

Point-by-point reply on RC3

Black font = Reviewers' comments

Blue font = Response to reviewers' comments

Red font = Changes in the text of the manuscript

We thank the editor and reviewers for the valuable feedbacks on our manuscript. We have addressed all point-by-point reply below.

Reviewer Comments

Review of: MSDT v1.0: An Open-source Multi-method Snow Drought Toolbox (egusphere-2026-4290)

General Comments

I read this manuscript with great interest and genuine enthusiasm. The snow drought research community has long needed a unified computational platform, and MSDT represents exactly the kind of foundational infrastructure that facilitates reproducible, intercomparable science. The authors have done commendable work in harmonizing seven different snow drought identification methods across multiple temporal scales, packaging them with automated data access and interactive visualization tools, and demonstrating their application through well-chosen and geographically diverse case studies. The modular architecture and standardized NetCDF output are particularly thoughtful design choices that will serve the community well.

That said, I believe the manuscript can be substantially strengthened in several areas. My comments below are intended as constructive suggestions to help the authors realize the full potential of this contribution.

Major Recommendations

1. Add a quantitative method intercomparison framework to demonstrate MSDT's unique value proposition.

One of the most exciting promises of MSDT is that it enables systematic comparison among snow drought definitions. The current manuscript demonstrates this qualitatively (Figs. 3-11) but stops short of providing quantitative metrics that would help users understand when and why methods diverge. I encourage the authors to consider adding:

- A spatial agreement map (e.g., the fraction of methods that agree on drought classification at each grid cell) for a representative year, which would immediately reveal regions of high vs. low methodological consensus.
- A pairwise confusion matrix or Cohen kappa among methods for drought/no-drought classification, perhaps stratified by climate zone (e.g., maritime, continental, alpine, Arctic).

- Timeseries of the "ensemble spread" in drought area across methods, which would quantify how much the choice of method affects the global drought trend narrative.

These additions would transform Sect. 3 from a demonstration of functionality into a genuine scientific contribution - a quantitative assessment of methodological uncertainty in snow drought research. This is where MSDT can truly add unique value beyond being a software package.

Response: Thank you for this constructive suggestion. We agree that quantitative intercomparison is important for demonstrating the value of MSDT and for characterizing the dependence of snow-drought assessments on the selected definition. We have therefore performed the requested analyses and present them in a dedicated section of the Supplementary Material (Text S4), including spatial consensus maps, pairwise Cohen's kappa and drought Jaccard matrices, and time series of drought-area fractions and between-method spread. The Supplement also documents the temporal harmonization, common valid domains, area weighting, and sensitivity to the criteria used to summarize SWEI and PDM annually.

2. Provide quantitative software performance metrics to substantiate efficiency claims.

Section 2.3.5 describes an impressive optimization strategy, and I appreciate the care taken with lazy loading, float32 precision, adaptive parallelism, and chunked output. However, the evaluation remains qualitative. I suggest adding a brief benchmarking section or table:

- Test platform specifications (CPU model, RAM, OS, Python version).
- For a representative global ERA5-Land dataset (1980-2024, 0.25 deg, snow-covered domain): wall-clock time and peak RAM for each of the seven methods.
- Storage comparison: NetCDF4 compressed vs. uncompressed file sizes for a standard output set.

These benchmarks would transform the optimization section from a description of intent into a documented achievement, greatly strengthening the software contribution.

Response: We thank the comments for requesting standardized quantitative performance evidence, which was mentioned by all three reviewers. The supplementary benchmarking attachment now reports all seven methods, temporal and spatial scaling, and lossless output compression.

All tests used the same Intel Core i9-12900K workstation (16 cores, 24 logical processors), 128 GB RAM, Windows 11 Pro, and Python 3.11.7. The full-domain inputs cover 20–84°N at 0.25° over 1980–2024 with the supplied extent mask. Annual methods and PDM use daily inputs, whereas SWEI uses monthly SWE. Each configuration was evaluated in three independent processes. Reported medians include input reading, calculation, compression and output completion; peak RAM is the maximum process resident memory.

