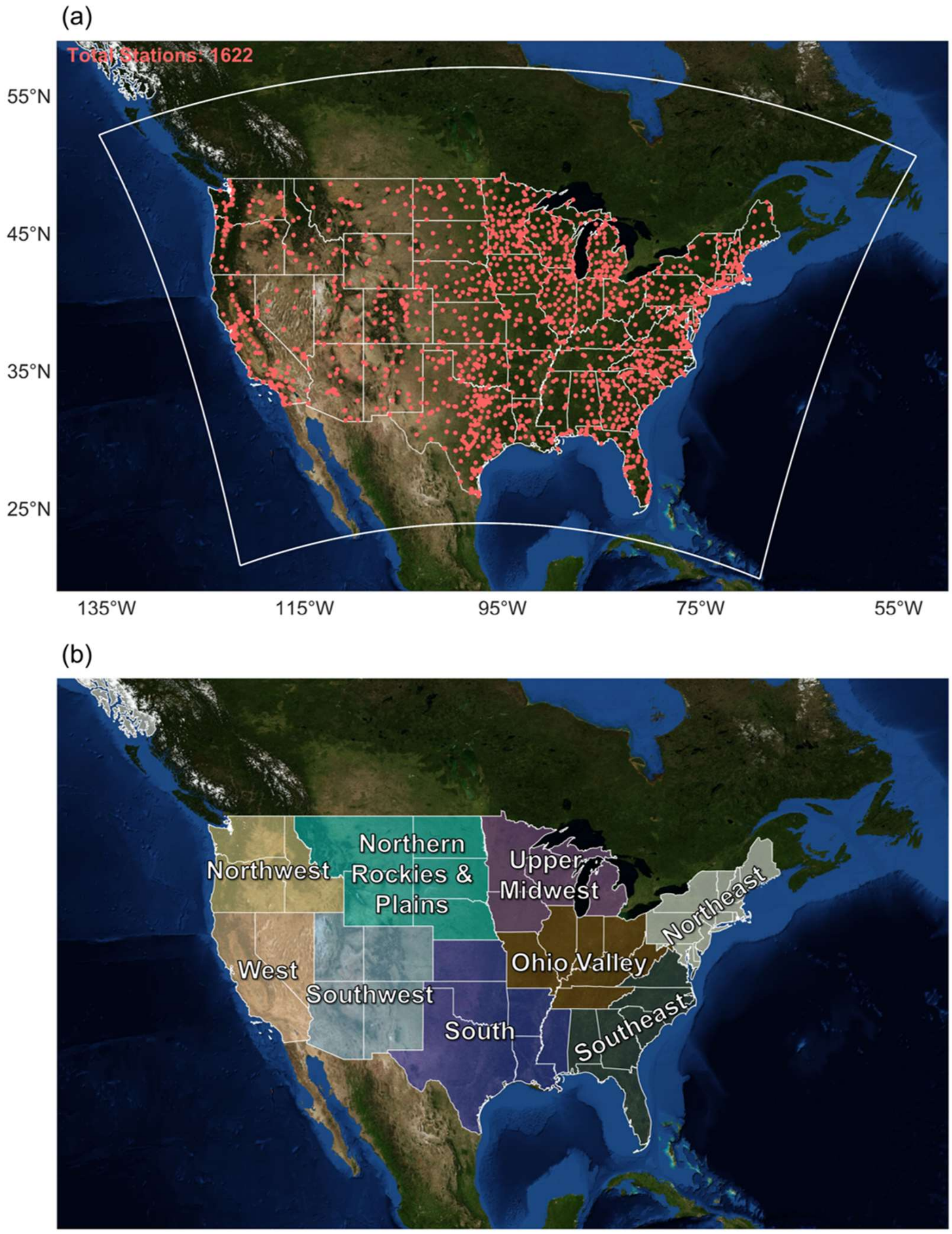

Figure 2: Spatial coverage of surface observational datasets, WRF model domain, and regional divisions: (a) locations of NCDC surface meteorological stations (red dots) across CONUS domain; the white box outlines the WRF simulation domain; (b) delineation of nine climatologically coherent subregions. Base map imagery from NASA Worldview, Earth Observing System Data and Information System (EOSDIS).

It was unclear if 69% at 5 significant digits in the input data (WRF_5) meant that:

1) original data at full precision are further compressed using lossless gzip

2) 5-digit truncated input data are further compressed using lossless gzip

3) the ratio of the storage size 2) over 1) is computed.

Thank you for the comment. Your interpretation is entirely correct. The "relative compression ratio" is explicitly designed to isolate the additional space savings attributable strictly to decimal significant-digit rounding. It is formally calculated as the storage size of the precision-reduced data (subsequently compressed via a lossless codec) divided by the storage size of the full-precision data (compressed using the identical lossless codec). To avoid any ambiguity, we have provided an explicit and formal definition of the relative compression ratio in the revised manuscript (Section 3.1: Lines 304–306).

Lines 304–306:

"Here, the relative compression ratio is defined as the compressed size of the precision-reduced configurations divided by the compressed size of the full-precision baseline. This metric isolates the additional space savings attributable specifically to decimal significant-digit rounding."

When the authors write that the baseline dataset has 837GB of input data, are these full precision data or data compressed using gzip / bzip2?

Thank you for the question. The reported 837 GB baseline explicitly refers to the original, completely uncompressed WRF input forcings stored in the standard 32-bit single-precision floating-point format (with the output also adhering to the same standard 32-bit single-precision floating-point format). To permanently eliminate any ambiguity, we have provided a clearer expression in the revised manuscript (Section 3.1: Lines 300–302).

During this revision process, we proactively identified and rectified two reporting discrepancies in the original manuscript. First, the uncompressed baseline output volume was previously misstated as 2154.3 GB; the correct physical full-precision size is 7395.8 GB. Second, in the original Figure 2a, the input relative compression ratios were inadvertently calculated against the uncompressed raw data, rather than the lossless compressed baseline. This visualization error has been fully rectified in the updated Figure 3a to strictly align with the formal definition of relative compression. We emphasize that these corrections do not alter any scientific conclusions.

Lines 300–302:

"All calculations are benchmarked against a full-precision baseline (stored in the standard 32-bit single-precision floating-point format) comprising 837.0 GB of uncompressed input forcings and 7395.8 GB of uncompressed output data."

A table summarizing the effective storage space after various truncation strategies, as well as compression, would be very useful.

Thank you for the suggestion. As recommended, we have added a Table 5 (see below) and related description in the revised manuscript (Abstract: Lines 18–20; Section 3.1: Lines 303–304, 311–312; Conclusion: Lines 710–716).

Lines 18–20:

"From a storage perspective, combining precision reduction (retaining 5–3 digits) with bzip2 compression reduces model outputs to 19.2%–7.5% of their original uncompressed sizes and model inputs to 52.4%–18.5%."

Lines 303–304:

"As summarized in Table 5 and illustrated in Fig. 3, data compressibility improves progressively with more aggressive precision reduction."

Lines 311–312:

"On average, bzip2 outperforms gzip by 15–30 percentage points across all precision-reduced configurations (Table 5)."

Table 5. Storage and compression performance of input and output data across diverse precision-reduced configurations.

Data Type	Configurations	gzip			bzip2		
		Compressed size (GB)	% of Orig. Size	Relative compression ratio (%)	Compressed size (GB)	% of Orig. Size	Relative compression ratio (%)
INPUT (Original size: 837 GB)	WRF_bl	628.4	75.1	-	621.5	74.3	-
	WRF_5	577.4	69.0	91.9	438.5	52.4	70.6
	WRF_4	457.5	54.7	72.8	274.5	32.8	44.2
	WRF_3	291.7	34.9	46.4	155.1	18.5	25.0
OUTPUT (Original size: 7395.8 GB)	WRF_bl	2229.5	30.1	-	2212.6	29.9	-
	WRF_5	2214.0	29.9	99.3	2196.6	29.7	99.3
	WRF_4	2223.3	30.0	99.7	2203.8	29.8	99.6
	WRF_3	2217.0	30.0	99.4	2228.8	30.1	100.7
	WRF_fx5	1879.4	25.4	84.3	1418.5	19.2	64.1
	WRF_fx4	1489.8	20.2	66.8	926.0	12.5	41.9
	WRF_fx3	977.2	13.2	43.8	551.6	7.5	24.9
	WRF_5fx5	1879.6	25.4	84.3	1418.8	19.2	64.1
	WRF_5fx4	1489.9	20.2	66.8	926.1	12.5	41.9
	WRF_5fx3	977.3	13.2	43.8	551.4	7.5	24.9
	WRF_4fx5	1876.1	25.4	84.2	1414.8	19.2	64.1
	WRF_4fx4	1489.5	20.1	66.8	925.8	12.5	41.8
	WRF_4fx3	972.1	13.1	43.6	545.8	7.4	24.7
	WRF_3fx5	1872.2	25.3	84.0	1433.2	19.4	64.8
	WRF_3fx4	1486.5	20.1	66.7	937.6	12.7	42.4
	WRF_3fx3	978.2	13.2	43.9	560.5	7.6	25.3

*Notes: Relative compression ratio = (compressed size of precision-reduced configurations/compressed size of WRF_bl).

Lines 710–716:

"From a storage perspective, removing redundant numerical precision prior to compression yields substantial additional storage savings beyond conventional lossless compression. For model outputs, retaining 5 to 3 significant digits prior to bzip2 compression reduces data volumes to 19.2%–7.5% of their original uncompressed sizes (corresponding to 64.1%–24.9% relative to lossless compression alone). Comparable reductions relative to the original uncompressed dataset are achieved with Zstd (21.8%–10.2%), zlib (25.6%–13.0%), and gzip (25.4%–13.2%). Similarly, precision-reduced model inputs paired with bzip2 compress files to 52.4%–18.5% of their original volumes (corresponding to 70.6%–25.0% relative to lossless compression alone), while Zstd, zlib and gzip achieve 58.3%–25.5%, 69.1%–33.6% and 69.0%–34.9%, respectively."

The color scheme of Fig. 6 is confusing and does not correspond to previous figures.

Thank you for this comment. We fully agree that the categorical color mapping in the original Figure 6 lacked visual consistency with the preceding analyses. To resolve this, we have completely redesigned this visualization (now presented as Figure 9 in the revised manuscript). The updated figure strictly employs a standardized matrix of colors and marker shapes that systematically maps to specific climate regions and seasons, ensuring seamless visual continuity across the entire paper. Furthermore, during this graphical revision, we recognized that our original palette was not scientifically optimized for red-green color vision deficiencies. Consequently, we have systematically overhauled the color schemes across multiple visualizations in the manuscript (e.g., the newly revised Figure 5).

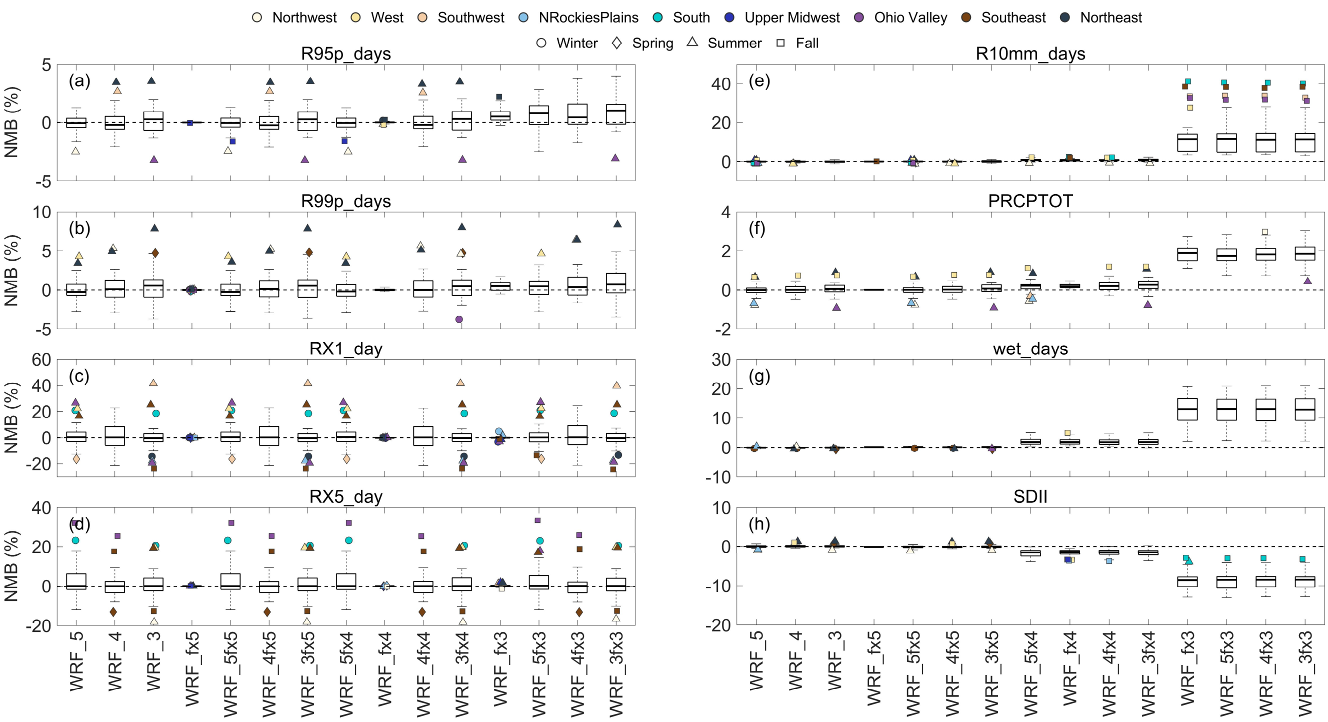

Figure 9: Box plots showing the distribution of NMB values for extreme precipitation indices across climate regions and seasons. Each box spans the interquartile range (IQR; 25th–75th percentiles), with

the median shown as a horizontal line. Whiskers extend to 1.5 × IQR, and outliers are plotted as colored shapes corresponding to different climate regions and seasons. Dashed horizontal lines indicate zero bias.

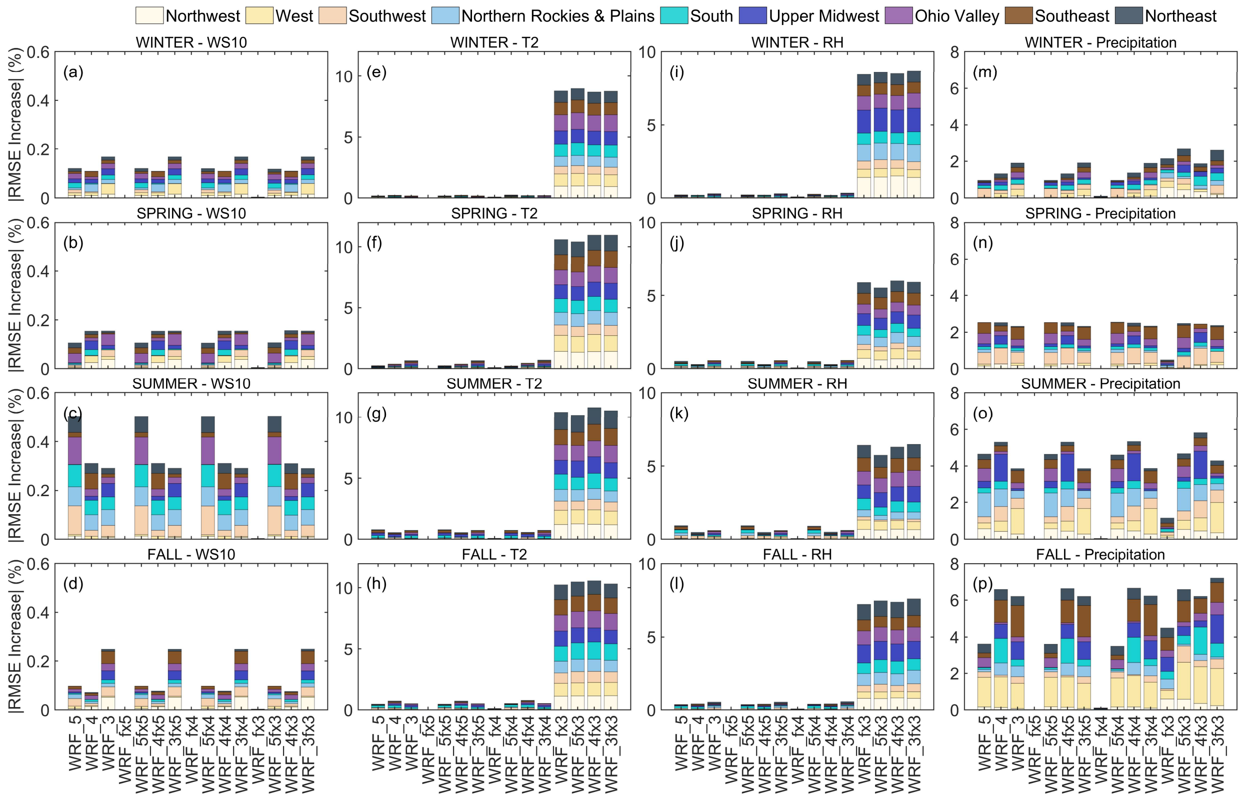

Figure 5: Magnitude of seasonal and regional relative changes in RMSE for meteorological variables across 15 precision-reduced configurations. Stacked bars illustrate the regional breakdown of |RMSE increase| relative to the WRF_bl across nine climate regions. Note that the total bar heights serve solely to visualize relative regional contributions and do not represent a mathematically aggregated domain-wide RMSE percentage. Results are presented separately for (a–d) wind speed, (e–h) temperature, (i–l) relative humidity, and (m–p) precipitation.

What is also missing is a visual showing one measurement (e.g., temperature at 2m) with the FORTRAN representation and corresponding truncations.

Thank you for this suggestion. In the revised manuscript, we have added a table (Table 2) and corresponding description in Section 2.2 (Lines 182–184). The table presents variables (T2, Q2, U10, and V10) at a single grid point and time step, showing the original values together with the corresponding values after retaining 5, 4, and 3 significant digits.

