Response to Review Comment #1

The manuscript presents "RTSEvo v1.0," a dynamic evolution model for Retrogressive Thaw Slumps (RTS). The authors address a critical gap in current research by advancing from static susceptibility mapping to spatiotemporal simulation. I find the methodological framework to be highly innovative. This hybrid approach offers valuable insights into the morphological evolution of RTS and significantly contributes to our understanding of abrupt thaw processes and permafrost degradation mechanisms on the Qinghai-Tibet Plateau. The study is well-structured and tackles a timely issue in permafrost science. However, I have several concerns regarding the spatial scale of the input data, the transferability of the model, and the standardization of the source code that need to be addressed before publication.

Response：We sincerely thank the reviewer for the constructive and insightful comments and for the positive evaluation of our manuscript. We have revised the manuscript and the source code accordingly to address these concerns and to improve the quality and clarity of the manuscript.

1. Scale Mismatch and Resolution Limitations: Retrogressive Thaw Slumps (RTS) are local-scale periglacial landforms. In this study, the selected RTS features have an average area of 2.61 ha, which corresponds to an approximate diameter of 160 m (assuming a square geometry). However, the spatial resolution of several predictor variables used in the model is relatively coarse; for instance, the NDVI data is at 250 m resolution. Even utilizing 30 m resolution datasets may be insufficient for simulating such small-scale geomorphological processes. Consequently, this "scale gap" represents a significant constraint on the model's precision. I strongly recommend that the authors include a serious discussion regarding this limitation in the manuscript.

Response: Thank you. We acknowledge that RTS are local-scale landforms, and that the use of coarser-resolution environmental predictors may limit the representation of fine-scale heterogeneity. The RTSEvo framework is designed to partially mitigate this scale gap through model structure rather

than relying solely on predictor resolution. Coarse-resolution predictors (e.g., NDVI, climate variables) are used to characterize background environmental suitability, whereas the fine-scale spatial expression of RTS expansion is primarily governed by the constrained spatial allocation module.

Higher resolution meteorological data are often not available for the entire study area, and cannot be reliably downscaled without introducing substantial uncertainty. With respect to vegetation data, although higher-resolution Landsat 8 NDVI (30 m) is available, it suffers from temporal limitations in the study region, including frequent cloud contamination, irregular acquisition intervals, and incomplete annual coverage. In contrast, the MODIS NDVI product (250 m) provides temporally consistent, annually complete observations that are better aligned with the annual RTS inventory. For this reason, MODIS NDVI was retained as the primary vegetation predictor in the main experiment.

To explicitly quantify the impact of predictor variable resolution, we conducted an additional supplementary experiment in which the MODIS NDVI was replaced by a reconstructed higher-resolution Landsat 8 NDVI dataset (missing or cloud-affected areas interpolated using MODIS NDVI). The full RTSEvo workflow was re-run using this alternative input. Validation against independently mapped RTS expansion in 2021 shows that higher-resolution NDVI leads to a modest but consistent improvement in simulation accuracy: the Figure of Merit increased from 12.00% to 12.53% for the Logistic Regression Evolution Model and from 10.77% to 11.93% for the Random Forest Evolution Model (Figure S1). The simulated RTS spatial distributions are shown in the Figure S2.

These results indicate that finer-scale vegetation information does enhance the model's ability to resolve local spatial contrasts in RTS occurrence probability. However, the relatively small magnitude of improvement demonstrates that the use of 250 m NDVI does not substantially constrain overall model performance.

We therefore conclude that although scale mismatch remains an inherent limitation of regional RTS evolution modeling, its impact on RTSEvo

simulations is partly mitigated by the model's hybrid structure and is quantitatively shown to be limited with respect to NDVI resolution. We have revised the Discussion section to clearly articulate this limitation and present the supplementary analysis as a sensitivity test. Future work will focus on integrating multi-scale remote sensing products and improved high-resolution climate forcing as such datasets become available.

The above discussion and analysis is included in the Discussion (Section 4.3).

Figures S1 and S2 is added to the supplementary file.