The five annual methods required 124.6–319.3 s, with peak RAM medians of 1.76–

27.11 GiB. SWEI required 117.3 s and 5.28 GiB using 3-month accumulation and empirical Gringorten plotting positions. PDM required 27.4 min and 28.45 GiB, producing 44 annual files totalling 7.38 GiB. Lossless compression reduced the tested output sizes by 88.2–98.2%, with identical decoded values.

The scaling tests cover 10-, 20- and 45-year records and regional grids at 0.25°, 0.1° and 0.01°. The finer regional grids are computational stress tests derived by nearest-neighbour replication, not native observations or demonstrations of full-domain high-resolution feasibility. Based on the measured peaks, 64 GB RAM is a practical planning recommendation for larger workloads, with headroom for the operating system.

For more details, please refer to Text S3 in the Supplement.

3. Situate MSDT more explicitly in the broader drought informatics landscape.

The Introduction appropriately mentions meteorological and hydrological drought indices (Hao et al., 2017; Hayes et al., 2011) as models for MSDT. I suggest expanding this comparison in the Discussion by drawing concrete parallels with successful drought software packages (e.g., SPEI R package, scPDSI, xclim). What design lessons from these mature packages informed MSDT? What gaps remain? This would help readers understand MSDT not as an isolated effort but as part of a broader movement toward standardized, community-vetted climate hazard software.

Response: Thank you for this constructive suggestion. We have expanded the Discussion to position MSDT within the broader ecosystem of open-source drought and climate-analysis software. The revised text draws concrete parallels with the SPEI R package, scPDSI, and xclim, highlighting transparent statistical configuration, explicit calibration settings, and the separation of computational routines from user-facing interfaces. We explain how these principles are reflected in MSDT's configurable SWEI calculations, baseline-dependent identification methods, and modular architecture. We also distinguish MSDT's snow-drought-focused graphical workflow from these established packages. This comparison clarifies both MSDT's complementary contribution and its priorities for further development.

“The development of MSDT can be situated within the broader movement towards standardized, reusable drought and climate-analysis software. The SPEI R package exposes accumulation scales, probability distributions, fitting methods, and reference periods through documented arguments, illustrating the importance of making statistical choices explicit and reproducible (Beguería and Vicente-Serrano, 2023). MSDT follows a similar principle by allowing users to select SWEI accumulation scales and probability distributions rather than relying on a single fixed configuration. The scPDSI package provides explicit calibration-period settings, highlighting the importance of reference conditions when interpreting drought severity across locations and periods (Zhong et al., 2018). This principle is reflected in MSDT's baseline-dependent thresholds and configurable classification rules, although these are not equivalent to the self-calibration procedure used by scPDSI. Meanwhile, xclim

separates numerical index calculations from an indicator framework that handles metadata, units, and missing-data checks (Bourgault et al., 2023). MSDT's separation of analysis routines from its graphical interface is consistent with this modular approach, while xclim provides a more developed model for systematic data validation. MSDT's complementary contribution is a graphical workflow integrating seven snow drought identification methods with data acquisition, visualization, and result export."

Specific Suggestions

Introduction (L64-67): The paragraph on existing methods is effective. A small suggestion: explicitly tabulate which method combinations have never been compared in a single study before MSDT. This would sharpen the novelty claim by showing the gap MSDT fills.

Response: Thank you for this helpful suggestion. We have highlighted the limited number of studies of snow drought across multiple identification methods at the end of the third paragraph of the Introduction:

"Consequently, many snow drought studies rely on a single identification method, leaving the sensitivity of their conclusions to alternative definitions insufficiently explored. Although existing intercomparisons have demonstrated that method choice can affect inferred drought characteristics (Gottlieb and Mankin, 2022), systematic comparisons spanning annual, monthly, and daily approaches remain limited."