Table 2. An example of retaining different significant digits for WRF output variables.

Variables	Full precision	5-digit	4-digit	3-digit
T2 (K)	290.3516	290.35	290.4	290
Q2 (kg kg^{-1})	0.01852282	0.018523	0.01852	0.0185
U10 (m s^{-1})	0.5504633	0.55046	0.5505	0.551
V10 (m s^{-1})	-1.483089	-1.4831	-1.483	-1.48

Lines 182–184:

"To provide an intuitive visualization of the rounding results, Table 2 contrasts the original full-precision values against those retained at 5, 4, and 3 significant digits for key near-surface variables at a single grid point and time step: 2-m temperature (T2), 2-m specific humidity (Q2), and 10-m wind components (U10 and V10)."

References

Baker, A. H., Hammerling, D. M., and Turton, T. L.: Evaluating image quality measures to assess the impact of lossy data compression applied to climate simulation data, Computer Graphics Forum, 38, 517–528, https://doi.org/10.1111/cgf.13707, 2019.

Baker, A. H., Hammerling, D. M., Mickelson, S. A., Xu, H., Stolpe, M. B., Naveau, P., Sanderson, B., Ebert-Uphoff, I., Samarasinghe, S., and De Simone, F.: Evaluating lossy data compression on climate simulation data within a large ensemble, Geoscientific Model Development, 9, 4381–4403, https://doi.org/10.5194/gmd-9-4381-2016, 2016.

Davenport, F. V., Burke, M., and Diffenbaugh, N. S.: Contribution of historical precipitation change to US flood damages, Proceedings of the National Academy of Sciences, 118, e2017524118, https://doi.org/10.1073/pnas.2017524118, 2021.

Delaunay, X., Courtois, A., and Gouillon, F.: Evaluation of lossless and lossy algorithms for the compression of scientific datasets in netCDF-4 or HDF5 files, Geoscientific Model Development, 12, 4099-4113, https://doi.org/10.5194/gmd-12-4099-2019, 2019.

Han, T., Chen, Z., Guo, S., Xu, W., and Bai, L.: CRA5: Extreme Compression of ERA5 for Portable Global Climate and Weather Research via an Efficient Variational Transformer, arXiv:2405.03376, https://doi.org/10.48550/arXiv.2405.03376, 2024.

Huang, L. and Hoefler, T.: Compressing multidimensional weather and climate data into neural networks, arXiv:2210.12538, https://doi.org/10.48550/arXiv.2210.12538, 2023.

Klöwer, M., Razinger, M., Dominguez, J. J., Düben, P. D., and Palmer, T. N.: Compressing atmospheric data into its real information content, Nature Computational Science, 1, 713-724, https://doi.org/10.1038/s43588-021-00156-2, 2021.

Mirowski, P., Warde-Farley, D., Rosca, M., Grimes, M. K., Hasson, Y., Kim, H., Rey, M., Osindero, S., Ravuri, S., and Mohamed, S.: Neural Compression of Atmospheric States, arXiv:2407.11666, https://doi.org/10.48550/arXiv.2407.11666, 2024.

Poppick, A., Nardi, J., Feldman, N., Baker, A. H., Pinard, A., and Hammerling, D. M.: A statistical analysis of lossily compressed climate model data, Computers & Geosciences, 145, 104599, https://doi.org/10.1016/j.cageo.2020.104599, 2020.

Roebber, P. J.: Visualizing multiple measures of forecast quality, Weather and Forecasting, 24, 601-608, https://doi.org/10.1175/2008WAF2222159.1, 2009.

Seneviratne, S. I., Zhang, X., Adnan, M., Badi, W., Dereczynski, C., Di Luca, A., Ghosh, S., Iskandar, I., Kossin, J., Lewis, S., Otto, F., Pinto, I., Satoh, M., Vicente-Serrano, S. M., Wehner, M., and Zhou, B.: Weather and Climate Extreme Events in a Changing Climate, in: Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change, edited by: Masson-Delmotte, V., Zhai, P., Pirani, A., Connors, S. L., Péan, C., Berger, S., Caud, N., Chen, Y., Goldfarb, L., Gomis, M. I., Huang, M., Leitzell, K., Lonnoy, E., Matthews, J. B. R., Maycock, T. K., Waterfield, T., Yelekçi, O., Yu, R., and Zhou, B., Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 1513–1766,

https://doi.org/10.1017/9781009157896.013, 2021.
Silver, J. D. and Zender, C. S.: The compression–error trade-off for large gridded data sets, Geoscientific Model Development, 10, 413–423, https://doi.org/10.5194/gmd-10-413-2017, 2017.
Underwood, R., Bessac, J., Di, S., and Cappello, F.: Understanding the effects of modern compressors on the community earth science model, 2022 IEEE/ACM 8th International Workshop on Data Analysis and Reduction for Big Scientific Data (DRBSD), https://doi.org/10.1109/DRBSD56682.2022.00006, 2022.
Walters, M. S. and Wong, D. C.: The impact of altering emission data precision on compression efficiency and accuracy of simulations of the community multiscale air quality model, Geoscientific Model Development, 16, 1179–1190, https://doi.org/10.5194/gmd-16-1179-2023, 2023.
Wang, Z., Bovik, A. C., Sheikh, H. R., and Simoncelli, E. P.: Image quality assessment: from error visibility to structural similarity, IEEE Transactions on Image Processing, 13, 600–612, https://doi.org/10.1109/TIP.2003.819861, 2004.
Zender, C. S.: Bit Grooming: statistically accurate precision-preserving quantization with compression, evaluated in the netCDF Operators (NCO, v4.4.8+), Geoscientific Model Development, 9, 3199–3211, https://doi.org/10.5194/gmd-9-3199-2016, 2016.

Response to the comments of Reviewer 2

Thank you for reviewing our manuscript thoroughly and providing constructive comments and valuable suggestions. The reviewer's comments are in black *Italics* followed by our responses (in blue). The red text within the quotation marks is the revisions in the manuscript, while the black text is the unmodified content.

The authors present a study on the impacts of lossy compression in WRF simulations over the contiguous United States. The precision of input or output fields is truncated (rounding) and errors are analyzed across different variables and derived diagnostics. The authors conclude with a strategy who to truncate WRF data in their case.

The paper addresses lossy compression the prevailing solution to growing data archives in weather and climate modelling. Given that lossy compression is either not applied or analyses hardly reach the level conducted here I also see this as a timely contribution as it suggests safe levels of compression errors. Novel is the analyses of errors arising from rounding input data, most compression studies deal with output data and therefore ignore the effect of rounding errors on a simulation. In that sense I am generally happy to accept this study but only after major changes: The methods aren't well reported, some subjective choices are unjustified and much of the literature in this field from recent years is not cited and their conclusions therefore in relation to results here not discussed. I find much around the choice of error metrics problematic and insufficiently motivated and discussed particularly when they are used to define an "optimal truncation strategy".

I start with some major points, note that some of the minor points may repeat issues as I wrote the minor points first. Feel free to cross-reference your answers.

Thank you for your careful reading of the manuscript and for recognizing both the relevance and novelty of this work. We are encouraged that you consider this analysis timely and potentially useful for identifying safe levels of compression. In accordance with your suggestions, we have carefully and substantially revised the manuscript to address the concerns you raised. Please find our point-by-point responses listed below.

1) Precision truncation method

It is not stated how the precision of floating-point numbers is actually truncated. Generally, this is known as rounding (given you apply the truncation to numbers and not e.g. spectra) so I suggest you to adapt the term "rounding". There are many rounding modes available, so please state the details here. I highly suggest you to apply IEEE 754 standard round to nearest tie to even, it's an IEEE standard for a reason. Other rounding modes are bit shaving, setting or grooming (as terminology used by Zender et al.) however they have a towards or away from zero and don't deal with ties properly. You never talk about bits so it's unclear to me whether you actually round in *binary (zeroing trailing mantissa bits) or whether you round in decimal like round(x*10^N)/10^N. I can't find this in the provided Fortran code (no readme provided). If you don't round in binary then you truncate the precision without actually setting bits to 0, e.g. 0.1 has a 001100110011... mantissa in binary. The lossless compressors will still be able to compress this somewhat but you're essentially giving the compressors a much harder time than if it was rounded in binary. Please state your methods, also the cited Walters and Wong (2023) don't seem to explain this in detail. Also how to deal with ties is important, IEEE 754 introduced alternating rounding of ties up and down (to the nearest even number) for the reason to avoid an away from zero bias. Given you talk about biases in your error analyses, the role of ties is unclear a priori.*

Thank you for your thorough and insightful review.

According to your suggestion, we have revised the relevant terminology in the manuscript to "decimal significant-digit rounding" or "significant-digit rounding". The term "precision truncation" used in the previous version has been replaced with "precision reduction" or "decimal significant-digit rounding," and expressions such as "truncation strategies" have been replaced by "precision reduction configurations." Concerning the significant-digit rounding method, we have added a clear statement in the revised manuscript that we employ a decimal significant-digit rounding approach which was developed in the Walters and Wong (2023) paper, and added a new Table 2 in Section 2.2 to demonstrate the idea.

The decimal significant-digit rounding tool is written in Fortran. The basic idea of the decimal significant-digit rounding tool contains three steps: (1) formulate a specific Fortran scientific format statement with respect to a given n, where n is the number of significant digits you want to keep; (2) write a real number to a character string based on that specific format statement created in Step 1, and; (3) read a new real number from the character string which was created in Step 2.

Here are the actual lines of code:

```
character (len = 30) :: fmt, str
write (fmt, '(a2, I2.2, a1, I2.2, a1)') '(e', n+7, '.', n, ')')
write (str, fmt) real_number
read (str, *) new_real_number
```

Here is a sample toy code for testing purposes:

```
program test
implicit none
integer :: ndigits
real :: old, new
character (len = 30) :: fmt, str
write (6, *) ' enter a real number and the number of significant digits
               you want to keep'
read *, old, ndigits
```

```fortran
    if ((ndigits > 8) .or. (ndigits < 0)) then
        write (6, *) ' Abort: Invalid n_digits setup (proper range [ 0, 8 ]'
        write (6, *)
        stop
    end if

    write( fmt, '(a2, i2.2, a1, i2.2, a1)') '(e', ndigits+7, '.', ndigits, ')'
    write (str, fmt) old
    read (str,*) new

    write (6, '(a14, 2e18.8)') ' ==d== result ', old, new

end program test
```

Please note that rounding (traditional rounding technique, values ≥ 5 round up, and values < 5 round down) will be performed in Step#2 by Fortran internally. In mathematical scientific notation, $a \times 10^n$, where $1 \leq a \leq 10$ but in Fortran it is $0 \leq a \leq 1$. As a result, there are 7 reserved spaces (indicated in the above line of code) in the Fortran scientific notation to hold, the sign of the value, the notation of "0.", and exponent "E±mn". Follow your suggestion, we have added the details of rounding method in the revised manuscript (Section 2.1: Lines 153–158) and added a detailed README file in public released code.

We agree with you that conducting more operations at the binary level can yield data patterns with greater compressibility, as evidenced in Delaunay et al. (2019). However, we refrain from making any specific claim regarding the general superiority of binary-level over decimal-level compression effectiveness dealing with WRF related data without providing a comprehensive evaluation, which is beyond the scope of this work, of these two underlying methodologies. Providing a comprehensive evaluation of these underlying binary mechanisms remains beyond the scope of the present study. The primary motivation for our adoption of the decimal-based significant digit method lies in its intuitive nature and ease of implementation. This approach allows users to intuitively grasp the mechanism of precision reduction, thereby lowering the barrier to adoption within the community and facilitating the integration of compression strategies into the practical workflows of a broad range of model users. In the revised Section 2.2 (Lines 162–167), we have incorporated a detailed methodological discussion to clarify the rationale for adopting the decimal rounding method and to highlight its inherent limitations in compression efficiency compared to further manipulation on the binary level.

Thank you for highlighting the importance of using the IEEE 754 round-to-nearest, ties-to-even rounding rule to avoid potential systematic bias associated with exact tie cases. We agree that this rounding mode is theoretically preferable and represents best practice for unbiased rounding. To assess the practical impact of this issue in the context of our study, we conducted an additional statistical evaluation to quantify how frequently exact tie cases occur in the variables analyzed. Across the 16 examined variables (e.g., T2, Q2, U10, V10, and precipitation), each containing approximately 1.4 billion data points, we evaluated the proportion of values for which conventional decimal rounding and ties-to-even rounding would produce different results. The results indicate that such cases are rare. Depending on the number of retained significant digits, the proportion of affected values is approximately:

~0.003–0.005% when keeping 3 significant digits,
~0.03–0.04% when keeping 4 significant digits, and
~0.2–0.4% when keeping 5 significant digits.

This means that more than 99.6–99.99% of values are rounded identically under both rounding rules, indicating that the practical impact of rounding-mode selection on the datasets analyzed in this study is negligible. Given this very small difference and the fact that the current implementation relies on the default rounding behavior of the Fortran environment used in our study, we retained the existing implementation. However, following your suggestion, we have now added a quantitative assessment and discussion of this issue in the revised manuscript to clarify its potential implications. The corresponding explanation has been added to (Section 2.2: Lines 168–174).

<u>Lines 153–158:</u>

"This scheme regulates numerical precision by retaining a user-specified number of significant digits. The rounding procedure was carried out using a dedicated Fortran routine based on internal character formatting: each floating-point value is first converted into a string in scientific notation using a dynamically specified Fortran format descriptor (i.e., Ew.d, where E denotes the exponential scientific notation, the total width w = n + 7, and the number of decimal places d = n; here, n denotes the targeted number of significant digits to retain), and the formatted string is subsequently read back into floating-point representation."

<u>Lines 162–167:</u>

"The deliberate adoption of this decimal-based approach, instead of direct binary mantissa manipulation, is motivated by the following considerations: decimal significant-digit rounding enables users to intuitively understand and verify the exact numerical modifications applied to the data, thereby lowering the adoption barrier for practical modeling workflows. Nevertheless, we acknowledge the inherent compression advantages of bit-level operations. As demonstrated by Delaunay et al. (2019), algorithms that map decimal precision requirements to the binary mantissa can maximize sequences of consecutive zero bits, thus generating data patterns with higher compressibility."

<u>Lines 168–174:</u>

"Furthermore, traditional decimal rounding (rounding half away from zero), which is natively handled by Fortran, may theoretically introduce systematic biases compared to IEEE 754 round-to-nearest, ties-to-even rules (IEEE, 1985). We conducted a statistical evaluation and found that exact tie cases are rare when we practically apply the method of rounding to significant digits. Across all examined core variables (e.g. surface meteorological field and precipitation, each comprising over 1.4 billion data points), exact tie situations occurred at low frequencies: approximately 0.003–0.005% for retaining 3 significant digits, 0.03–0.04% for 4 significant digits, and 0.2–0.4% for 5 significant digits. Because over 99.6% of values are rounded identically under both approaches, the practical impact of rounding-mode selection on the datasets analyzed herein is negligible."

Table 2. An example of retaining different significant digits for WRF output variables.

Variables	Full precision	5-digit	4-digit	3-digit
T2 (K)	290.3516	290.35	290.4	290
Q2 (kg kg^{-1})	0.01852282	0.018523	0.01852	0.0185
U10 (m s^{-1})	0.5504633	0.55046	0.5505	0.551
V10 (m s^{-1})	-1.483089	-1.4831	-1.483	-1.48

Related is that you don't state whether you're dealing with single or double precision numbers. If you're doing binary rounding then the compressor can take care of the additional 4 byte of zeros but if you round in in another base that may matter and also is relevant to the stated size of your baseline dataset being ~3TB.

Thank you for your comment. It was our oversight that we did not state that we were dealing with single-precision netCDF data. The reported baseline dataset size of ~3 TB is computed on the basis of this single-precision storage format. We have clarified in the revised Section 2.2 (Lines 151–153) and Section 3.1 (Lines 300–302).

<u>Lines 151–153:</u>

"The input and output data of the WRF model are stored in single-precision (float32) format, which supports a numerical precision of approximately 7 significant digits. To control the precision of such single-precision netCDF datasets, this study implements a decimal significant-digit rounding method following Walters and Wong (2023)."

<u>Lines 300–302:</u>

"All calculations are benchmarked against a full-precision baseline (stored in the standard 32-bit single-precision floating-point format) comprising 837.0 GB of uncompressed input forcings and 7395.8 GB of uncompressed output data."

2) Lossless codecs

There are many other lossless codecs available yet you focus on bzip2 and gzip both are about 30 years old. Newer ones are Zstandard, Blosc, LZ4, or pcodec. Why don't you compare with those? There might be a reason why you want to use those older codecs but if they have an advantage over newer one, please state them? Also related is that bzip2 just seems to be the better choice in your case but you don't actually report on (de)compression speed (gzip might be faster). I suggest to compare against at least one compressor of the newer generation, e.g. Zstandard. Each of them also has compression options to trade in speed for compression that need to be stated. With Zstandard for example you can choose a really high compression level, which produces smaller compressed files but takes forever. So generally, you should always state this tradeoff and the limits you apply, e.g. you don't want a compressor to be slower than 100MB/s or so.

Thank you for this helpful suggestion. Following this recommendation, we have expanded our compression benchmark analysis to include Zstandard (Zstd), which is widely regarded as a modern high-performance codec and is representative of the newer generation of compressors designed to balance compression efficiency and runtime performance.

In the revised manuscript, we now compare gzip, bzip2, and Zstd across multiple precision-reduction configurations. To ensure a realistic comparison for WRF workflows, we conducted targeted benchmark experiments on both input forcing datasets and model output files. Specifically, we selected seven 5-day segments of WRF input data (starting on January 1, 6, 11, 16, 21, 26, and 31) and applied compression to four datasets: the original full-precision data and datasets with 5, 4, and 3 retained significant digits. The same methodology was applied to seven daily WRF output files (January 1–7) to evaluate compression performance for model outputs. Because the performance of Zstd depends strongly on compression parameters, we evaluated multiple configurations. Zstd typically supports compression levels ranging from 1 to 19, which allow users to trade computational cost for improved compression ratios. In our benchmarks, we tested compression levels 3, 7, and 19 using a single CPU core, as well as level 19 using 8 threads to evaluate the benefits of parallel execution.