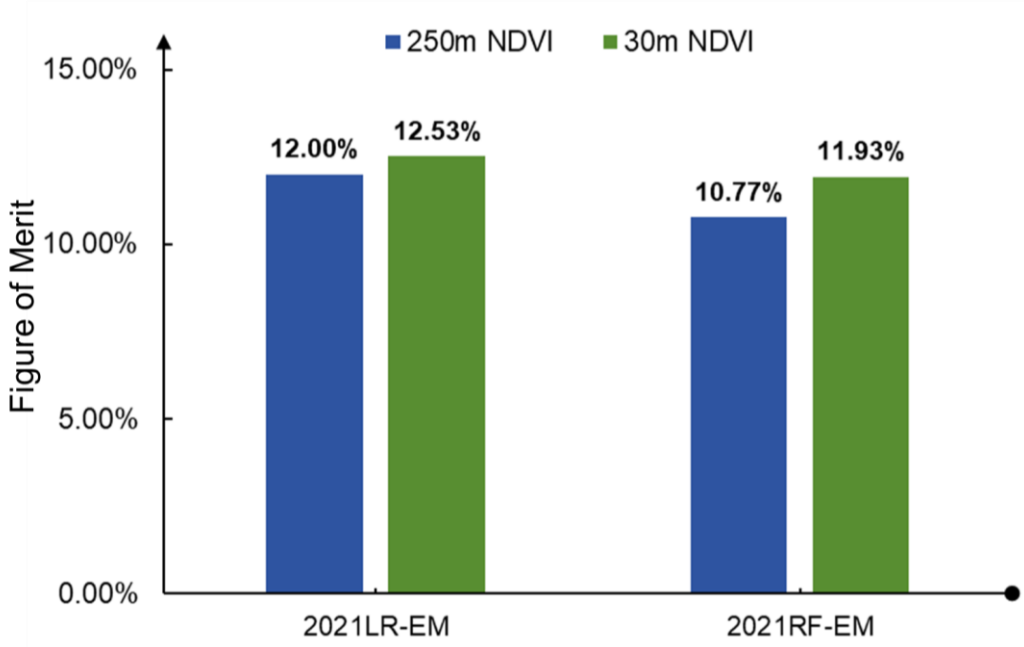

Figure S1. The impact of different NDVI resolutions on model performance.