L134-171: SWEI/PDM descriptions are vague. How is probability distribution selected? What smoothing is applied? These details matter for reproducibility.

Response: Thank you for pointing this out. We have expanded the descriptions of SWEI and PDM in the revised manuscript and added detailed pseudocode and flowcharts in the Supplementary Material. For SWEI, we now specify that the probability distribution is selected by the user from the options implemented in MSDT. We also describe the fitting and rejection rules: parametric fitting is performed for each grid cell and calendar month, and when fitting is not possible because of insufficient finite distinct values or numerical failure, the empirical Gringorten method is used as a fallback. Probabilities are clipped before transformation to avoid infinite standardized values. For PDM, we now specify that daily SWE and precipitation data are first restricted to the selected winter period and complete hydrological years, after which a centered 7-day rolling mean with at least three valid observations is applied. The revised text also documents the percentile-threshold calculation, SWE and precipitation percentile ranking, snow drought type classification, missing-value handling, and output variables.

For more details, please refer to Text S1 in the Supplement.

Section 2.2 (L172-178): The snow cover masking procedure is important but described too briefly. Please add: (a) MODIS product version and collection number; (b) regridding method to 0.25 degrees; (c) justification for the 5% and 75% thresholds with sensitivity tests or supporting citations.

Response: Thank you for this helpful comment. We have revised Section 2.2 to provide

a clearer description of the snow cover masking procedure:

“In this study, we first calculate the 3-month rolling snow cover fraction climatology from MODIS/Terra Monthly Snow Cover data set (MOD10C1) observations (2000-2022), then apply a 5% threshold to remove ephemeral snow regions. Monthly SWE records from the land component of the fifth generation of European ReAnalysis (ERA5-Land, 1980-2024) are further used to retain pixels where snow is present for at least 75% of a given month across the study period (Huning and AghaKouchak, 2020; Li and Wang, 2022). All data are regridded to 0.25° spatial resolution using the bilinear interpolation method before identifying snow-covered area.”

Section 2.3.3 (Visualization): Basemap is deprecated in favor of Cartopy. Please justify this dependency choice for long-term maintainability, or migrate to Cartopy.

Response: Thank you for the comment. Basemap was used during the initial stage of development. In the version submitted, Matplotlib and Cartopy were employed. We have revised the relevant statements:

“Interactive geospatial visualization is implemented using Matplotlib and Cartopy, ...”

Sect. 2.3.5: Adaptive parallelism description is hand-waving. How is available memory detected programmatically? What is the exact partitioning heuristic? Provide pseudocode.

Response: Thank you for this comment. We have provided detailed pseudocode and parameter definitions in Text S1 of the Supplement to document the daily PDM parallel computing strategy. The workflow integrates memory-aware input loading, spatially partitioned calculation, and incremental annual output.

After restricting SWE, reference SWE, and precipitation inputs to the selected winter period, MSDT detects available system memory using `psutil.virtual_memory().available`. Input sizes are estimated from the filtered array dimensions assuming float32 storage. The default working-memory budget is approximately 50% of available memory, initially subject to a minimum of 1 GiB and subsequently capped to retain system headroom equivalent to the larger of 2 GiB or 20% of available memory, where possible. The streaming planner allocates an estimated working-memory allowance of 16 times the raw input-block size to accommodate input arrays, intermediate calculations, and block-level results. Small inputs are processed as a single block when they satisfy both the adaptive in-memory limit, which does not exceed 2 GiB by default, and the working-memory budget divided by this allowance. Larger inputs are partitioned into latitude blocks, with nominal input-block targets constrained to 256 MiB–1 GiB. Each block retains the complete filtered temporal dimension required for reference-period percentile estimation.