The benchmark results reveal a clear trade-off between compression efficiency and computational cost. For raw full-precision datasets, Zstd achieves compression ratios comparable to or slightly better than bzip2 while operating substantially faster when multi-threading is enabled. However, once decimal significant-digit rounding is applied, the relative performance changes: bzip2 consistently produces smaller compressed files than Zstd across all tested configurations.

This behavior likely reflects differences in the underlying compression algorithms. Zstd is based on LZ77-style dictionary matching, which is highly effective when repeated or closely matching byte sequences occur. In precision-reduced floating-point datasets, rounding introduces structured regularity mainly in the lower-order mantissa bits, while higher-order bits retain physically meaningful variability. Such patterns may reduce the frequency of exact byte-level matches, limiting the effectiveness of dictionary-based compression. In contrast, bzip2 uses the Burrows–Wheeler Transform, a block-sorting approach that reorders data to cluster similar patterns prior to entropy encoding and appears to better exploit the structured redundancy introduced by decimal significant-digit rounding.

Based on these results, we use bzip2 as the analytical baseline for estimating maximum storage savings, and explicitly recommend it for long-term cold archiving where absolute space storage savings supersedes compression/decompression speed. Furthermore, as a ubiquitous standard utility natively available across virtually all high-performance computing (HPC) environments, bzip2 ensures seamless data portability and zero-dependency deployment. Conversely, Zstd is recommended for practical, active workflows. Zstd provides a favorable balance between compression ratio and processing speed, particularly when multi-threading is enabled, making it well suited for active research environments where datasets are frequently accessed.

The new benchmark experiments and discussions have been incorporated into the revised manuscript in Section 2.2 (Lines 199–205), Section 3.1 (Lines 323–364), and Conclusions (Lines 725–727) with the detailed compression ratios and configuration settings reported in Tables 6 and 7.

There is currently no discussion on data formats in the manuscript. While the rounding can be applied to any format as it just operates on floating-point numbers in some arrays the use of different lossless codecs may be integrated into the format (e.g. netCDF with zlib compression, compresses fields but not the header so you don't have to decompress to read the metadata). So, for the tool you built, which data formats does it operate on?

The decimal significant-digit rounding method developed in this study is fundamentally format-agnostic, as it operates directly on floating-point data arrays. In principle, it can be applied to any data format that stores numerical fields in standard floating-point representation (e.g., netCDF, HDF5, or binary formats). However, in the current implementation, the tool is configured to operate on netCDF datasets, which are the standard input and output format of WRF simulations.

Within the netCDF framework, the rounding procedure is applied exclusively to floating-point data variables, while the file structure, metadata, and header information remain unchanged (Section 2.2: Lines 158–161). As a result, the method preserves full compatibility with format-native compression schemes, including netCDF internal zlib compression. In this case, compression is applied only to the data variables, whereas metadata remains uncompressed and directly accessible, allowing efficient metadata inspection without triggering data decompression.

We have expanded the manuscript to explicitly discuss the interaction between precision reduction and format-integrated compression (Section 3.1: Lines 315–356). We now include a comparison between netCDF internal compression (zlib) and external file-level compressors (bzip2, Zstd, gzip), highlighting their different operational characteristics (Conclusions: Lines 714–724). While netCDF internal compression provides advantages in metadata accessibility and seamless integration, external compressors (e.g. bzip2 and Zstd) offer improved compression performance for precision-reduced datasets. These revisions clarify that our approach is not tied to a specific data format, but can be flexibly integrated into existing data workflows, with the choice of format and compression strategy guided by practical use requirements.

Table 6. Compression performance of a 5-day WRF input file for the full-precision data and data retaining 5, 4, and 3 significant digits, respectively. Numbers in parentheses indicate specific configurations: (compression level, number of threads) for Zstd, and (deflation level) for zlib.

	Compressor	Full Precision	5 digits	4 digits	3 digits
Size (% of uncompressed)	bzip2	75.11	52.98	33.11	18.56
	gzip	75.61	69.33	55.04	34.95
	zlib (1)	75.08	68.18	55.78	38.59
	zlib (9)	74.80	69.10	54.42	33.63
	Zstd (3,1)	76.87	73.19	58.08	36.70
	Zstd (7,1)	76.10	65.29	50.95	31.16
	Zstd (19,1)	74.16	58.32	43.15	25.54
	Zstd (19,8)	74.14	58.27	43.08	25.47

Compression (sec)	bzip2	821.83	713.64	635.17	656.60
	gzip	811.32	609.17	760.91	893.29
	zlib (1)	301.00	247.00	199.00	143.00
	zlib (9)	571.00	516.00	743.00	1289.00
	Zstd (3,1)	34.48	64.47	67.52	64.35
	Zstd (7,1)	169.25	293.75	266.06	189.56
	Zstd (19,1)	3544.12	2905.94	3301.46	3345.15
	Zstd (19,8)	401.65	502.20	536.14	521.97
Decompression (sec)	bzip2	468.92	404.08	308.98	239.09
	gzip	76.91	77.76	76.80	63.09
	zlib (1)	0.00	0.00	0.00	0.00
	zlib (9)	0.00	0.00	0.00	0.00
	Zstd (3,1)	10.96	13.39	14.56	13.30
	Zstd (7,1)	12.17	15.76	15.21	12.82
	Zstd (19,1)	16.11	23.98	20.14	13.09
	Zstd (19,8)	16.71	23.89	20.43	13.23

Table 7. Compression performance of a 1-day WRF output file for the full-precision data and data retaining 5, 4, and 3 significant digits, respectively. Numbers in parentheses indicate specific configurations: (compression level, number of threads) for Zstd, and (deflation level) for zlib.

	Compressor	Full Precision	5 digits	4 digits	3 digits
Size (% of uncompressed)	bzip2	30.46	19.68	12.88	7.65
	gzip	30.66	25.93	20.66	13.58
	zlib (1)	30.98	26.13	21.30	15.08
	zlib (9)	30.34	25.63	20.18	12.96
	Zstd (3,1)	30.41	26.60	21.26	14.34
	Zstd (7,1)	30.18	24.27	18.95	12.42
	Zstd (19,1)	29.79	21.82	16.12	10.17
	Zstd (19,8)	29.78	21.81	16.11	10.16
Compression (sec)	bzip2	940.15	855.28	839.07	879.25
	gzip	410.83	569.10	578.58	576.38
	zlib (1)	253.00	207.00	176.00	142.00
	zlib (9)	614.00	749.00	905.00	1125.00
	Zstd (3,1)	38.90	54.08	62.68	45.06
	Zstd (7,1)	116.27	186.48	188.44	130.43
	Zstd (19,1)	1664.30	1711.50	1963.40	2517.14
	Zstd (19,8)	326.24	337.98	352.54	325.12
Decompression (sec)	bzip2	361.47	306.41	251.31	206.12
	gzip	99.43	101.65	99.04	90.94
	zlib (1)	0.00	0.00	0.00	0.00
	zlib (9)	0.00	0.00	0.00	0.00
	Zstd (3,1)	10.99	12.19	12.63	11.79
	Zstd (7,1)	11.18	13.51	12.69	11.24
	Zstd (19,1)	12.67	17.66	15.44	11.28
	Zstd (19,8)	12.66	17.60	15.03	11.34

Lines 158–161:

"The precision-reduction method implemented in this study operates directly on the floating-point variables, except latitude and longitude, in the netCDF data part, which are the standard format used by WRF for both input forcing and model outputs. The rounding procedure modifies only the numerical values stored in the variable arrays and does not affect the netCDF header, metadata, or structural attributes."

Lines 199–205:

"In this study, gzip and bzip2 were selected as the primary external compressors for evaluation, given their widespread availability in operational high-performance computing (HPC) environments and compatibility with existing modeling workflows. To account for format-integrated compression, we also evaluated the internal zlib compression of netCDF, which is commonly used within the netCDF/HDF5 framework and enables direct metadata access without full data decompression. In addition, to align with contemporary data storage practices, we included targeted evaluations using the newer-generation codec Zstandard (Zstd). A comparative assessment of compression performance across these compressors is presented in Section 3.1."

Lines 315–356:

"We additionally evaluated format-native netCDF internal compression (utilizing zlib deflation) alongside external compressors, including gzip, bzip2, and the newer-generation, multi-threaded Zstd codec. Because our precision-reduction procedure strictly modifies only the floating-point data arrays while preserving the original file structure, the resulting datasets remain fully compatible with the internal compression functionality of netCDF. This native integration provides an important operational advantage: because netCDF internal compression applies strictly to data variables, the metadata headers remain uncompressed and directly accessible. This enables rapid metadata inspection and lazy loading without triggering decompression of the underlying data variables, resulting in effectively zero decompression time in our benchmarks (reported as 0.00 s). Benchmark experiments were conducted for both input and output datasets across different precision-reduction configurations, and the compression performance for a representative 5-day WRF input dataset and a 1-day WRF output dataset is summarized in Tables 6 and 7, respectively. Throughout the following discussion, specific compressor configurations are denoted by parentheses indicating their settings; for instance, Zstd (19,8) denotes Zstandard applied at compression level 19 using 8 threads, and zlib (9) refers to zlib deflation at compression level 9.

The benchmark results reveal a clear trade-off between storage efficiency and computational cost, and this trade-off differs between full-precision and precision-reduced datasets. For the raw full-precision data, the best storage reduction is achieved by high-level Zstd, which slightly outperforms both bzip2 and zlib. For example, for the WRF output file (Table 7), Zstd at level 19 reduces the file to 29.78 % of its original size, compared with 30.34 % for zlib (9), 30.46 % for bzip2, and 30.66 % for gzip. A similar but smaller advantage is found for the full-precision input file (Table 6), where Zstd (19,8) reduces the file to 74.14 % of its original size, compared with 74.80 % for zlib (9) and 75.11 % for bzip2. In addition, low-level Zstd provides markedly faster compression than bzip2 and gzip, while multi-threaded Zstd (19,8) recovers much of the runtime cost associated with high compression levels.

Once significant-digit rounding is applied, however, the relative ranking changes substantially. Across all tested precision-reduced configurations, bzip2 consistently achieves the smallest final file sizes, while Zstd remains the strongest practical alternative, and zlib and gzip occupies an intermediate position. For example, for the WRF input file with 3 significant digits retained (Table 6), bzip2 reduces the dataset to 18.56 % of its original size, compared with 25.47 % for Zstd (19,8), 33.63 % for zlib (9), and 34.95 % for gzip. For the WRF output file with 3 significant digits retained (Table 7), the corresponding values are

7.65 % for bzip2, 10.16 % for Zstd (19,8), 12.96 % for zlib (9), and 13.58 % for gzip. Thus, after precision reduction, the practical ordering is generally bzip2 > Zstd > zlib ≈ gzip in terms of compression ratio. Notably, zlib level 1 provides a balanced native option, but maximum native compression at zlib level 9 becomes increasingly expensive as precision is reduced, with compression times rising sharply for heavily rounded datasets (e.g., 1289 s for the input with 3 significant digits retained and 1125 s for the output with 3 significant digits retained) without proportional storage benefits.

These differences reflect both the underlying compression mechanisms and their algorithmic implementations. Within the LZ77-based family (Zstd, gzip, zlib), Zstd employs more efficient entropy coding (Finite State Entropy, FSE), which explains its superior performance on raw full-precision data (Collet and Kucherawy, 2021). However, decimal rounding transforms long repeated byte sequences into quasi-repetitive, imperfect patterns primarily reflected in the lower-order mantissa bits, reducing the effectiveness of dictionary-based matching. For zlib, this limitation is further amplified by its operation within HDF5 chunk boundaries, which limits the exploitation of redundancy beyond local chunks, and by deeper dictionary hash-chain traversals at high compression levels. As a result, the algorithm incurs substantial computational overhead while searching for longer matching sequences among near-matches, without proportional gains in compression efficiency. In contrast, compressors such as bzip2 process data in larger blocks and employ the Burrows–Wheeler Transform (Burrows and Wheeler, 1994). By globally reordering data prior to entropy encoding, BWT effectively clusters scattered, non-contiguous bit patterns introduced by precision reduction, enabling more efficient exploitation of redundancy than LZ77-based methods. Consequently, bzip2 consistently achieves the highest compression ratios for these rounded datasets."

Lines 714–724:

"Similarly, precision-reduced model inputs paired with bzip2 compress files to 52.4%–18.5% of their original volumes (corresponding to 70.6%–25.0% relative to lossless compression alone), while Zstd, zlib and gzip achieve 58.3%–25.5%, 69.1%–33.6% and 69.0%–34.9%, respectively. The overall effectiveness of precision reduction is intrinsically coupled to the choice of the backend lossless compression codec. Our findings demonstrate that for data processed via decimal significant-digit rounding, a distinct operational trade-off exists between storage efficiency and compression/decompression execution time. For long-term archival storage, where maximum compression ratio is the primary objective, bzip2 remains the preferred choice despite its higher computational cost. In contrast, for active research workflows involving frequent data access and post-processing, Zstd provides a more balanced solution, offering near-optimal compression efficiency together with substantially faster compression and decompression speeds. The internal compression zlib of netCDF represents a complementary option, particularly in environments where direct metadata accessibility and seamless format integration are required."

3) Subjective choice on the significant digits

You round input and or output data to 3, 4, or 5 significant digits. However, this range seems to come out of nowhere. Why not 6 significant digits or 2? For some variables 3 digits might be an overkill if the uncertainty is only within 2x of the values. For others like CO2 5 significant digits might be at the edge of that's acceptable as it's a well-mixed concentration with the variance being relatively small compared to its mean value. For a more systematic analysis on this see Kloewer et al. 2021 who also advocate for the round+lossless compression method but employ IEEE rounding, information theory to determine the number of bits to keep and use newer lossless *codecs that generally achieve 10-20x compression. You achieve at most 4x compression (Fig 2), which isn't better than ECMWF's/ERA5s linear packing (also called quantization) into 16-bit integers (but bounds an absolute rather than a relative error).*

Thank you for this insightful comment. We agree that the choice of retained significant digits should ideally be guided by the intrinsic information content of each variable and the scientific requirements of downstream analyses.

In this study, the range of 3–5 retained significant digits was chosen primarily as a practical evaluation window rather than a universal recommendation. WRF model inputs and outputs are stored in single-precision (float32) format, which provides approximately at most 7 significant digits of precision. In our initial design, we considered that retaining 6 significant digits would provide very limited compression benefit while remaining close to full precision, making it less relevant for evaluating meaningful storage reductions. Conversely, retaining only 1–2 significant digits would clearly destroy the scientific meaning of key variables, such as temperature and pressure, and therefore was not considered meaningful for diagnostic evaluation. Hence, we selected 3–5 digits as a balanced experimental range that spans realistic levels of precision reduction while still preserving interpretable meteorological signals. We have added additional discussion to clarify this point in the revised Section 2.2 (Lines 175–181).

We agree that the acceptable precision level can vary substantially across variables. For example, some variables, such as surface wind speed, exhibit a relatively narrow dynamic range and may tolerate stronger precision reduction. In our analysis, retaining 3 significant digits has demonstrated negligible numerical impact on wind fields. Furthermore, we now explicitly discuss in the revised manuscript that a 2-digit configuration could theoretically serve as a viable alternative when wind speeds remain below 10 m s^{-1}, as it mathematically preserves a single-decimal resolution. We have added additional discussion to clarify this point (Section 4.2: Lines 647–657). In contrast, some variables require higher retained precision. A representative example is precipitation, which is stored as the cumulative sum of grid-resolved (RAINNC) and convective (RAINC) components. Over long simulations, accumulated precipitation values may reach several thousand millimeters within one year and potentially exceed ten thousand millimeters in multi-year integrations. In such cases, insufficient retained digits would produce excessively coarse quantization steps when hourly precipitation is obtained through temporal differencing. For this reason, cumulative variables such as precipitation requires 4-5 or more significant digits depending on simulation duration. We have added additional discussion in the revised manuscript to clarify this point (e.g., Section 4.2: Lines 676–682).

We fully agree with the reviewer that the information-theoretic framework proposed by Klöwer et al. (2021) provides a more systematic way to determine the optimal precision for specific variables. Their work demonstrates that the intrinsic information content varies significantly across atmospheric fields. This work is incorporated as a supplementary reference to address the limitations of our study. In the revised manuscript we now discuss this perspective and emphasize that this systematic variable-specific precision reduction method represents an important direction for future work (Section 4.2: Lines 664–666).

We would like to clarify the interpretation of the reported compression ratios. The four-fold compression mentioned in the original manuscript refers to the additional compression achieved relative to lossless compression alone, rather than compression relative to the original uncompressed datasets. Combining precision reduction (retaining 5–3 digits) with bzip2 compression reduces model outputs to 19.2%–7.5% of their original uncompressed sizes and model inputs to 52.4%–18.5%. To avoid confusion, we have revised the manuscript to explicitly distinguish between direct compression ratios (relative to the original data) and additional compression gains introduced by precision reduction (Section 3.1: Lines 297–300, Lines 312–314; Conclusions: Lines 710–714).

Lines 175–181:

"Given that single-precision format inherently provides approximately at most 7 significant digits, reducing precision to 6 digits yields only marginal compression gains. Conversely, retaining only 1 or 2 significant digits would severely degrade key variables such as temperature and pressure, rendering such configurations physically unrealistic. As a result, this study conservatively focuses on retaining 3–5 significant digits. The range of 3–5 retained significant digits is not intended to represent a universal optimal precision level for each variable but rather to provide a practical experimental window spanning meaningful levels of precision reduction for single-precision WRF data. More systematic, variable-specific optimal precision levels could be determined using information-theoretic approaches (Klöwer et al., 2021)."