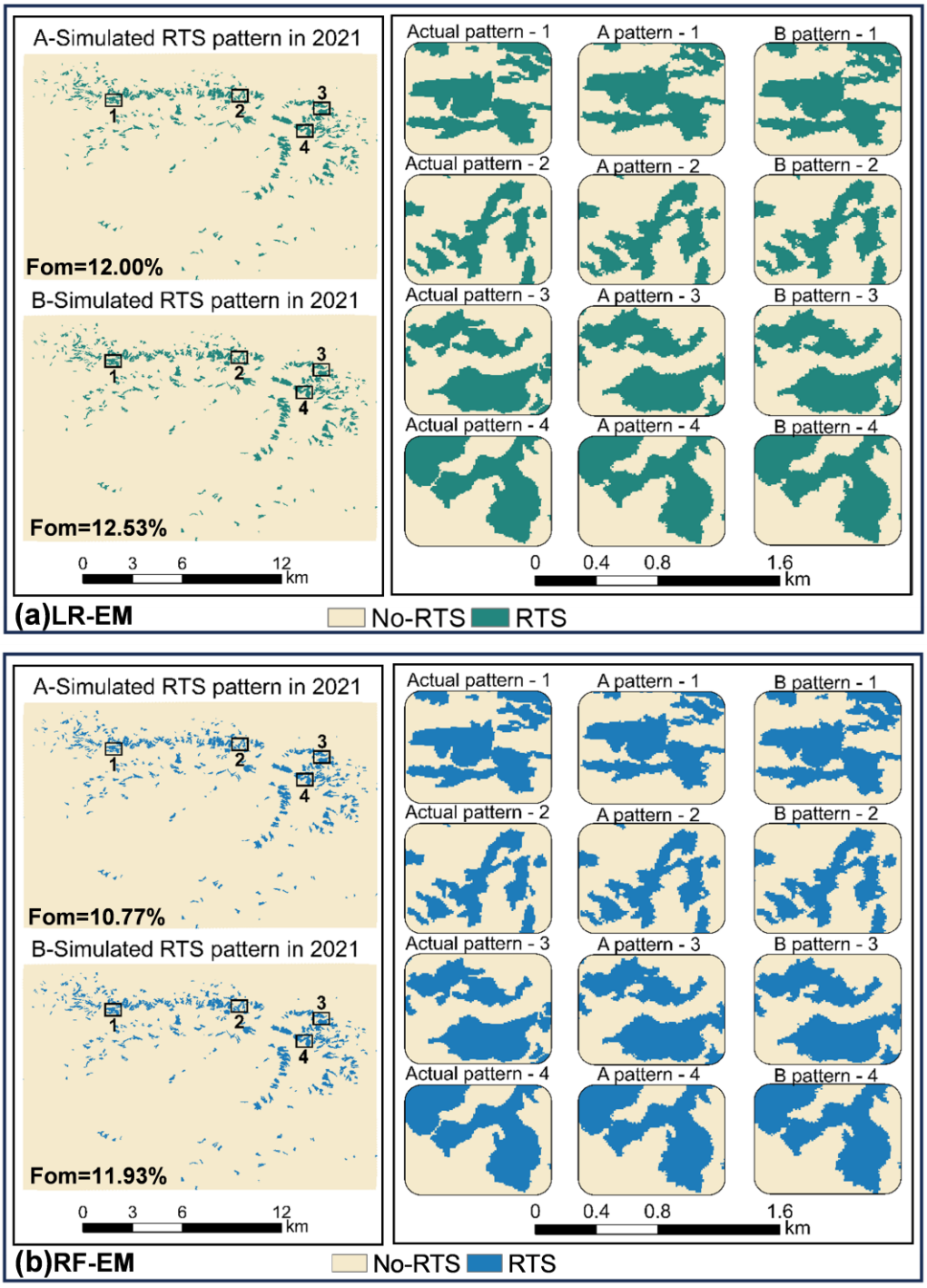

Figure S2. Comparison of RTS spatial distribution at different NDVI resolutions. (a) LR-EM in 2021, (b) RF-EM in 2021. The inset panels provide a detailed visual comparison between the observed (“Actual pattern”) and simulated ("A pattern" represents the simulation results using 250m NDVI, and "B pattern" represents the simulation results using 30m NDVI.) slump morphology for four representative sub-regions (numbering 1-4), randomly selected from areas with intense RTS expansion.

2. Model Transferability and Generalizability The study constructs the RTSEvo model by combining physical process characterization with machine learning parameter calibration. While the model was validated using over 450 RTS inventory records with positive results, the spatial distribution of these samples is highly concentrated within the Beiluhe Basin. This spatial clustering implies that the climatic conditions, which are key drivers used to predict RTS state and evolution，are highly homogeneous across the training and validation datasets. Therefore, it is questionable whether the parameters calibrated in this specific region are transferable to other permafrost regions with different environmental settings. The authors must strictly clarify the model's transferability and discuss this potential lack of generalizability in the text.

Response: Thank you. Indeed, the RTS samples used for calibration and validation in this study are spatially concentrated within the Beiluhe Basin, where environmental conditions, particularly climate forcing and permafrost characteristics are homogeneous. This may limits the direct transferability of region-specific calibrated parameters derived from this case study.

In the revision, we explicitly state that the current parameter set is optimized for the Beiluhe Basin. When implementing the model in permafrost regions with different climate patterns and subsurface ice conditions, we recommend additional regional calibration in order to achieve best performance.

In addition, to quantitatively assess model generalization, we conducted an independent transferability experiment in an additional permafrost region where RTS activity has been observed (location shown as the “Test area” in Figure 2. The test area’s landscape is dominated by alpine meadows and alpine grasslands. The number of RTS increased from 52 to 92 between 2016 and 2022. Compared with the Beiluhe Basin (Figure A2), the RTS in the Test area is distributed in areas with lower DEM and NDVI, and the silt content at 30-60cm and TDD are also higher (Figure S3).

In this experiment, we ran the RTSEvo framework under two scenarios: (1) using parameters calibrated exclusively from the Beiluhe Basin, and (2) using parameters recalibrated locally based on the RTS inventory of the test area.