Block calculations use a `ThreadPoolExecutor` with at most two concurrent workers. Concurrency is determined by the estimated working-memory requirement and the number of active blocks. A bounded submission queue controls the number of outstanding tasks, and replacement tasks are submitted after preceding block results

have been written. Dask uses a single-threaded scheduler within this workflow. NetCDF access is protected by a shared library lock, and a single writer incrementally stores each block at its corresponding spatial position in annual output files. Block results are released after writing. Blocks containing no valid extent-mask cells are skipped.

Further details are provided in Text S1 of the Supplement. The relevant manuscript text reads:

“For the computationally intensive daily PDM method, MSDT uses a memory-aware calculation and output workflow. Inputs are first restricted to the selected winter period. Small datasets are processed as a single spatial block, whereas larger datasets are partitioned into latitude blocks that retain the complete temporal sequences needed for percentile estimation. Block size and concurrency are determined from the filtered input dimensions and available system memory, with explicit parallel execution limited to two workers. A bounded task queue controls the number of resident block results, and blocks outside the extent mask are omitted. Floating-point intermediate arrays use float32 precision where appropriate. A single writer incrementally stores block results in annual NetCDF files, and these results are released after writing. Outputs use lossless NetCDF4 compression at level 4, with shuffle enabled and storage chunks aligned with spatial blocks. This workflow controls memory demand throughout calculation and output and reduces storage requirements.”

Section 3.1 (L296-330): 22-year split comparison is too crude. Suggest using Theil-Sen slopes with Mann-Kendall significance testing to provide rigorous, comparable trend characterization.

Response: Thank you for this helpful suggestion. For the annual-scale methods, each grid cell yields 44 annual binary records, with 1 indicating a snow drought year and 0 indicating a non-drought year. The numerous tied values can make the Theil–Sen slope uninformative, potentially yielding a zero slope despite differences in occurrence frequency between periods. Although the Mann–Kendall test can accommodate ties, the purpose of this analysis is to compare snow drought occurrence frequencies between two equal-length periods rather than to estimate a monotonic trend. We therefore retain the comparison between the two 22-year periods in the revised manuscript.

Discussion (L415-443): Add a paragraph on the limitations of SWE-only approaches vs. multi-variable drought indicators (e.g., incorporating temperature for warm vs. dry snow drought discrimination). This would guide users and set the stage for planned future extensions.

Response: Thank you for your comment. In the revised discussion, we have envisioned that snow drought identification methods that take into account non-temperature and non-precipitation factors should be integrated into MSDT:

“An important direction is to integrate snow drought identification methods that account for drivers beyond air temperature and precipitation, such as radiation, wind

redistribution, sublimation, and blowing-snow processes. These factors can strongly influence snow accumulation, redistribution, and ablation, particularly in open, alpine, and high-latitude environments where wind exposure and surface energy balance play major roles.”

References: Multiple encoding errors in accented names. Several missing DOIs. Journal name abbreviations are inconsistent (full names vs. abbreviated).

Response: Thank you for pointing this out. We have carefully checked and revised the reference list throughout the manuscript. Encoding errors in accented author names have been corrected, and all references were inspected to ensure that special characters are displayed properly. We also added missing DOIs where available and verified the DOI links against the corresponding publications. In addition, journal names have been standardized to a consistent format following the journal style, avoiding mixed use of full journal names and abbreviations. References without DOIs, such as reports, datasets, or web resources, have been formatted consistently with access information where appropriate.

Summary and Recommendation

MSDT addresses a genuine and well-motivated need in the snow hydrology community. The software design is thoughtful, the multi-method integration is valuable, and the demonstration cases are well-chosen. I am optimistic about this contribution for snow drought research.

My primary recommendation is that the authors consider the major additions suggested above—particularly quantitative method intercomparison and software performance benchmarks — which would elevate the manuscript from a solid software announcement into a contribution that simultaneously advances both snow drought science and community cyberinfrastructure.

Recommendation: Major Revision. I look forward to seeing a strengthened version of this manuscript.

Response: Thank you for your insightful comments and the time you've taken.