Lines 297–300:

"To quantify the impact of decimal significant-digit rounding on data compressibility, we analyze compression performance using two complementary metrics. First, we evaluate the additional compression gain relative to lossless compression alone, which isolates the contribution of rounding. Second, we assess the overall storage reduction relative to the original uncompressed datasets to provide an intuitive measure of practical data savings."

Lines 312–314:

"From a practical storage perspective, when retaining 3 significant digits and using bzip2, the size of the compressed input and output datasets is approximately one-quarter of that achieved with lossless compression alone. Specifically, the compressed datasets account for 18.5% and 7.5% of the original uncompressed data volume, respectively."

Lines 647–657:

"For fundamental state variables, compressibility is primarily governed by their numerical distributions and physical measurement limits. Atmospheric state variables exhibit markedly different compressibility characteristics determined by their numerical distributions. Dynamic variables such as WS10 are typically concentrated within a narrow numerical range, retaining 3 significant digits completely guarantees the preservation of at least one decimal place (e.g., a resolution of 0.1 m s^{-1}). Therefore, applying a universal 3-digit retention for wind fields represents the most conservative compromise between storage reduction and precision, introducing negligible numerical distortion in our evaluation. Theoretically, retaining 2 significant digits could also be a viable configuration. Because 2 digits mathematically preserve single-decimal resolution for values below 10 m s^{-1}, perceptible quantization errors would primarily be expected to emerge only when wind speeds exceed this threshold. Consequently, to maximize storage benefits without compromising physical fidelity, future applications could implement a magnitude-aware adaptive strategy for wind fields: dynamically toggling between 2 and 3 retained significant digits based on a 10 m s^{-1} threshold."

Lines 664–666:

"This pronounced heterogeneity in compressibility aligns profoundly with Klöwer et al. (2021), which demonstrates that a variable-specific precision reduction strategy emerges as a key paradigm for optimizing output data compression."

Lines 676–682:

"Specifically, for grid cells where the accumulated precipitation remains below 1000 mm, retaining 4 significant digits is generally sufficient to preserve a scientifically viable sub-millimeter resolution. Once the accumulation surpasses the 1000 mm threshold, the precision should be dynamically increased to 5 significant digits to explicitly prevent the decimal resolution from degrading. For multi-year simulations, where total accumulations routinely exceed 10000 mm, retaining 6 significant digits becomes strictly necessary. It is worth noting that designing a lower-tier threshold (e.g., at 100 mm) is operationally unnecessary; most simulated regions rapidly exceed this baseline shortly after initialization, rendering any lower-precision tier computationally transient and practically redundant."

Lines 710–714:

"From a storage perspective, removing redundant numerical precision prior to compression yields substantial additional storage savings beyond conventional lossless compression. For model outputs, retaining 5 to 3 significant digits prior to bzip2 compression reduces data volumes to 19.2%–7.5% of their original uncompressed sizes (corresponding to 64.1%–24.9% relative to lossless compression alone). Comparable reductions relative to the original uncompressed dataset are achieved with Zstd (21.8%–10.2%), zlib (25.6%–13.0%), and gzip (25.4%–13.2%)."

4) Missing literature

The word "compression" occurs in only 3 independent (not from the authors) studies cited. So, you are clearly not discussing well your findings against the existing literature. Under References below I'm listing a few studies that are relevant for this study here. It's not just about citing them but actually writing a manuscript (especially introduction and discussion) that builds on top of their findings. Please rewrite your manuscript to account for the results of these studies.

Thank you for pointing out that the original manuscript did not sufficiently situate our study within the broader literature on atmospheric data compression. Following this suggestion, we conducted a substantially expanded literature review and revised the Introduction, Methodology and Results sections to better position our work relative to previous studies.

First, we enriched the research background in the Introduction (Lines 54–80) by incorporating additional studies on precision reduction and compression methods. The revised manuscript now discusses classical precision-reduction techniques and related evaluation frameworks, including Zender (2016) on bit grooming and bit shaving, Delaunay et al. (2019) on the Digit Rounding algorithm bridging decimal and binary representation, Silver and Zender (2017) on compression–error trade-offs in geophysical datasets, Baker et al. (2019) on structural fidelity diagnostics, technical investigations of compression performance in netCDF/HDF datasets (Delaunay et al., 2019; Underwood et al., 2022), and addressing the problem of atmospheric data compression from an information-theoretic perspective

(Klöwer et al., 2021). We also clarified the role of the IEEE-754 floating-point standard (IEEE, 1985) in governing numerical precision and rounding behavior. These revisions provide a more complete context for understanding how precision-reduction approaches have been developed and evaluated in previous studies.

Second, building on this expanded research context, we have explicitly integrated the perspective of Klöwer et al. (2021) into the dedicated discussion on variable-dependent precision requirements (Section 2.2: Lines 180–181; Section 4.2: Lines 664–666). This study is prominently cited in the revised manuscript, as its findings provide critical theoretical and methodological complements to our work: theoretically, it reveals substantial variations in the intrinsic precision requirements across different physical variables; methodologically, it offers support for the systematic determination of precision requirements for individual variables.

Finally, inspired by the evaluation frameworks used in Baker et al. (2019) and Klöwer et al. (2021), we strengthened the methodological analysis of structural fidelity in our study. Specifically, we introduced the Structural Similarity Index Measure (SSIM) as a complementary diagnostic metric. This metric is described in revised Section 2.4 and applied in the newly added Section 3.3 "Maximum Deviations and Structural Fidelity." For detailed information, please refer to the response to the next major comment.

Lines 54–80:

"A variety of compression techniques applicable to atmospheric model archives have been developed to alleviate the rapidly growing storage demands of numerical simulations. Lossless algorithms alone preserve bitwise reproducibility but generally achieve only a modest compression ratio for floating-point geophysical fields because the high entropy of mantissa bits limits compressibility (Poppick et al., 2020). To address this limitation, combining precision reduction with lossless compression has emerged as a widely explored strategy for improving storage efficiency while attempting to preserve scientific fidelity. Within this paradigm, many approaches focus on manipulating the least significant bits of floating-point data. Early approaches such as bit shaving (Zender, 2016) reduce precision by zeroing trailing mantissa bits, which can introduce systematic bias. Zender (2016) subsequently introduced bit grooming, a statistically robust quantization approach that alternates bit shaving and bit setting of trailing IEEE mantissa bits, thereby preserving the mean in expectation while improving compressibility. Building on this line of work, Delaunay et al. (2019) introduced the Digit Rounding algorithm, which bridges decimal precision control and binary representation. Compared to bit grooming, Digit Rounding optimizes the allocation of mantissa bits required, thereby improving compression efficiency while retaining controlled numerical precision. In contrast to hybrid methods like Digit Rounding, alternative approaches regulate precision purely within the decimal domain. For example, significant-digit rounding (Walters and Wong, 2023) directly constrains the number of significant digits, yielding a highly transparent and interpretable approach to precision reduction.

Alongside algorithmic developments, some studies have expanded the evaluation of precision reduction beyond simple error statistics toward broader notions of statistical and structural fidelity. Baker et al. (2016) proposed an ensemble-based framework to assess whether compressed datasets remain statistically indistinguishable from internal climate variability, while Baker et al. (2019) emphasized the importance of evaluating structural integrity in compressed climate data. Silver and Zender (2017) quantified the compression–error trade-off for gridded datasets, demonstrating that carefully designed precision reduction can remove substantial false precision with limited impact on scientific conclusions. Complementary technical investigations have examined how data representation and codec behavior influence achievable compression ratios within netCDF and HDF data formats (Delaunay et al., 2019; Underwood et al., 2022). Recently, Klöwer et al. (2021) reframed atmospheric data compression from an information-theoretic perspective, demonstrating that the intrinsic precision requirements of atmospheric variables vary substantially and that a large fraction of stored floating-point precision is redundant. In parallel, machine-learning-based compression methods have emerged that exploit nonlinear data manifolds to achieve very high compression ratios (Huang and Hoefler, 2023; Han et al., 2024; Mirowski et al., 2024)."

<u>Lines 180–181:</u>

"Establishing strict, variable-specific theoretical precision limits would require information-theoretic evaluations (Klöwer et al., 2021)."

<u>Lines 664–666:</u>

"This pronounced heterogeneity in compressibility aligns profoundly with Klöwer et al. (2021), which demonstrates that a variable-specific precision reduction strategy emerges as a key paradigm for optimizing output data compression."

5) Error metrics

You use 3 error metrics: an absolute error (RMSE), correlation and a bias. They seem to be subjectively chosen and their use isn't motivated. What about a relative error or an error in variance, a maximum error (yielding a much stronger bound), number of zeros preserved (important for precipitation). What about other suggested metrics like the structural similarity index measure (SSIM), see Baker et al 2019 or a spectral error (used to identify the introduction/removal of grid-scale variability). These error metrics are also important to discuss relative to the distribution of variables. E.g. temperature is more linearly distributed (with a higher entropy on a linear scale compared to a logarithmic scale) but other variables may be logarithmically distributed (wind speed, precipitation, global specific humidity). This affects the meaning of absolute vs relative error. Either error metric may be dominated by compression errors on outliers. Please include this into the discussion of your results.

Thank you for this very helpful and insightful comment. Following this suggestion, we substantially expanded the evaluation framework in the revised manuscript by incorporating additional diagnostics designed to quantify maximum absolute deviations (AD), structural fidelity, and the sensitivity of precipitation's lower-tail distribution.

First, we introduced Maximum AD and the Structural Similarity Index Measure (SSIM) as complementary diagnostics (Section 2.4: Lines 256–273). These analyses are presented in the newly added Section 3.3 (“Maximum Deviations and Structural Fidelity”) (Lines 434–516), with the inclusion of new Table 8, Figs. 6–7, and Supplementary Figs. S1–S4.

The results reveal extreme Maximum AD values under input perturbations, such as T2 reaching ~21.45 K and hourly precipitation hitting ~145.45 mm h^{-1} (Table 8). However, despite these apparently large local deviations, the annual minimum SSIM for T2 remains above 0.97 under WRF_4, and at the exact moment when T2 deviation reaches its peak, the SSIM is 0.998. The annual minimum SSIM for precipitation is approximately 0.86 under WRF_3, but at the exact moment when precipitation deviation reaches its peak, the SSIM approaches 0.99. To further clarify this behavior, five variables (T2, Q2, WS10, PSFC, and hourly precipitation) were examined to determine the relationship between the maximum AD at individual grid points and the SSIM computed at the corresponding time when these maximum deviations occur (Fig. 6). The results reveal a clear decoupling between grid-point maximum deviations and structural fidelity under input precision-reduced configurations. This indicates that large local errors under input precision reduction primarily represent spatial or temporal phase shifts of coherent meteorological features, rather than a fundamental systemic breakdown of the simulated atmospheric structures. Snapshot comparisons at the time of maximum AD for T2 and precipitation (Fig. S1 and Fig. S2) provide further visual evidence supporting this interpretation.

Furthermore, evaluating the temporal evolution of SSIM reveals two fundamentally different error modes associated with input and output precision reduction (Fig. 7). Input precision reduction produces a seasonally modulated “U-shaped” SSIM pattern, reflecting the amplification of small perturbations by nonlinear atmospheric dynamics during periods of strong convective activity. In contrast, output-only precision reduction produces a step-like and monotonic decline in SSIM, consistent with purely numerical rounding artifacts that do not interact with model dynamics.

Importantly, precipitation accumulation acts as a constantly increasing baseline, retaining a fixed number of significant digits for forces the absolute quantization error to grow over time. By late autumn, this expanding quantization interval completely overwhelms the physical signal of weak stratiform precipitation, leading to a catastrophic structural collapse (SSIM dropping to ~0.88). Based on this finding, in Discussion (Section 4: Lines 657–682) of the revised manuscript, we propose a dynamic precision reduction strategy for cumulative variable (e.g., precipitation). Specifically, for grid-scale accumulated precipitation values less than 1000 mm, 4 significant figures should be retained; for values greater than 1000 mm, the number of significant figures should be increased to 5.

Second, to systematically quantify the specific vulnerability associated with the lower-tail boundary of the precipitation distribution (i.e., zero-value preservation), we introduced categorical metrics including the Zero Preservation Ratio (ZPR), Probability of Detection (POD), False Alarm Ratio (FAR), Critical Success Index (CSI), and Frequency Bias (Bias). The definitions of these methods are detailed in the revised Section 2.4 (Lines 274–283) and revised Table 4. The analytical content is presented in the revised Section 3.4, “Impacts on Precipitation Diagnostics: Zero-Value Preservation and Extreme Precipitation Indices” (Lines 511–555). The results are summarized in the new added Table 9, which quantifies how different precision-reduced configurations affect the preservation of precipitation occurrence using the 0.1 mm h^{-1} wet-day threshold.

The results show that input-only precision reduction largely preserves the fundamental wet/dry morphology, maintaining a Frequency Bias near unity with symmetrically balanced error rates between false alarms and missed events. This indicates that input perturbations we introduced primarily induce spatial phase shifts in precipitation systems rather than fundamentally altering precipitation occurrence statistics. In contrast, aggressive output precision reduction (e.g., retaining only three significant digits) introduces substantial distortions at the lower tail of the precipitation distribution. Under the WRF_fx3 configuration, the POD drops to 71.94%, indicating that nearly 28% of valid light precipitation events are artificially truncated to zero due to the loss of numerical significance in cumulative precipitation arithmetic. Under the WRF_fx4 configuration, as the proportion of grids with cumulative precipitation reaching 1000 mm further increases (Fig. S3), a notable late-stage decline in metrics is observed (Fig. S4), which aligns with the SSIM identified in Section 3.3.

Ultimately, the integration of these multidimensional diagnostics transitions our analysis from bulk statistical approximations to a rigorously comprehensive evaluation. It now explicitly distinguishes the fundamentally different impacts of precision reduction on input forcings (which induce physical spatiotemporal phase shifts) versus static model outputs (which generate numerical artifacts, especially for cumulative variables).

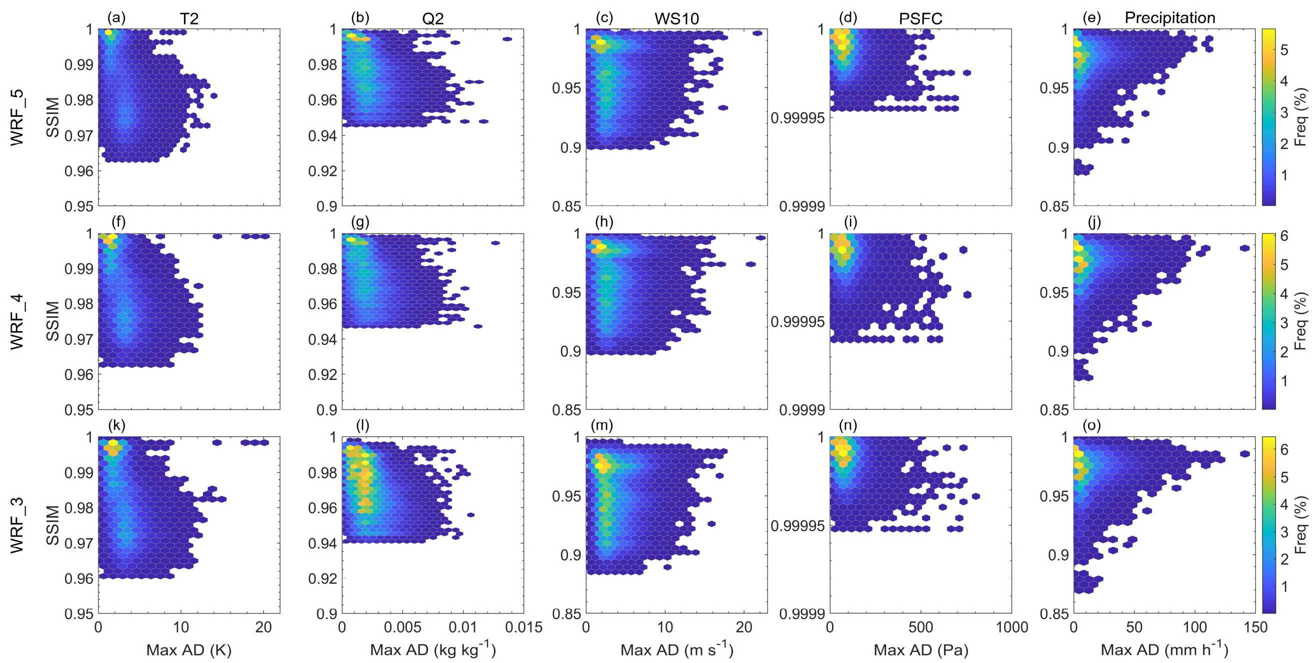

Figure 6: Hexbin density plots showing the grid-scale maximum AD and the corresponding SSIM. Results are presented for five variables (T2, Q2, WS10, PSFC, and precipitation) with input precision reduced to 5, 4, and 3 significant digits, respectively.