The RTS and driving data used in the test area are consistent with the data sources used in the Beiluhe Basin.

Taking the LR-EM method as an example, when Beiluhe-calibrated parameters were directly transferred to the test area, the FoM values for 2020, 2021, and 2022 were 15.81%, 5.27%, and 6.10%, respectively. After recalibrating parameters using the actual RTS distribution in the test area (based on 2020 data), the FoM values increased to 18.15% and 6.25% for 2020 and 2021, while showing a modest decrease to 4.29% in 2022 (Figure S4). These results demonstrate that although the model framework itself is transferable and remains functional across regions, local calibration leads to improved performance.

This new analysis is incorporated into the Discussion (Section 4.4).

Figures S3 and S4 is added to the supplementary file.

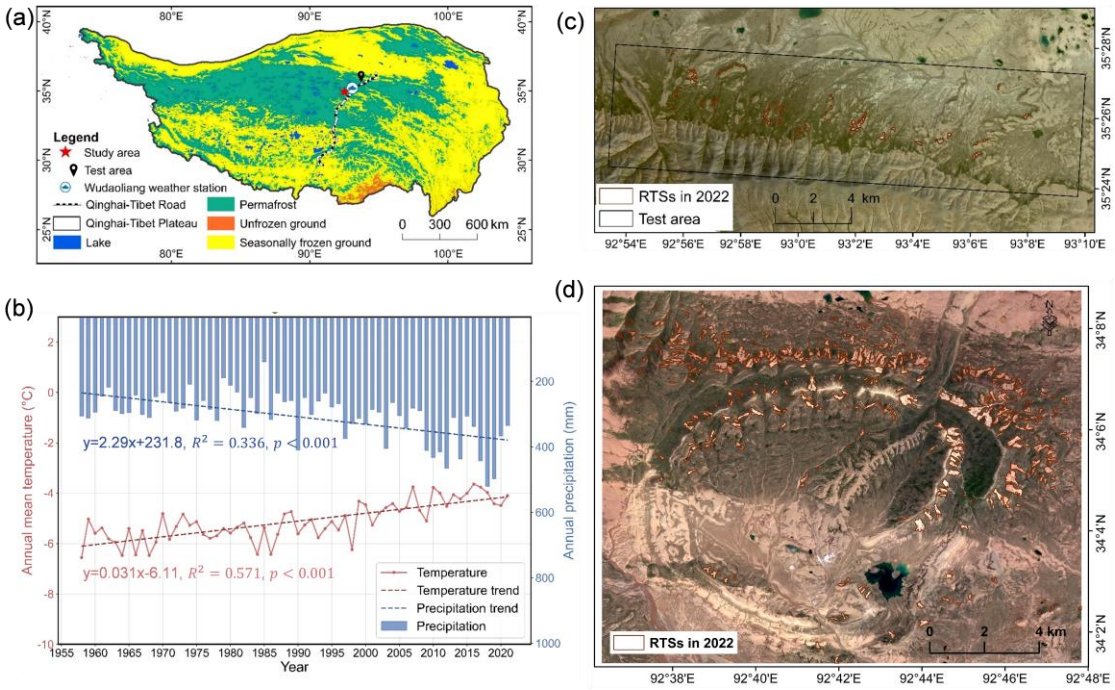

Figure 2. Location of the Beiluhe Basin study area, test area, regional climate trends, and RTS distribution. (a) Permafrost distribution on the QTP, with the study area marked by a red star. Permafrost data are from Cao et al. (2023). (b) Air temperature and precipitation records from the nearby Wudaoliang weather station (1955-2021), with trend lines illustrating regional warming and humidification. (c-d) 2022 PlanetScope satellite images (Planet Team, 2025) of the test area and the Beiluhe Basin overlaid with mapped RTS boundaries are from Xia et al. (2024).