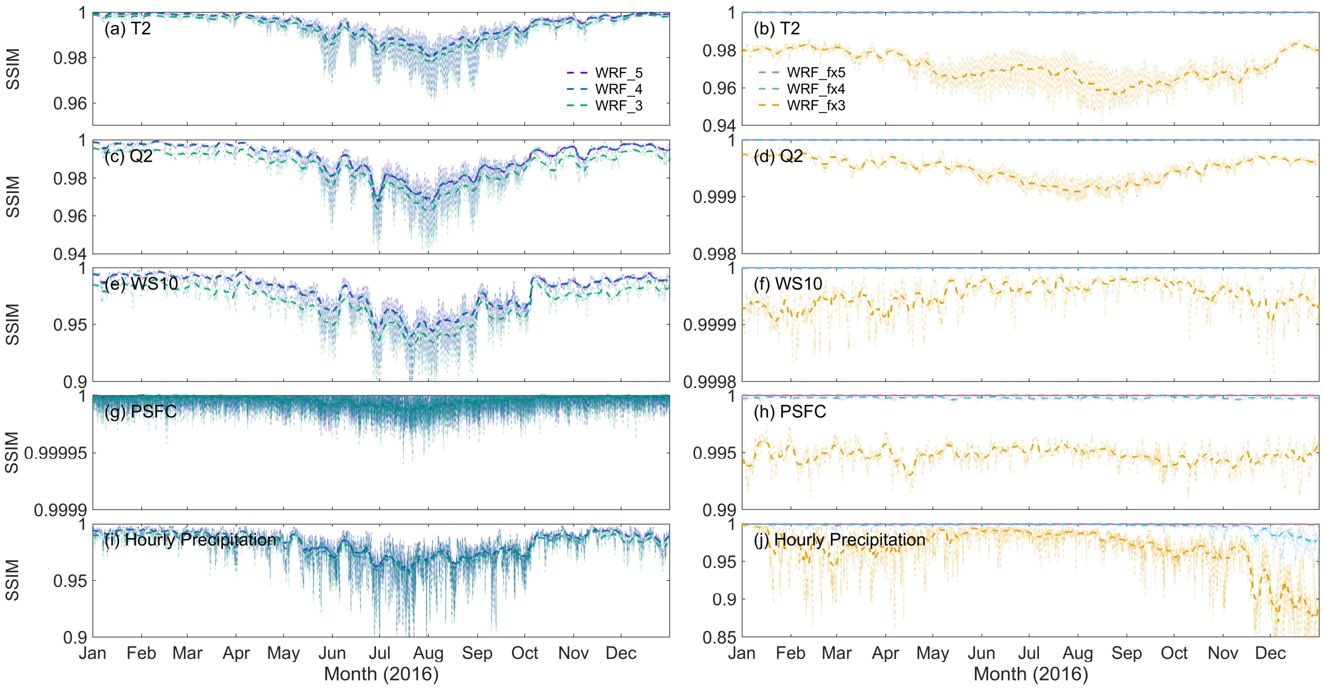

Figure 7: Time series of SSIM for five variables (T2, Q2, WS10, PSFC, and precipitation) under different precision-reduced configurations. The fine dotted lines represent the original value of hourly SSIM, and the thick dashed lines represent the rolling median of hourly SSIM.

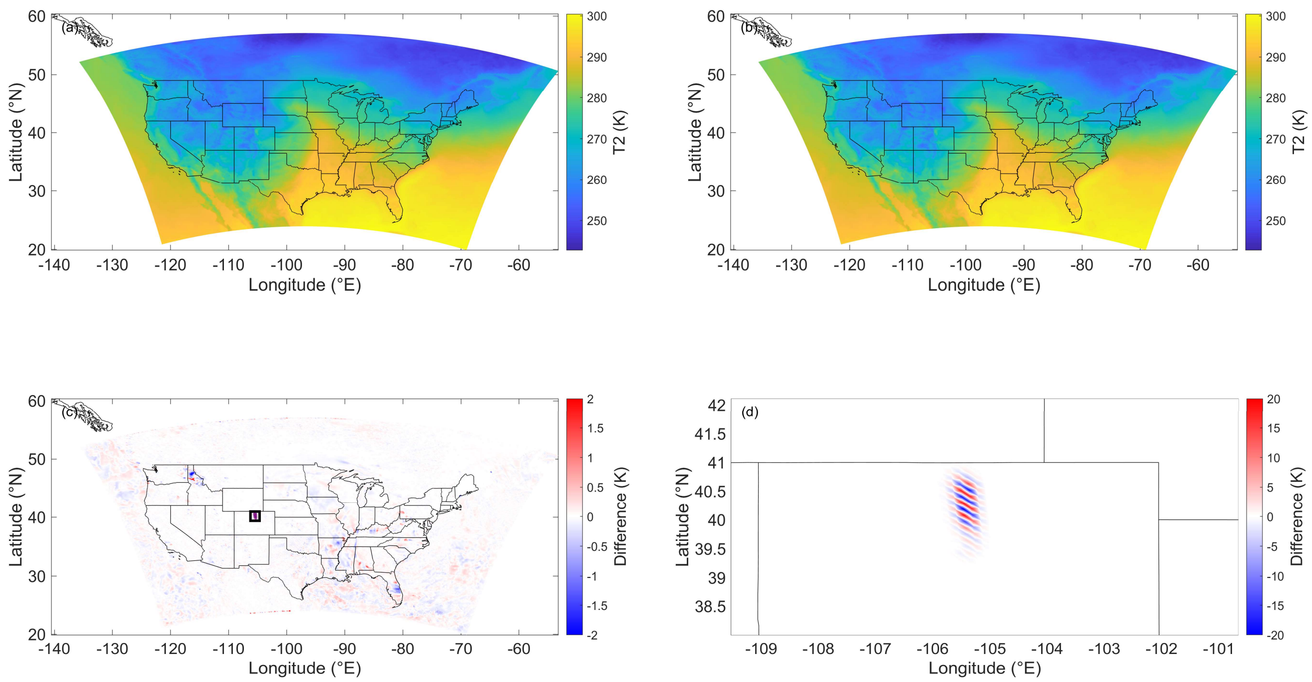

Figure S1: Spatial visualization of T2 at the time step exhibiting the maximum AD. (a) WRF_bl, (b) WRF_3, (c) the spatial difference fields (WRF_3 minus WRF_bl), and (d) a locally zoomed-in view of the difference field corresponding to the bold black box in (c). The black square marks the precise location of the maximum AD (21.12 K). At this specific time step, the domain-wide SSIM is 0.998.

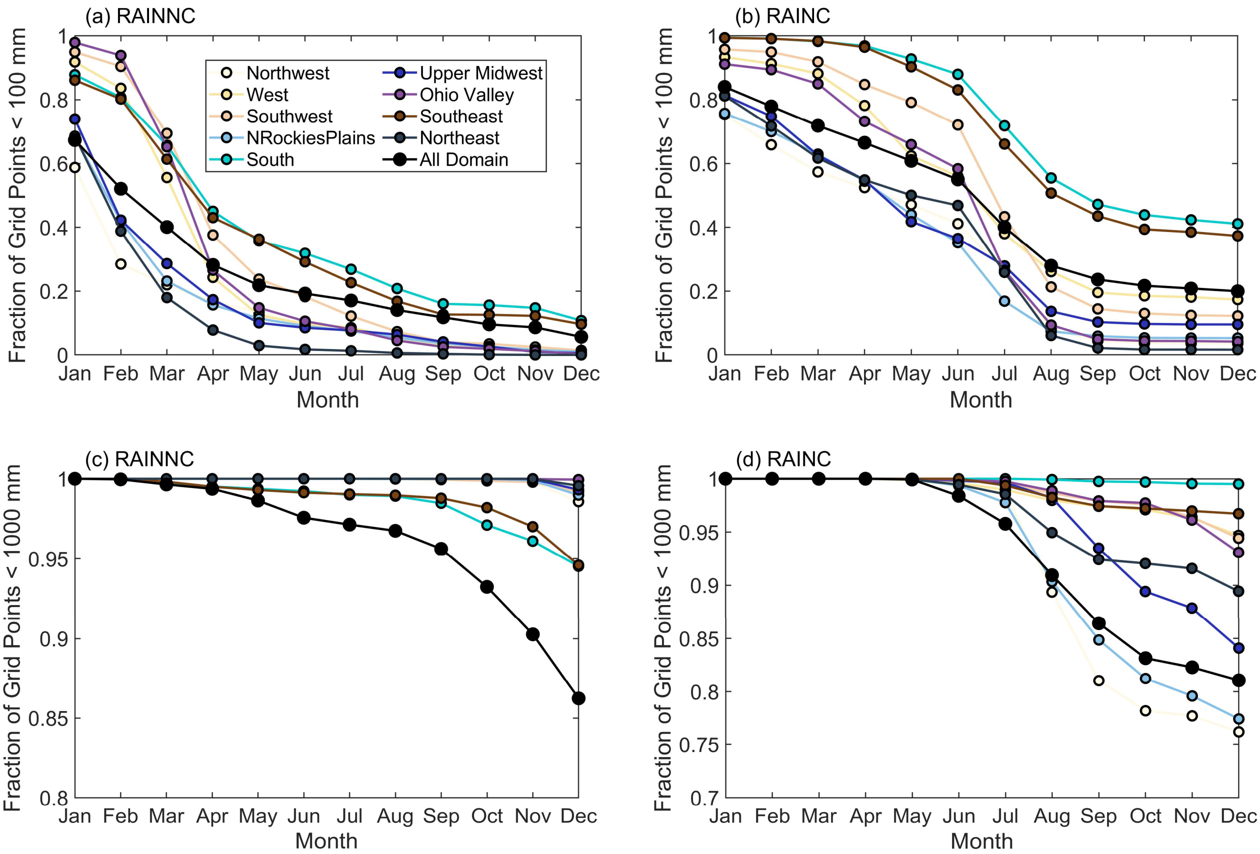

Figure S3: Monthly evolution of the fraction of grid points with accumulated non-convective (RAINNC; a, c) and convective (RAINC; b, d) precipitation remaining below the 100 mm (top row) and 1000 mm (bottom row) thresholds. The lines denote different climate regions and the entire WRF simulation domain average.

Table 8. Statistical summaries of grid-scale AD and domain-scale SSIM relative to WRF_bl.

Variable	Precision Configuration	Mean AD	Maximum AD	Mean SSIM	Minimum SSIM	1st Percentile SSIM
T2 (K)	WRF_5	0.0581798	14.25198	0.9939092	0.9630990	0.9709463
	WRF_4	0.0594887	21.44626	0.9938635	0.9626499	0.9705832
	WRF_3	0.0894812	21.11664	0.9921307	0.9605753	0.9682225
	WRF_fx5	0.0025001	0.0050000	0.9999945	0.9995806	0.9998766
	WRF_fx4	0.0250000	0.0500000	0.9997513	0.9981659	0.9991125
	WRF_fx3	0.2500606	0.5000000	0.9713302	0.9423122	0.9506720
Q2 (kg kg^{-1})	WRF_5	0.0000477	0.0134958	0.9898861	0.9457591	0.9574129
	WRF_4	0.0000486	0.0125344	0.9897875	0.9471009	0.9578402
	WRF_3	0.0000677	0.0120278	0.9853344	0.9410679	0.9523994
	WRF_fx5	0.0000001	0.0000005	0.9999999	0.9999989	0.9999996
	WRF_fx4	0.0000010	0.0000050	0.9999960	0.9999886	0.9999907
	WRF_fx3	0.0000105	0.0000500	0.9994877	0.9988874	0.9989862
WS10 (m s^{-1})	WRF_5	0.0717549	22.79240	0.9779398	0.8997450	0.9211134
	WRF_4	0.0726728	21.88551	0.9777849	0.8985054	0.9208920
	WRF_3	0.1003181	18.21674	0.9678928	0.8853601	0.9100209
	WRF_fx5	0.0000324	0.0007067	0.9999999	0.9999995	0.9999998

	WRF_fx4	0.0003241	0.0070658	0.9999995	0.9999979	0.9999986
	WRF_fx3	0.0032415	0.0705795	0.9999494	0.9998048	0.9998650
PSFC (Pa)	WRF_5	1.6698395	811.9844	0.9999944	0.9999550	0.9999727
	WRF_4	1.6841504	878.2109	0.9999943	0.9999399	0.9999720
	WRF_3	2.3376082	804.9219	0.9999937	0.9999480	0.9999702
	WRF_fx5	1.2597312	5.0000	0.9999974	0.9998826	0.9999572
	WRF_fx4	12.595335	50.0000	0.9997887	0.9991961	0.9994622
	WRF_fx3	127.72333	500.0000	0.9947414	0.9909650	0.9919420
Precipitation ($mm\ h^{-1}$)	WRF_5	0.0443069	130.6014	0.9816086	0.8783776	0.9302602
	WRF_4	0.0444704	139.8082	0.9815308	0.8769866	0.9299442
	WRF_3	0.0495302	145.4452	0.9782254	0.8696560	0.9271690
	WRF_fx5	0.0014635	0.195557	0.9998999	0.9946446	0.9983964
	WRF_fx4	0.0105326	1.972900	0.9963600	0.9353753	0.9608235
	WRF_fx3	0.5544670	19.90356	0.9632875	0.7775496	0.8290028

Table 4. Definitions of categorical verification metrics for hourly precipitation and daily extreme precipitation indices.

Name	Definition	Units
ZPR	Percentage of dry grids (< 0.1 $mm\ h^{-1}$) in WRF_bl that remain below the threshold after precision reduction.	%
POD	Percentage of wet grids (≥ 0.1 $mm\ h^{-1}$) in WRF_bl that are correctly retained after precision reduction.	%
FAR	Percentage of wet grids after precision reduction that are false alarms relative to WRF_bl.	%
CSI	A comprehensive metric for the spatial fidelity of the precipitation field, penalizing both the artificial elimination (missed events) and generation (false events) of wet events.	-
Bias	Ratio of the total number of wet grids in the precision-reduced configurations to that in WRF_bl. Values > 1 indicate an artificial inflation of precipitation spatial extent.	-
R95p_days	Number of days per year with daily precipitation exceeding the 95th percentile of wet-day amounts (≥ 1 mm), thresholds derived from the 2001–2015 baseline period.	days
R99p_days	Same as R95p_days, but for the 99th percentile threshold.	days
Rx1_day	Maximum 1-day precipitation total in a year.	mm
Rx5_day	Maximum total precipitation accumulated over any consecutive 5-day period.	mm
R10mm_days	Annual count of days with daily precipitation ≥ 10 mm.	days
PRCPTOT	Total annual precipitation from wet days.	mm
wet_days	Annual count of wet days (≥ 1 mm).	days
SDII	Simple Daily Intensity Index, calculated as PRCPTOT divided by wet_days.	$mm\ day^{-1}$

Table 9. Categorical verification metrics for hourly precipitation across different precision reduction configurations relative to the WRF_bl.

Configuration	ZPR (%)	POD (%)	FAR (%)	CSI	Bias
WRF_5	98.7086	92.1787	7.8366	0.8548	1.00017
WRF_4	98.7078	92.1526	7.8431	0.8545	0.99995
WRF_3	98.5641	91.2276	8.7198	0.8391	0.99942
WRF_fx5	99.8289	99.2410	1.0357	0.9822	1.00280
WRF_fx4	98.4663	94.3254	8.9820	0.8629	1.03634
WRF_fx3	99.1221	71.9411	6.8957	0.6830	0.77269
WRF_5fx5	98.6073	91.8428	8.4280	0.8468	1.00296
WRF_5fx4	97.4814	88.3594	14.7486	0.7664	1.03646
WRF_5fx3	98.5405	68.4151	11.4639	0.6285	0.77274
WRF_4fx5	98.6069	91.8190	8.4321	0.8466	1.00274
WRF_4fx4	97.4812	88.3404	14.7522	0.7663	1.03628
WRF_4fx3	98.5400	68.4076	11.4680	0.6284	0.77269
WRF_3fx5	98.4697	90.9292	9.2680	0.8320	1.00217
WRF_3fx4	97.3682	87.6111	15.4209	0.7554	1.03585
WRF_3fx3	98.4717	67.9713	12.0081	0.6220	0.77247

*Note: A threshold of 0.1 mm h^{-1} is applied to distinguish between dry and wet grids. ZPR is the Zero Preservation Ratio, POD denotes the Probability of Detection, FAR is the False Alarm Ratio, CSI is the Critical Success Index, and Bias represents the Frequency Bias.

Lines 256–273:

"Second, to evaluate the impacts of precision reduction on both local numerical accuracy and spatial structures, we analyzed point-wise deviations together with spatial similarity metrics. At the grid scale, we computed the absolute grid-scale deviation (AD), defined as the absolute difference between the precision-reduced configurations and the full-precision baseline simulation WRF_bl at each grid point and hourly time step. Because point-wise metrics alone cannot capture changes in the spatial organization of meteorological fields, we additionally employed the Structural Similarity Index Measure (SSIM) (Wang et al., 2004; Baker et al., 2019; Klöwer et al., 2021). This metric is particularly important for evaluating simulations driven by precision-reduced inputs, where small numerical perturbations may propagate through nonlinear dynamics and alter evolving mesoscale structures. Unlike AD, which measures local differences, SSIM quantifies the preservation of large-scale spatial patterns and structural textures. For a full-precision reference spatial window x and the corresponding window y from precision-reduced configurations, the SSIM is calculated as:

$$\mathrm{SSIM}(x,y)=\frac{(2\mu_x\mu_y+C_1)(2\sigma_{xy}+C_2)}{(\mu_x^2+\mu_y^2+C_1)(\sigma_x^2+\sigma_y^2+C_2)} \tag{3}$$

where μ_x and μ_y are the local means, σ_x^2 and σ_y^2 are the local variances, and σ_{xy} is the local covariance between the two fields. The stabilization constants C_1 and C_2 are scaled by the dynamic range of the specific meteorological variable being evaluated to accommodate diverse atmospheric fields. SSIM values range from 0 to 1, with unity indicating perfect structural similarity. Structural similarity was computed using an 11×11 Gaussian weighting window sliding across the domain. Given the 12 km horizontal grid spacing, this corresponds to a spatial footprint of approximately 130 km, enabling assessment of mesoscale structural consistency. To avoid artificial inflation of scores caused by large, structurally uniform dry regions, a standard wet-day mask with a threshold of 0.1 mm h^{-1} was applied prior to the calculation of precipitation SSIM."

Lines 274–283:

"Finally, the evaluation of precipitation for both the lower-tail thresholds that control the occurrence of light rainfall and the behavior of upper-tail extreme events is necessary given the highly skewed distribution of precipitation fields. Moreover, because total precipitation in WRF is represented as the cumulative sum of grid-resolved (RAINNC) and parameterized convective (RAINC) components, it exhibits a compounded sensitivity to numerical precision loss.

To assess structural fidelity at the lower bound of the precipitation spectrum, we first performed a categorical verification for hourly precipitation, using a wet–dry threshold of 0.1 mm h^{-1}. For each precision-reduced configuration, grid cells were classified relative to WRF_bl. Based on this comparison, several categorical verification metrics were calculated (Roebber, 2009), including the Probability of Detection (POD), False Alarm Ratio (FAR), Critical Success Index (CSI), and Frequency Bias (Bias). In addition, a Zero Preservation Ratio (ZPR) was introduced to quantify the fraction of dry grid cells that remain correctly classified after precision reduction."

Lines 511–555:

"3.4 Impacts on Precipitation Diagnostics: Zero-Value Preservation and Extreme Precipitation Indices

While the SSIM analysis reveals how these numerical artifacts alter the spatial structure of precipitation fields, an equally important question is how distortions introduced by precision reduction in cumulative precipitation fields propagate into downstream scientific diagnostics. Precipitation exhibits a highly skewed spatial and temporal distribution, characterized by two distinct and sensitive regimes: a widespread lower tail dominated by zero or near-zero values, and a rare but high-impact upper tail driven by extreme rainfall events. Both regimes are particularly vulnerable to numerical artifacts introduced by precision reduction. To quantify these impacts, the following section evaluates how precision-reduced configurations affect precipitation diagnostics, including zero-value preservation and extreme precipitation indices.

We first evaluate the preservation of the precipitation occurrence spectrum, specifically focusing on the zero-value boundary (using the 0.1 mm h^{-1} wet-day threshold). Standard contingency metrics, including ZPR, POD, FAR, CSI, and Bias, are employed to diagnose how different compression configurations alter the fundamental wet/dry morphology (Table 9).

For input-only precision reduction configurations, the metrics indicate that the overall precipitation spectrum is largely preserved. The Frequency Bias remains close to unity (e.g., 0.99942 for WRF_3), confirming that the total precipitation area is conserved. Furthermore, the nearly symmetric error rates, where the moderate FAR (7.84%–8.72%) closely balances the corresponding miss rates (1 - POD), together with a stable CSI (~0.84) indicate that input precision reduction induces spatial phase shifts in precipitation systems rather than systematically disrupting their structure. In contrast, aggressive output precision reduction introduces substantial numerical distortions that directly erode the bottom end of the precipitation spectrum, most prominently in the WRF_fx3 configuration. Under this configuration, the POD drops sharply to 71.94%, indicating that approximately 28% of valid light precipitation events are artificially truncated to zero due to the loss of significance in the arithmetic of cumulative precipitation variables. As a result, a pronounced systematic dry bias emerges (Bias ≈ 0.77).

Interestingly, WRF_fx3 exhibits an apparently higher ZPR (99.12%) and a lower FAR (6.90%)

compared with WRF_fx4 (98.5% and 8.98%, respectively). However, this does not indicate improved structural fidelity. Instead, it reveals that WRF_fx3 suppresses both numerical noise (reducing FAR) and genuine light stratiform precipitation (reducing POD), effectively converting them into widespread non-physical dry regions.

The structural degradation previously indicated by the late-stage decline in SSIM is further corroborated by the temporal evolution of precipitation occurrence metrics under the WRF_fx4 configuration (Fig. S4). At the domain scale, the ZPR remains high (> 97 % throughout the year), largely reflecting the overwhelming dominance of climatologically dry background grids. However, metrics that evaluate performance specifically within precipitation events (CSI, POD, and FAR) reveal a pronounced late-stage deterioration. Consistent with the SSIM decline observed in November, the CSI decreases from its summer peak (~0.89) to 0.769 by December. This degradation reflects a dual numerical distortion associated with the expanding accumulation baseline. As the quantization step increases during November and December, the FAR rises substantially (reaching 11.4%), indicating the growing occurrence of spurious drizzle signals introduced by rounding artifacts. At the same time, the POD declines to 85.4%, implying that approximately 15% of genuine winter stratiform precipitation events are artificially truncated to zero. The Frequency Bias further highlights this shift in behavior. While the system exhibits a slight wet bias in spring, it gradually transitions to a systematic dry bias by December (Bias = 0.964). The simultaneous increase in false alarms and the loss of weak precipitation events progressively disrupt the spatial consistency of precipitation structures, definitively explaining the late-stage SSIM collapse.

The analyses above focus on the lower boundary of the precipitation spectrum, where precision reduction primarily affects the occurrence of light rainfall and the preservation of dry conditions. However, precipitation diagnostics are also strongly influenced by the opposite end of the distribution.”

<u>Lines 657–682:</u>

“In stark contrast to instantaneous state variables, the WRF precipitation variable, a cumulative quantity with a highly skewed distribution, can reach thousands of millimeters annually. This continuously growing accumulation introduces fundamentally different numerical vulnerabilities, making it the primary bottleneck for precision reduction design. Specifically, retaining a fixed 3 significant digits critically widens the effective quantization interval over time (e.g., approaching ~20 mm h^{-1}). When hourly or daily precipitation is derived through temporal differencing, this coarse quantization introduces contradictory artifacts: it suppresses very light (0.1 mm h^{-1}) rainfall increments (producing a systematic dry frequency bias) while intermittently generating large, discrete step-increments, which artificially elevates the frequency of false wet_days and R10mm_days threshold exceedances. To maintain realistic precipitation statistics without indiscriminately inflating file sizes, a magnitude-aware dynamic precision strategy, scaling the retained digits according to the evolving cumulative total, emerges as a highly practical solution. Specifically, for grid cells where the accumulated precipitation remains below 1000 mm, retaining 4 significant digits is generally sufficient to preserve a scientifically viable sub-millimeter resolution. Once the accumulation surpasses the 1000 mm threshold, the precision should be dynamically increased to 5 significant digits to explicitly prevent the decimal resolution from degrading. For multi-year simulations, where total accumulations routinely exceed 10000 mm, retaining 6 significant digits becomes strictly necessary. It is worth noting that designing a lower-tier threshold (e.g., at 100 mm) is operationally unnecessary; most simulated regions rapidly exceed this baseline shortly after initialization, rendering any lower-precision tier computationally transient and practically redundant.”

Then for the "optimal truncation strategy" you decide that NMB < 1% is a sufficient condition for an acceptable compression error. Why is that? If I have [1.5001, 0.4999] and truncate this to [2, 0] (round to nearest integer) then NMB = 0 but we have increased the variance (from 1/2 to 2) and the maximum absolute error is 1/2 which might be unacceptably high. I generally propose to use a (normalized) mean and maximum absolute error or a (normalized) mean and maximum relative error depending on the data distribution. Furthermore, a spectral error is often used to investigate the impact on the small scales as rounding can introduce artificial gradients (jumps from one representable number to the next) or smooth out gradients if neighboring cells are rounded to the same value. I would reject the idea to formulate an "optimal strategy" based on solely one metric (NMB) and definitely expect a discussion around the chosen error metrics (what they measure and what they don't) and a strong justification of why you choose what you choose.

Thank you for this insightful comment. We agree with you that determining an "optimal truncation strategy" based solely on a single metric such as NMB is insufficient and potentially misleading. As you correctly pointed out, a metric such as NMB can mask important distortions in other statistical properties, including variance changes, extreme deviations, or structural distortions introduced by rounding. Furthermore, we acknowledge that the decimal-based precision reduction method employed in our study inherently carries methodological limitations when attempting to define an 'optimal' strategy.

In the revised manuscript, we have removed the previous Section 3.4 ("Optimal Truncation Strategy") entirely. Instead, we now treat this topic more cautiously and present it within a broader discussion framework. Specifically, we introduced a new section: Section 4 "Practical Guidelines and Broader Perspectives for Precision Reduction Configuration." Rather than proposing a single universal precision rule, this section synthesizes the findings from the analyses in Section 3 and translates them into practical guidance for users of WRF simulations. We believe this revised approach is both more scientifically robust and more useful for users who need to balance storage efficiency with scientific fidelity under different application contexts. The new discussion section highlights several key insights derived from our expanded evaluation framework:

First, we clarify that different precision reduction pathways introduce fundamentally different error mechanisms. Precision reduction applied to time-varying inputs interacts with nonlinear atmospheric dynamics and can introduce spatial or temporal phase shifts in simulated weather systems. Although these perturbations may lead to large grid-scale deviations, the structural similarity analysis shows that the overall morphology of atmospheric fields remains largely preserved. This implies that input precision reduction may be acceptable for applications focusing on large-scale statistics, but it may be unsuitable for studies requiring strict deterministic spatiotemporal correspondence. In addition, applying precision reduction in time-varying input forcings results in the dynamic amplification of errors during summer. Whether maintaining full precision during summer and implementing precision reduction in other seasons can yield superior overall simulation performance warrants further investigation in future studies.

Second, we emphasize that output precision reduction must be tailored to the mathematical characteristics of different variables. Instantaneous state variables (e.g., surface wind speed) are relatively tolerant to precision reduction due to their limited magnitude range (2–3 significant digits), whereas thermodynamic variables such as temperature require higher precision (at least 4 significant digits) to avoid systematic information loss. In contrast, cumulative quantities such as precipitation exhibit fundamentally different vulnerabilities because rounding errors interact with the cumulative arithmetic used in WRF outputs. This can produce quantization artifacts that distort precipitation occurrence statistics and threshold-based diagnostics. To maintain statistical fidelity while maximizing storage efficiency, we propose a magnitude-aware dynamic precision strategy. By adaptively scaling retention from 4 significant digits (below 1000 mm) to 5 or 6 digits for extreme accumulations (>1000 mm and >10000 mm, respectively), this approach strictly preserves reliable hourly precipitation information throughout the entire simulation.

Finally, we emphasize that the appropriate precision level is inherently application-dependent, particularly for diagnostics involving fine spatial gradients or temporal trends. Because these calculations rely on extracting small differences, the retained precision must be sufficient to resolve subtle physical perturbations from the background state. Consequently, the acceptable extent of precision reduction depends on how well future data usage can be anticipated. When downstream applications are broad and unpredictable, as is typically the case for static reanalysis datasets distributed by institutional centers, providers are typically required to maintain conservative, high-precision baselines. Conversely, in the context of active model post-processing (e.g., targeted WRF simulations), individual researcher often has well-defined scientific objectives. This allows them to adopt more aggressive precision reduction without compromising the physical signals of interest. Such application-driven flexibility enables targeted projects to maximize storage efficiency while preserving the fidelity required for their specific analyses.

6) Inconsistent conclusions

On line 205 you write "truncation impacts are variable-dependent" highlighting the need for precisions chosen differently by variable (supported by Fig 3). However, when you present your "optimal truncation strategy" while still mentioning the variable-dependency, you don't conclude that an adaptive strategy should adjust to different variables. Instead, you talk only about seasons and regions. Applying a different precision by variables clearly seems to be the more optimal way, so why call your strategy "optimal" when you're leaving potential on the road? If this isn't possible for practical reasons then state this. But any modern data format (netCDF, HDF5, Zarr, ...) would allow you to round variables differently, see connection to (1).

Thank you for pointing out this important inconsistency in the original manuscript. We agree with the reviewer that the previous presentation of an "optimal truncation strategy" did not fully reflect the variable-dependent nature of precision requirements that we discussed earlier in the paper. As you correctly noted, our results clearly demonstrate that the impacts of precision reduction differ substantially across variables. Therefore, a variable-specific precision configuration is indeed a more appropriate strategy for data compression in atmospheric modeling.

Following this comment, and consistent with the concerns raised in the previous major comment regarding the formulation of an “optimal truncation strategy,” we have removed the previous Section 3.4 (“Optimal Truncation Strategy”) entirely in the revised manuscript. Instead, the revised manuscript now introduces a broader discussion section: Section 4: “Practical Guidelines and Broader Perspectives for Precision Reduction Configuration.”

In this revised section, we explicitly emphasize that precision reduction should not be implemented uniformly with a specific reduction, but rather should be adapted to the characteristics of individual variables. Establishing strict, variable-specific theoretical precision limits would require distinct information-theoretic evaluations (Klöwer et al., 2021), which fall outside the scope of this operational analysis.

Finally, the precision-reduction tool developed and applied in this study supports variable-specific precision configurations. The decimal significant digits rounding algorithm allows users to assign different numbers of retained significant digits to individual variables.

In summary, the revised manuscript abandons the notion of an "optimal" truncation configuration and instead presents a flexible precision reduction framework applicable to the entire WRF workflow. We contend that this revised formulation more accurately reflects the findings of our study and better addresses your concerns.

Minor

40: Not sure what you mean by "shift" here. Both global and regional models are used operationally?

Thank you for the question. Our original thought was expressing the idea that despite the smaller spatial domain of regional models, the enhancement in spatiotemporal resolution still renders data volume, which is a significant storage management challenge. In the revised manuscript, we have revised this statement accordingly (Lines 42–43).

Lines 42–43:

“These storage constraints are not unique to global simulations. In regional models, although they focus on smaller spatial domains, data storage management is still an inevitable challenge as spatiotemporal resolution increases.”

52: But this is not the fault of gzip, bzip2, to problem is that tailing mantissa bits are high entropy and hence incompressible. Rephrase this to make this clear to the reader? Especially because you're using lossless compression later.

Thank you for the suggestion. In the revised manuscript (Lines 55–57), we clarified this point.

Lines 55–57:

"Lossless algorithms alone preserve bitwise reproducibility but generally achieve only a modest compression ratio for floating-point geophysical fields because the high entropy of mantissa bits limits compressibility (Poppick et al., 2020)."

54: terminology: absolute and relative error? (Or generally any error metric?) Yes, a relative error can be expressed in significant digits but that's just the unit, significant bits would be another?

Thank you for the question. This is an imprecise use of terminology. Following the substantial rewrite of the Introduction, this original sentence is no longer present in the revised manuscript.

56: State that this is also known as (bit) rounding? ", both operational workflows": what does "both" refer to here?

Thank you for the suggestion. To improve clarity and avoid potential confusion, we standardized the terminology throughout the manuscript. The term "precision truncation" used in the previous version has been replaced with "precision reduction" or "decimal significant-digit rounding," based on context, and expressions such as "truncation strategies" has been replaced by "precision reduction configurations." In addition, we acknowledge that the phrase "both operational workflows" contained a grammatical error. Following the substantial rewrite of the Introduction, this original sentence is no longer present in the revised manuscript.

57: What does "lightweight" mean? It surely doesn't consume memory but maybe you want to say "fast" or "cheap"?

Thank you for the question. Our original intention was to convey that the method is "computationally inexpensive". Following the substantial rewrite of the Introduction, this original sentence is no longer present in the revised manuscript.

57: "Straightforward to implement" I somewhat disagree: IEEE-754 round to nearest tie has its complexities but it's certainly an (IEEE) standard and therefore widely available and accepted.

Thank you for the comment. You are correct that implementing the IEEE-754 default "round to nearest, ties to even" rule has its complexities. Our original phrasing introduced ambiguity. In this study, we employ decimal significant-digit rounding (specifically the standard "Round Half Up" method). We originally described this as "straightforward" because it is implemented using standard Fortran intrinsic functions. Following the substantial rewrite of the Introduction, this original sentence is no longer present in the revised manuscript.

58: "utilities" -> "compressors"?

Your point of view is well taken and thank you. Following the substantial rewrite of the Introduction, this original sentence is no longer present in the revised manuscript. This term is employed in other contexts (e.g., Lines 205, 208, and 344).

59: You certainly make that statement based on previous publications. References?

Thank you for pointing this out. Our original statement lacked specific citations. Following the substantial rewrite of the Introduction, this original sentence is no longer present in the revised manuscript.

64: Certainly, agree with this statement but for an analysis across variables see Klöwer et al. 2021

Thank you for the comment. The rewritten Introduction has summarized the findings of Klöwer et al. 2021. Please see the response to Major comment 4.

66: climate -> weather events? This sounds like floods, storms or heatwaves/cold snaps to me?

Your point of view is well taken and thank you.

67: Given this sentence I don't know what you mean in the former. Please clarify?

Thank you for the question. Our original intent was simply to emphasize that extreme precipitation events are high-impact, and that evaluating the biases introduced to diagnostics by precision reduction is of critical importance. To improve clarity, the expression is revised (Lines 92–94).

Lines 92–94:

"Given the disproportionate societal and economic impacts of extreme precipitation (Seneviratne et al., 2021; Davenport et al., 2021), it is therefore essential to assess whether precision reduction introduces artificial biases in extreme-event diagnostics, such as inflated dry-area coverage or altered peak precipitation intensities, that exceed expected analysis uncertainty."

69: "can alter" -> this discussion is missing that the changes introduced by lossy compression may not be statistically significant? If lossy compression is applied right then the compression error should be masked by the analysis error.

Thank you for the comment. You are correct that carefully designed lossy compression of model outputs typically does not introduce statistically significant artifacts. The original expression was ambiguous. Our initial consideration focused on the error induced by input precision reduction. Following the substantial rewrite of the Introduction, this original sentence is no longer present in the revised manuscript.

76: on "nonlinear sensitivity" missing the point here that you also apply lossy compression to the initial conditions used for WRF. State that explicitly? Otherwise, the "nonlinear sensitivity" of data compression is confusing

Your point of view is well taken and thank you. In the revised manuscript (Lines 83–85), we have rewritten this sentence to explicitly specify the nonlinearity stems from the reduction in the precision of input data.

Lines 83–85:

"A fundamental distinction exists between the precision reduction of input data and that of output data. Perturbations introduced into the input forcings may be amplified through nonlinear model physics, potentially influencing simulated trajectories and downstream diagnostics."

104: I'm missing here a discussion how significant digits are translated to bits, see major point

Your point of view is well taken and thank you. For detailed responses, please refer to Major point 1.

104: It's unclear why you chose 3-5 significant digits and not more or less. For some variables 5 significant digits are more clearly an overkill, say instantaneous cloud cover of e.g. 0.12345 whereas for others it's not, e.g. CO2 at 428.63ppm. Motive this range here?