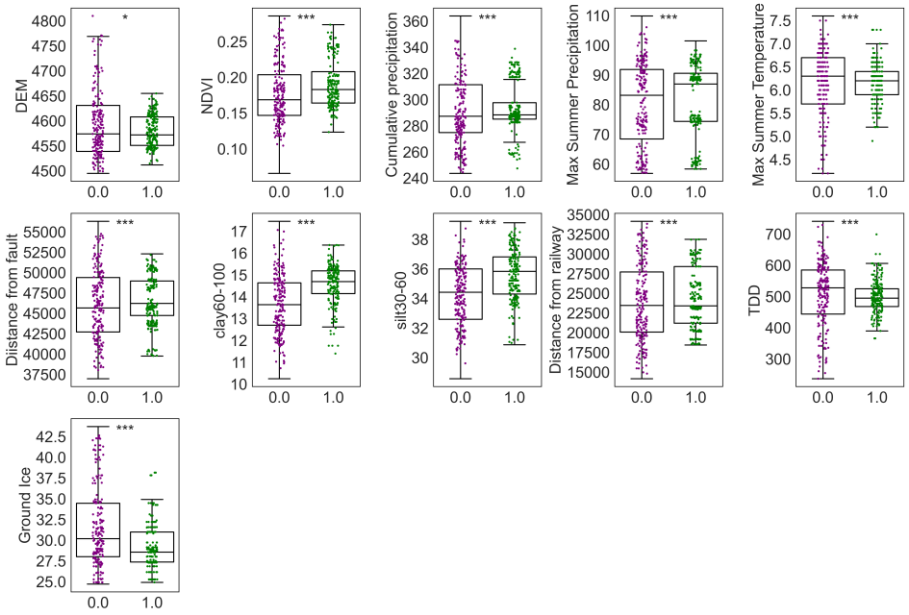

Figure S3. Distribution of numerical predictor variables for RTS versus non-RTS locations. The box-and-whisker plots compare the distributions for zones where RTSs occurred (RTS=1) and where they did not (RTS=0). Significance levels between groups, determined by the Mann-Whitney U test, are indicated as follows: ns, not significant; *p < 0.05; **p < 0.01; ***p < 0.001.

Figure S4. Comparison of simulated (based on LR-EM) and observed RTS

spatial distributions in the test area for 2020–2022. (a) Observed RTS spatial distributions. (b) Beiluhe-calibrated parameters were used to simulate RTS spatial distributions in the test area. (c) Simulation of RTS spatial distributions using parameters calibrated in the test area.

3. Codes: Geoscientific Model Development (GMD) mandates that code be open-source and sufficiently documented to facilitate reuse by the community (and for the purpose of peer review). Upon reviewing the provided repository. I found that the current documentation is non-standard and lacks rigor. Notably, the code contains comments in languages other than English (e.g., Chinese characters are visible in comments. I strongly suggest the authors perform a thorough revision of the code and documentation to ensure it meets international standards and undergoes rigorous testing before final publication.
Response: Thanks. We conducted a thorough revision of the code repository to meet these standards. Specifically, all non-English comments (including Chinese annotations) will be removed and rewritten in clear English, and the overall code documentation will be improved.

The source code for the thaw slump evolution model is publicly available on GitHub (https://github.com/nanzt/RTSEvo) and the exact version used to generate the results presented here is archived on Zenodo (https://doi.org/10.5281/zenodo.17850641)

4. Figure A2: Regarding the final plot in Figure A2 (Active Layer Thickness): Why does the distribution exhibit a bimodal pattern with values clustered at two extremes (showing a difference of up to 1 m)? Please clarify the physical or data-processing reason for this distinct separation.

Response：Thanks. The bimodal distribution of Active Layer Thickness (ALT) in Figure A2 is real and physically meaningful, rather than an artifact.

The bimodal distribution of ALT shown in Figure A2 reflects the presence of two dominant ground thermal regimes in the Beiluhe Basin. Continuous vegetation and organic layers maintain a shallow active layer through strong thermal insulation, whereas RTS-affected surfaces experience abrupt deepening of the active layer due to vegetation loss, reduced insulation, and enhanced soil heat flux (Wang et al., 2018; Yi et al., 2018). Once surface insulation is sufficiently disturbed, ALT increases rapidly rather than gradually,

resulting in a threshold-like response. Consequently, intermediate ALT values are spatially rare, producing a distinct gap between the two modes. This pattern indicates regime separation rather than gradual spatial variability of active layer thickness. ALT observation data (Zhao et al., 2021) from adjacent stations along the Qinghai-Tibet Railway also showed significant differences ranging from 120 cm to 300 cm, which supports the existence of this phenomenon (Figure S5).