Your point of view is well taken and thank you. For detailed responses, please refer to Major point 3.

108: Delete "simultaneously"? There's a 1-yr simulation run in between "input" and "output", so hardly simultaneous?

Your point of view is well taken and thank you.

139: Why do you use RMSE as an error metric (which quantifies an absolute error) although your precision truncations yield relative compression errors? Sure, the relative error is therefore somewhat predictable (given it's bounded) but I suggest a discussion here whether all variables should be evaluated using an absolute error. Surely a wind speed of 0 vs 1m/s makes a difference but if it's 80 or 81 m/s probably not?

Thank you for raising this important point. We agree with you that decimal significant-digit rounding introduces bounded relative numerical errors, and therefore relative error metrics may also provide useful insights when evaluating compression artifacts.

In this study, we primarily used RMSE together with correlation *R* because these metrics are widely adopted in meteorological model evaluation and provide a physically interpretable measure of deviations in the original units of each variable. For example, in operational forecast verification and model evaluation, RMSE directly reflects the magnitude of deviations relative to typical atmospheric variability (e.g., m s^{-1} for wind speed or K for temperature), which allows straightforward comparison with observational uncertainties and natural variability.

However, as the reviewer perceptively noted, absolute metrics such as RMSE are mathematically insensitive to the background magnitude of the physical variable. A deviation of 1 m s^{-1} contributes identically to the RMSE regardless of whether the underlying flow corresponds to calm conditions or an intense storm. For example, a difference between 80 and 81 m s^{-1} represents a very small relative perturbation (~1.25%), even though the absolute deviation remains 1 m s^{-1}. Relative error metrics can indeed provide useful context for large-magnitude values. However, they become mathematically unstable near physical lower bounds. In the case of wind speed, comparing 0 and 1 m s^{-1} leads to a singularity when computing relative errors due to division by zero. Even when values are merely close to zero, relative metrics can produce artificially inflated percentages. Consequently, purely relative metrics can be problematic for variables with highly skewed distributions or frequent near-zero states, such as wind speed and precipitation.

Motivated by your prior suggestion, we expanded the evaluation framework in the revised manuscript. In Section 3.3, we introduced the SSIM as a complementary diagnostic that evaluates the preservation of spatial structures and field textures. In addition, for precipitation we incorporated categorical diagnostics focusing on the lower tail of the distribution, such as Probability of Detection (POD) and False Alarm Ratio (FAR). These metrics are particularly effective in capturing changes near the wet–dry threshold and therefore provide a more robust assessment of rounding effects near the physical lower bound.

Thank you again for this valuable suggestion, which helped us broaden the diagnostic framework used in the study.

165: "efficiency" -> "factor" or "ratio". If you actually mean efficiency, then introduce what efficiency means. Intuitively I think of efficiency as performance per resource. So it's unclear what this refers to here, could also be compressed size per (de)compression speed/time?

Thank you for the comment. We have revised the entire text and changed the compression efficiency uniformly to the relative compression ratio. Here, the relative compression ratio is defined as the compressed size of the precision-reduced configurations divided by the compressed size of the full-precision baseline.

171: I don't understand why input compression should affect output compression? Isn't there an entire simulation in between introducing high-entropy mantissa bits again?

Thank you for the question. As you correctly point out, the numerical integration within WRF acts as a generator of high-entropy information. Therefore, the low entropy (zeros) introduced by precision-reduced input should not propagate to the output data structure. To avoid potential misunderstandings, we have explicitly clarified this point in the revised Section 3.1 (Lines 309–311).

Lines 309–311:

"As physically expected, the final compressed volume of the output is dictated by the precision applied during the output post-processing stage; varying the input precision exerts no immediate influence on the final compressed wrfout file size."

181: This has been highlighted by Zender et al. and others, please cite those and present your result in discussion to their findings?

Your point of view is well taken and thank you. We have incorporated the relevant content into the revised Section 3.1 (Lines 357–361).

Lines 357–361:

"In conclusion, consistent with previous studies (Zender, 2016; Silver and Zender, 2017), precision reduction substantially improves subsequent standard lossless compression, which is also reflected in our results obtained with the decimal significant-digit rounding method. The substantial reduction in storage footprint achieved by reducing input and output precision underscores the potential of this coupled precision reduction and lossless compression strategy within the WRF workflow."

182: Do you want to add a discussion about compression speed here? How fast are both *compressors in your case?*

Your point of view is well taken and thank you. We have added a discussion in Section 3.1, as detailed in our response to Major Point 2.

190: Is Temperature in ˚C or Kelvin? O(300K) rounded to 3 significant digits rounds to 1K/˚C increments? Units are otherwise not relevant but an offset from ˚C to Kelvin is. You state this later, state it here?

Thank you for the question. The precision reduction is applied to the raw WRF output, which is in Kelvin (*K*). However, for validation against Observation data from NCDC stations (which is typically reported in ℃), we converted the temperature fields to ℃ before calculating RMSE.

Anyway, we have clarified the statements concerning temperature units and modified them in the revised Section 3.2 (Lines 390–391).

Lines 390–391:

"This behavior is largely a numerical artifact stemming from the Kelvin scale used in WRF outputs, where retaining 3 significant digits effectively removes all sub-degree decimal precision."

190: Relative humidity is likely only a post-processed output variable, calculated from temperature to get the saturation vapor pressure. So, if you see an error in relative humidity, are you sure it's not due to errors in temperature? Please clarify this dependency. 200: How do you know it's not the relationship to temperature? I would see your point if it was specific humidity, but you are analyzing relative humidity here. Also, the errors between Fig 3b&c are very similar but not to precipitation?

Thank you for raising this important point regarding the dependency between relative humidity and temperature. We agree with you that relative humidity is not an independent prognostic variable in WRF but a derived thermodynamic quantity, computed from temperature and water vapor through the saturation vapor pressure relationship. Therefore, deviations in relative humidity may indeed arise from upstream errors in temperature rather than representing independent numerical distortions in moisture.

Because RH depends directly on temperature through the saturation vapor pressure formulation, rounding-induced perturbations in temperature propagate into RH calculations. In contrast, precipitation exhibits fundamentally different behavior because it is a cumulative variable derived from model microphysics and convective parameterizations, rather than a direct thermodynamic diagnostic.

We agree that specific humidity/ water vapor mixing ratio would provide a more physically direct diagnostic. However, robust observational validation of such variables was not feasible in this study. Deriving specific humidity or water vapor mixing ratio requires concurrent high-quality surface pressure measurements. After quality control, the available dataset was reduced to only a few hundred valid stations across the entire domain, which is insufficient for statistically robust spatial verification.

We have therefore clarified these methodological constraints and explicitly discussed the thermodynamic dependency between RH and temperature in the revised Section 3.2 (Lines 391–398).

Lines 391–398:

"Importantly, the pronounced sensitivity of RH, when output precision is reduced to 3 significant digits, is intrinsically tied to these temperature deviations. Governed by the non-linear Clausius–Clapeyron relationship, numerical artifacts from temperature directly propagate into saturation vapor pressure calculations, rendering the observed RH deviations largely secondary thermodynamic artifacts. Additionally, direct observational validation of absolute moisture was precluded: deriving specific humidity or mixing ratio from dew point temperature requires concurrent pressure data, yet rigorous quality control of these pressure records limited the available data to merely a few hundred valid stations domain-wide, rendering the network statistically insufficient for robust spatial verification."

205: The meaning of "5 significant digits" also depends on whether precipitation is accumulated or a rate in the output. Can you clarify this?

Thank you for raising this important point. In WRF, total precipitation is not output as an instantaneous rate but rather as the cumulative sum of grid-scale precipitation (RAINNC) and convective precipitation (RAINC). Hourly or daily precipitation rates used in diagnostics are subsequently obtained through temporal differencing of these cumulative fields. The precision retention for precipitation mentioned in our study is based on the native output of cumulative precipitation.

To clarify this dependency, we have revised the manuscript to explicitly explain the cumulative nature of precipitation and its implications for precision reduction. The following clarifications were added:

Lines 90–91: We now explicitly state that precipitation is output as a cumulative quantity and exhibits intermittent and highly skewed characteristics. Lines 275–277: We clarify that total precipitation in WRF is computed as the cumulative sum of RAINNC and RAINC, which introduces compounded sensitivity to numerical precision loss. Lines 487–490: We introduce the concept of a dynamic precision strategy for cumulative precipitation variables, in which the retained number of significant digits increases as accumulated precipitation grows. Lines 667–682: We provide a detailed explanation of how precision reduction affects cumulative precipitation fields and how rounding errors propagate into derived hourly or daily precipitation through temporal differencing.

Lines 90–91:

"… this uncertainty becomes particularly critical in downstream diagnostic studies of precipitation, which is output as a cumulative quantity and exhibits intermittent and highly skewed characteristics."

Lines 275–277:

"Moreover, because total precipitation in WRF is represented as the cumulative sum of grid-resolved (RAINNC) and parameterized convective (RAINC) components, it exhibits a compounded sensitivity to numerical precision loss."

Lines 487–490:

"More generally, if scientific accuracy is prioritized, a dynamic precision reduction configuration can be adopted for cumulative precipitation variables. In such scheme, the compression algorithm monitors accumulated precipitation and initially retains 4 significant digits, automatically upgrading to 5 significant digits once cumulative precipitation exceeds 1000 mm at the regional average or even grid scale."

Lines 667–682:

"In stark contrast to instantaneous state variables, the WRF precipitation variable, a cumulative quantity with a highly skewed distribution, can reach thousands of millimeters annually. This continuously growing accumulation introduces fundamentally different numerical vulnerabilities, making it the primary bottleneck for precision reduction design. Specifically, retaining a fixed 3 significant digits critically widens the effective quantization interval over time (e.g., approaching ~20 mm h^{-1}). When hourly or daily precipitation is derived through temporal differencing, this coarse quantization introduces contradictory artifacts: it suppresses very light (0.1 mm h^{-1}) rainfall increments (producing a systematic dry frequency bias) while intermittently generating large, discrete step-increments, which artificially elevates the frequency of false wet_days and R10mm_days threshold exceedances. To maintain realistic precipitation statistics without indiscriminately inflating file sizes, a magnitude-aware dynamic precision strategy, scaling the retained digits according to the evolving cumulative total, emerges as a highly practical solution. Specifically, for grid cells where the accumulated precipitation remains below 1000 mm, retaining 4 significant digits is generally sufficient to preserve a scientifically viable sub-millimeter resolution. Once the accumulation surpasses the 1000 mm threshold, the precision should be dynamically increased to 5 significant digits to explicitly prevent the decimal resolution from degrading. For multi-year simulations, where total accumulations routinely exceed 10000 mm, retaining 6 significant digits becomes strictly necessary. It is worth noting that designing a lower-tier threshold (e.g., at 100 mm) is operationally unnecessary; most simulated regions rapidly exceed this baseline shortly after initialization, rendering any lower-precision tier computationally transient and practically redundant."

Fig. 3: Swap red-blue colors in e-h to signal worse with red and better with blue?

Done.

Fig. 3: Why are errors being reduced for wind? IEEE rounding is theoretically bias-free (due to round to nearest tie to even) so I don't understand what's happening here.

Thank you for the question. We agree with you that IEEE rounding is theoretically unbiased, and therefore precision reduction alone should not systematically reduce model errors.

The reduction in wind errors arises from the construction of the evaluation metric in Fig. 4, where model outputs are evaluated relative to observational reference data rather than relative to the uncompressed WRF baseline (WRF_bl). Because the baseline simulation itself contains inherent biases with respect to observations, the unbiased numerical perturbations introduced by precision reduction can occasionally act as a form of error compensation. In other words, if a rounding perturbation happens to partially offset an existing model bias, the resulting truncated value may appear closer to the observations, producing a small apparent improvement in the evaluation metrics.

Importantly, this does not indicate that precision reduction physically improves simulation. Rather, it reflects a statistical artifact of bias cancellation when comparing against observations. The purpose of this comparison is to demonstrate that precision reduction artifacts do not compromise the model's ability to pass standard observational validation benchmarks, rather than to suggest that precision reduction improves the physical realism of the simulation.

222: Please mention units earlier, see above.

Your point of view is well taken and thank you (Lines 390–391).

Lines 390–391:

"This behavior is largely a numerical artifact stemming from the Kelvin scale used in WRF outputs, where retaining 3 significant digits effectively removes all sub-degree decimal precision."

Fig 4: Why do you take the absolute value of RMSE changes? One is better the other one worse, could you clarify first why lower RMSE follows from rounding?

Thank you for the question. As discussed in our prior response, these small negative values are not physical improvements. Thus, we treat it as a perturbation magnitude. In addition, absolute values facilitate the visualization of stacked plots.

Fig 4: Can you please change the colors for the regions? The red and the green are pretty much indistinguishable for someone with deuteranopia (the most common color vision deficiency), make one brighter the other one darker for example

Your point of view is well taken and thank you. We have redrawn the color schemes of all the charts in the entire manuscript. The following revised Fig. 4 (Fig. 5 in the revised manuscript) provides as an illustration.

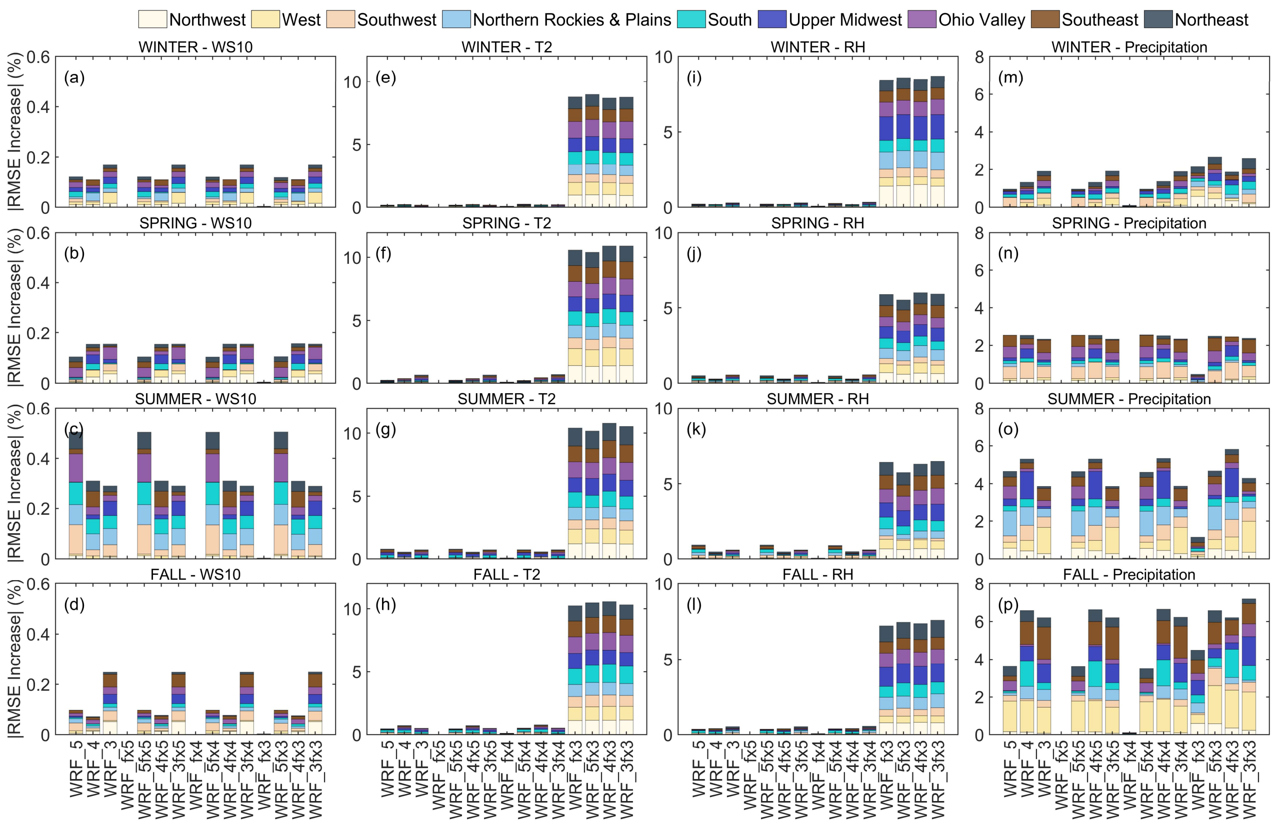

Figure 5: Magnitude of seasonal and regional relative changes in RMSE for meteorological variables across 15 precision-reduced configurations. Stacked bars illustrate the regional breakdown of |RMSE increase| relative to the WRF_bl across nine climate regions. Note that the total bar heights serve solely to visualize relative regional contributions and do not represent a mathematically aggregated domain-wide RMSE percentage. Results are presented separately for (a–d) wind speed, (e–h) temperature, (i–l) relative humidity, and (m–p) precipitation.

226: See comment above, can you show that humidity here is actually independent of temperature?

Thank you for the question. Please refer to the prior response (page 31).

250: How do you know that +-3% is modest? For temperature a 1% error was not acceptable.

Thank you for the comment. We have revised it to a more rigorous formulation in the Section 3.4 (Lines 565–563).

Lines 565–563:

"For the percentile-based index R99p_days, NMB fluctuates within ±3% during winter, spring, and fall, while it amplifies in summer to a maximum bias of 8.33% (Fig. 8a–d)."

316: Why is NMB < 1% a good (single) metric to decide whether your compression error is acceptable? See major point.

Thank you for the question. We agree that single metric is insufficient to identify the optimal. Please refer to our response to Major Point 6.

322: I don't disagree with that but that's the holy grail of lossy data compression: How do choose an acceptable compression error for a (not yet decided) set of scientific objectives.

Thank you for this thoughtful comment. We fully agree with you that determining an acceptable compression error for unknown future scientific applications is indeed one of the fundamental challenges in lossy data compression.

In practice, however, the acceptable level of precision reduction is inherently application-dependent. Large institutional archives (e.g., global reanalysis datasets) must support a wide range of unpredictable downstream uses, and therefore are typically required to adopt conservative precision baselines to avoid the risk of losing potentially relevant information.

In contrast, regional modeling workflows such as WRF are often goal-oriented. Simulations are typically conducted with clearly defined scientific objectives (e.g., extreme precipitation, boundary-layer processes, or regional climate diagnostics), and the required precision can therefore be aligned with the physical signals that the analysis aims to resolve. In such contexts, precision reduction can be selectively applied, preserving the precision required for scientific analysis for the variables and diagnostics of interest while reducing redundant numerical precision elsewhere.

We have clarified this principle in the revised manuscript (Section 4.2: Lines 683–702): WRF users can tailor precision levels according to their specific research objectives, allowing substantial storage savings without compromising the scientific signals relevant to their analyses.

Lines 683–702:

"Considering the specific application context, precision required for post-processed diagnostics is not universally applicable. The acceptable extent of precision reduction is ultimately determined by the relative magnitude between the physical signals that downstream analyses intend to resolve and the background state of the variable. The retained precision must be sufficient to resolve the order-of-magnitude contrast between the background state and the targeted physical perturbation, but only when resolving that perturbation is scientifically necessary. For example, atmospheric pressure typically has a background magnitude of ~10^5 Pa. If a study seeks to diagnose mesoscale pressure gradients on the order of ~10^1 Pa, retaining at least five significant digits becomes necessary to avoid numerical loss of information. Conversely, if the scientific objective focuses only on large-scale synoptic patterns, where variations of ~10^2 Pa are sufficient to characterize the system, retaining four significant digits may already provide adequate fidelity while significantly improving storage efficiency. Following analogous reasoning, water vapor mixing ratio typically exhibits background magnitudes of ~10^{-2}–10^{-4} kg kg^{-1}, while dynamically relevant perturbations associated with moisture advection may occur at ~10^{-5} kg kg^{-1}. Resolving such subtle signals requires at least four significant digits. However, studies concerned primarily with bulk moisture transport or large-scale moisture budgets may tolerate lower precision because these micro-scale perturbations contribute negligibly to the targeted diagnostics. Therefore, the level of precision reduction should ideally be chosen based on the requirements of the intended downstream scientific analysis, ensuring that numerical compression does not obscure the physical signals that the research aims to diagnose. When downstream applications are broad and unpredictable, as is typically the case for static reanalysis datasets distributed by institutional centers, providers are typically required to maintain conservative, high-precision baselines. Conversely, in the context of active model post-processing (e.g., targeted WRF simulations), researchers often have well-defined scientific objectives. This allows them to precisely tailor the extent of precision reduction to their specific needs, maximizing storage efficiency without compromising the physical signals of interest."

324: As far as I understand your nudging applied it's not just an initial or boundary condition but also a forcing term in the upper atmospheric levels that's altered.

Thank you for the comment. Here is a textual error. Precision reduction was applied prior to model execution and exclusively only to the time-varying input forcing files, encompassing atmospheric analysis nudging fields (wrffdda), surface analysis nudging fields (wrfsfdda), lateral boundary tendencies (wrfbdy), and lower boundary forcing (wrflowinp), such as sea surface temperature.

The initial atmospheric state is provided via either an initial condition file or a restart file (wrfrst) generated after model spin-up and serves as the starting point for the integration, providing the full three-dimensional atmospheric and land-surface prognostic states required for seamless temporal continuity. For consistency purposes, the annual run started with a 2-day spin-up on 12/30/2015 so every day in 2016 the model started with a restart file and four prescribed time-varying external forcing files, which include atmospheric and surface analysis nudging fields (wrffdda, wrfsfdda), lateral boundary condition tendencies (wrfbdy), and lower boundary inputs such as sea surface temperature (wrflowinp). The outputs are written to wrfout files and contain near-surface and surface variables, three-dimensional atmospheric state fields, cloud microphysical quantities, radiation and energy fluxes, and other diagnostics. We have systematically detailed the exact data structures of the WRF model's inputs and outputs in the revised manuscript (Section 2.1: Lines 136–143, 144–148).

The entire precision reduction workflow is visually anchored by the newly introduced Figure 1. As detailed in the revised manuscript (Section 2.2: Lines 185–194), we delineate the workflow into preprocessing, model integration, and post-processing stages.

Lines 136–143:

“To establish dynamical consistency prior to the evaluation period, the annual simulation for 2016 was preceded by a 2-day spin-up initialized on 30 December 2015. The annual integration was conducted in consecutive daily segments, with each day initialized from a model restart file (wrfrst) generated at the conclusion of the previous day. These restart files provide the three-dimensional atmospheric and land-surface state required for seamless temporal continuity. In addition to daily initialization, the model evolution was continuously constrained by prescribed time-varying external forcing fields, including atmospheric and surface analysis nudging (wrffdda, wrfsfdda), lateral boundary condition tendencies (wrfbdy), and lower-boundary updates such as sea surface temperature (wrflowinp).”

Lines 145–147:

“In contrast, model outputs (wrfout) contain near-surface variables, three-dimensional atmospheric states, cloud microphysical quantities and energy fluxes, representing the prognostic results of the completed simulation.”

Lines 185–194:

“The systematic integration of this rounding framework into the overall experimental design is summarized in Fig. 1. The schematic depicts the three core workflow stages: preprocessing on input with the precision-reduction tool, model integration with full-precision input and precision-reduction input, and post-processing on model integration result, wrfout file, with the precision-reduction tool. Lossless compression will be applied to all input and output. For input precision reduction configurations, the rounding procedure was applied during the preprocessing stage, exclusively targeting the time-varying forcing files (wrffdda, wrfsfdda, wrfbdy, and wrflowinp). Because these files define the dynamic external constraints on the model, evaluating each input precision level necessitated a fully independent WRF simulation to capture the non-linear propagation of rounding errors. Conversely, for output precision reduction, rounding was applied to the final wrfout files after simulation completion. As a purely static post-processing operation, output precision can be flexibly adjusted without requiring computationally expensive model re-runs.”

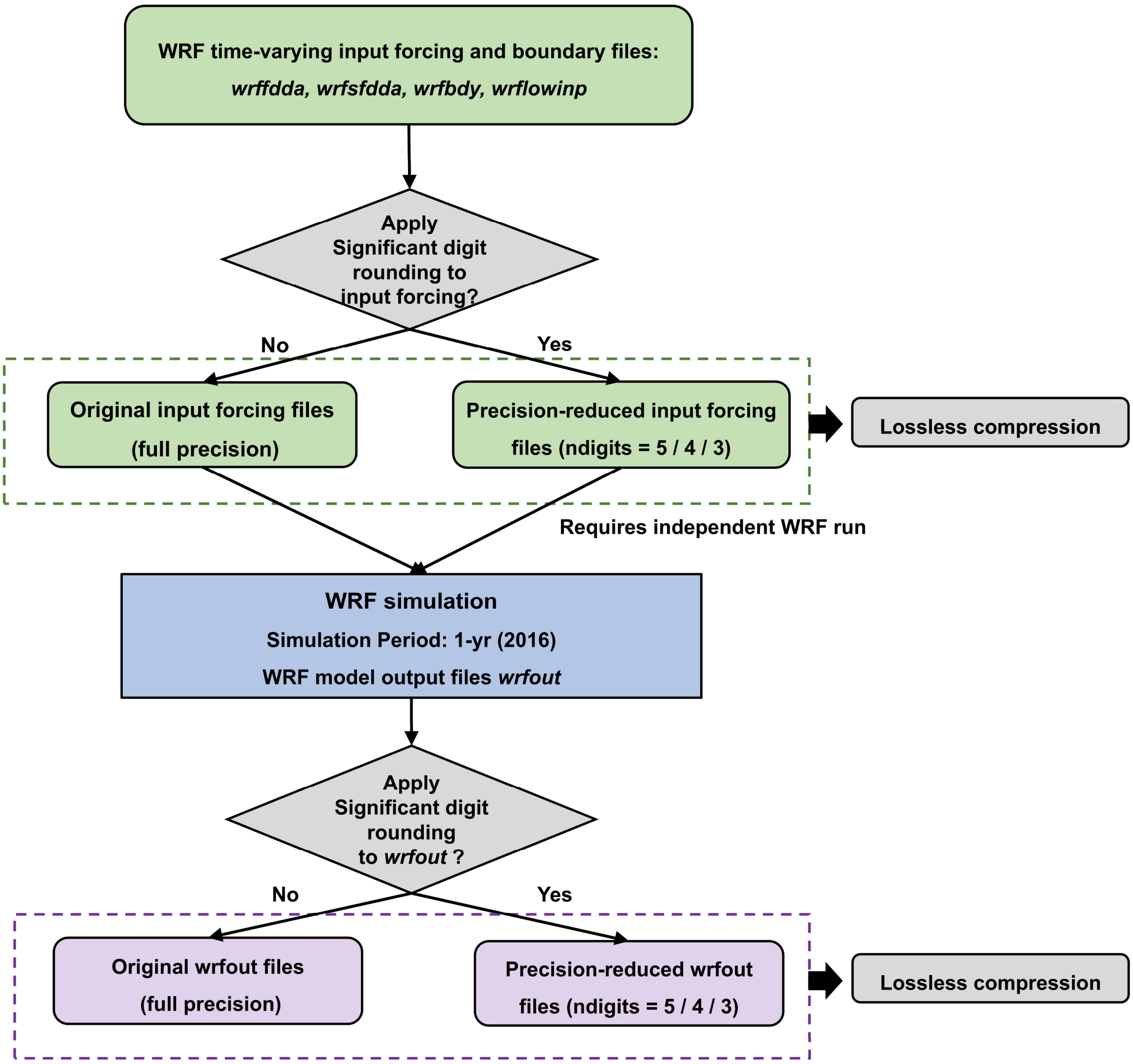

Figure 1: Schematic illustration of the experimental workflow for WRF data precision reduction and lossless compression.

331: fx6...15 would be even more conservative, given the subjective choice of 3...5 why do you conclude that 5 is the "most conservative"? Of course, higher values would defeat the point of lossy compression, but the conclusion drawn seems therefore very subjective.

Thank you for the comment. Please refer to our response to Major Point 3.

332: I don't understand why you don't suggest fx3 for winds, from Fig 3 that would follow as acceptable? But maybe you first have to explain the impact rounding has on the winds, see comment above.

Thank you for this helpful comment. We agree with you that the results indicate that wind fields can tolerate more aggressive precision reduction than the conservative configurations originally discussed in the manuscript.

In the original version of the manuscript, we adopted a conservative precision reduction level across all variables, primarily as a concession to precipitation-related considerations. However, as you correctly point out, this approach does not fully reflect the strong variable-dependent compressibility revealed by our analysis. We have revised the discussion (Lines 649–667) in the manuscript to explicitly emphasize variable-specific precision reduction strategies rather than a single uniform truncation configuration. Under this framework, retaining 3 significant digits for wind speed is recommended.

Lines 649–667:

"Dynamic variables such as WS10 are typically concentrated within a narrow numerical range, retaining 3 significant digits completely guarantees the preservation of at least one decimal place (e.g., a resolution of 0.1 m s^{-1}). Therefore, applying a universal 3-digit retention for wind fields represents the most conservative compromise between storage reduction and precision, introducing negligible numerical distortion in our evaluation. Theoretically, retaining 2 significant digits could also be a viable configuration. Because 2 digits mathematically preserve single-decimal resolution for values below 10 m s^{-1}, perceptible quantization errors would primarily be expected to emerge only when wind speeds exceed this threshold. Consequently, to maximize storage benefits without compromising physical fidelity, future applications could implement a magnitude-aware adaptive strategy for wind fields: dynamically toggling between 2 and 3 retained significant digits based on a 10 m s^{-1} threshold."

367: Thanks for providing the code but please provide a readme or documentation, I can hardly look through hundreds of lines of Fortran code to understand what you did or how you organize your code. I was looking for where you actually apply the rounding but struggled to find it.

Thank you for the comment. For the Fortran implementation, please refer to our response regarding Major Point 1. Also, we have updated the uploaded code and added a detailed Readme file. https://doi.org/10.5281/zenodo.19199806 (Wong and Wu, 2026).

References

Baker, A. H., Hammerling, D. M., and Turton, T. L.: Evaluating image quality measures to assess the impact of lossy data compression applied to climate simulation data, Computer Graphics Forum, 38, 517–528, https://doi.org/10.1111/cgf.13707, 2019.

Baker, A. H., Hammerling, D. M., Mickelson, S. A., Xu, H., Stolpe, M. B., Naveau, P., Sanderson, B., Ebert-Uphoff, I., Samarasinghe, S., and De Simone, F.: Evaluating lossy data compression on climate simulation data within a large ensemble, Geoscientific Model Development, 9, 4381–4403, https://doi.org/10.5194/gmd-9-4381-2016, 2016.

Burrows, M. and Wheeler, D. J.: A block-sorting lossless data compression algorithm, Digital Equipment Corporation SRC Research Report 124, 1994.

Collet, Y. and Kucherawy, M.: Zstandard Compression and the 'application/zstd' Media Type, Internet Engineering Task Force (IETF), RFC 8878, https://doi.org/10.17487/RFC8878, 2021.

Davenport, F. V., Burke, M., and Diffenbaugh, N. S.: Contribution of historical precipitation change to US flood damages, Proceedings of the National Academy of Sciences, 118, e2017524118, https://doi.org/10.1073/pnas.2017524118, 2021.

Delaunay, X., Courtois, A., and Gouillon, F.: Evaluation of lossless and lossy algorithms for the compression of scientific datasets in netCDF-4 or HDF5 files, Geoscientific Model Development, 12, 4099-4113, https://doi.org/10.5194/gmd-12-4099-2019, 2019.

Han, T., Chen, Z., Guo, S., Xu, W., and Bai, L.: CRA5: Extreme Compression of ERA5 for Portable Global Climate and Weather Research via an Efficient Variational Transformer, arXiv:2405.03376, https://doi.org/10.48550/arXiv.2405.03376, 2024.

Huang, L. and Hoefler, T.: Compressing multidimensional weather and climate data into neural networks, arXiv:2210.12538, https://doi.org/10.48550/arXiv.2210.12538, 2023.

IEEE Standard for Binary Floating-Point Arithmetic, Institute of Electrical and Electronics Engineers, 1–20, https://doi.org/10.1109/IEEESTD.1985.82928, 1985.

Klöwer, M., Razinger, M., Dominguez, J. J., Düben, P. D., and Palmer, T. N.: Compressing atmospheric data into its real information content, Nature Computational Science, 1, 713-724, https://doi.org/10.1038/s43588-021-00156-2, 2021.

Mirowski, P., Warde-Farley, D., Rosca, M., Grimes, M. K., Hasson, Y., Kim, H., Rey, M., Osindero, S., Ravuri, S., and Mohamed, S.: Neural Compression of Atmospheric States, arXiv:2407.11666, https://doi.org/10.48550/arXiv.2407.11666, 2024.

Poppick, A., Nardi, J., Feldman, N., Baker, A. H., Pinard, A., and Hammerling, D. M.: A statistical analysis of lossily compressed climate model data, Computers & Geosciences, 145, 104599, https://doi.org/10.1016/j.cageo.2020.104599, 2020.

Roebber, P. J.: Visualizing multiple measures of forecast quality, Weather and Forecasting, 24, 601-608, https://doi.org/10.1175/2008WAF2222159.1, 2009.

Seneviratne, S. I., Zhang, X., Adnan, M., Badi, W., Dereczynski, C., Di Luca, A., Ghosh, S., Iskandar, I., Kossin, J., Lewis, S., Otto, F., Pinto, I., Satoh, M., Vicente-Serrano, S. M., Wehner, M., and Zhou, B.: Weather and Climate Extreme Events in a Changing Climate, in: Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change, edited by: Masson-Delmotte, V., Zhai, P., Pirani, A., Connors, S. L., Péan, C., Berger, S., Caud, N., Chen, Y., Goldfarb, L., Gomis, M. I., Huang, M., Leitzell, K., Lonnoy, E., Matthews, J. B. R., Maycock, T. K., Waterfield, T., Yelekçi, O., Yu, R., and Zhou, B., Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 1513–1766, https://doi.org/10.1017/9781009157896.013, 2021.

Silver, J. D. and Zender, C. S.: The compression–error trade-off for large gridded data sets, Geoscientific Model Development, 10, 413–423, https://doi.org/10.5194/gmd-10-413-2017, 2017.

Underwood, R., Bessac, J., Di, S., and Cappello, F.: Understanding the effects of modern compressors on the community earth science model, 2022 IEEE/ACM 8th International Workshop on Data Analysis and Reduction for Big Scientific Data (DRBSD), https://doi.org/10.1109/DRBSD56682.2022.00006, 2022.

Walters, M. S. and Wong, D. C.: The impact of altering emission data precision on compression efficiency and accuracy of simulations of the community multiscale air quality model, Geoscientific Model Development, 16, 1179–1190, https://doi.org/10.5194/gmd-16-1179-2023, 2023.

Wang, Z., Bovik, A. C., Sheikh, H. R., and Simoncelli, E. P.: Image quality assessment: from error visibility to structural similarity, IEEE Transactions on Image Processing, 13, 600–612, https://doi.org/10.1109/TIP.2003.819861, 2004.

Wong, D. C. and Wu, S.: Precision Reduction tool for the paper 'A Flexible Framework for Precision Reduction of WRF Inputs and Outputs to Balance Storage Efficiency and Scientific Fidelity' [dataset], https://doi.org/10.5281/zenodo.19199806, 2026.

Zender, C. S.: Bit Grooming: statistically accurate precision-preserving quantization with compression, evaluated in the netCDF Operators (NCO, v4.4.8+), Geoscientific Model Development, 9, 3199–3211, https://doi.org/10.5194/gmd-9-3199-2016, 2016.