The following text is added to the revised manuscript:

“In particular, the Active Layer Thickness (ALT) exhibits a clear bimodal distribution. The bimodal distribution of ALT reflects the presence of two dominant ground thermal regimes in the Beiluhe Basin. Continuous vegetation and organic layers maintain a shallow active layer through strong thermal insulation, whereas RTS-affected surfaces experience abrupt deepening of the active layer due to vegetation loss, reduced insulation, and enhanced soil heat flux (Wang et al., 2018; Yi et al., 2018). ALT observation data (Zhao et al., 2021) from adjacent stations along the Qinghai-Tibet Railway also showed significant differences ranging from 120 cm to 300 cm.”

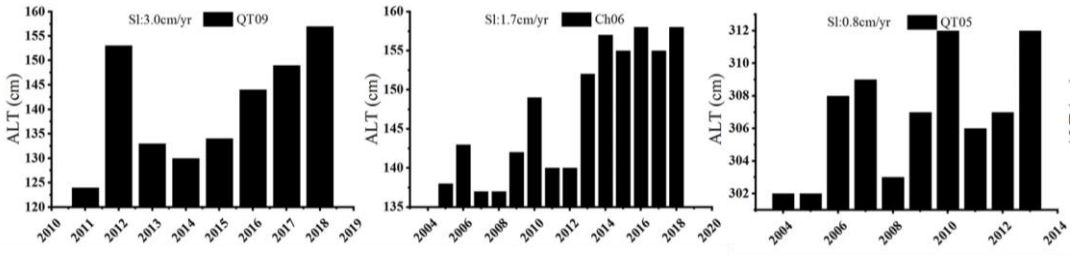

Figure S5. Active layer thickness at different adjacent stations along the Qinghai-Tibet highway (Zhao et al., 2021).

References

Cao, Z., Nan, Z., Hu, J., Chen, Y., and Zhang, Y.: A new 2010 permafrost distribution map over the Qinghai–Tibet Plateau based on subregion survey maps: a benchmark for regional permafrost modeling, Earth Syst. Sci. Data, 15, 3905-3930, https://doi.org/10.5194/essd-15-3905-2023, 2023.

Planet Team: Planet application program interface: In space for life on earth [dataset], Retrieved from https://api.planet.com, 2025.

Wang, Y., Sun, Z., and Sun, Y.: Effects of a thaw slump on active layer in

permafrost regions with the comparison of effects of thermokarst lakes on the Qinghai–Tibet Plateau, China, Geoderma, 314, 47-57, https://doi.org/10.1016/j.geoderma.2017.10.046, 2018.

Xia, Z., Liu, L., Mu, C., Peng, X., Zhao, Z., Huang, L., Luo, J., and Fan, C.: Annual inventories of retrogressive thaw slumps across the Qinghai-Tibet Plateau from 2016 to 2022, Zenodo [dataset], https://doi.org/10.5281/zenodo.10928346, 2024.

Yi, Y., Kimball, J. S., Chen, R. H., Moghaddam, M., Reichle, R. H., Mishra, U., Zona, D., and Oechel, W. C.: Characterizing permafrost active layer dynamics and sensitivity to landscape spatial heterogeneity in Alaska, Cryosphere, 12, 145-161, https://doi.org/10.5194/tc-12-145-2018, 2018.

Zhao, L., Zou, D., Hu, G., Wu, T., Du, E., Liu, G., Xiao, Y., Li, R., Pang, Q., Qiao, Y., Wu, X., Sun, Z., Xing, Z., Sheng, Y., Zhao, Y., Shi, J., Xie, C., Wang, L., Wang, C., and Cheng, G.: A synthesis dataset of permafrost thermal state for the Qinghai–Tibet (Xizang) Plateau, China, Earth Syst. Sci. Data, 13, 4207-4218, https://doi.org/10.5194/tc-12-145-2018/10.5194/essd-13-4207-2021, 2021.