Response to CC1

We sincerely thank the reviewer for dedicating time and effort devoted to reviewing our manuscript. We greatly appreciate the positive assessment of our approach and the practical value of the resulting dataset. The constructive comments have helped us improve the clarity, scientific rigor, and overall flow of the manuscript. In particular, we have added statistical tests for the environmental driver analysis, restructured parts of the Results and Discussion for better logical flow, expanded the explanation of the space-for-time substitution strategy, and refined the terminology and methodological descriptions. All comments have been carefully addressed in the revised manuscript.

Below, the reviewer's original comments are shown in **blue**, our responses are given in **black**, and the corresponding revisions added to the main text are highlighted in **red**.

General comments

I have reviewed EGUSPHERE-2026-345 by Chen et al. This paper has mapped permafrost distribution in 2020 over Tibetan plateau, using new strategy of a space-for-time substitution, using neural network and machine learning approaches. The map and estimated permafrost change since 2010 well explained ecohydrological changes in high resolution. It also explained well the limitation of applying neural network scheme for permafrost mapping. Accordingly I would be very positive to publish this in TC after following minor revisions.

Response: We sincerely thank the reviewer for the encouraging and constructive assessment. In this revised manuscript, we have added Welch's two-sample *t*-tests to quantify differences between the positive and negative response groups in the environmental driver analysis, restructured the manuscript by moving the soil parameter E variations and permafrost degradation mechanism analysis from the Results into the Discussion section, expanded the discussion on the applicability and limitations of the space-for-time substitution strategy, polished the language throughout, and updated relevant literature citations. Our point-by-point responses are provided below.

Specific comments

174-175: any reference for this threshold?

Response: Thank you for this comment. We have added the reference (Hu et al., 2020) and expanded the text to provide the physical and mathematical derivation of the $F = 0.5$ threshold.

In the revised manuscript we have added the following text:

"The parameter E is a dimensionless empirical factor that accounts for the local environmental modification of the ground thermal regime. Theoretically, E reflects the combined effects of soil thermal properties and moisture conditions in both frozen and thawed states. Based on this formulation, the threshold for permafrost occurrence is defined at $F > 0.5$. Pixels with $F \leq 0.5$ are classified as SFG or non-frozen ground (Hu et al., 2020). The threshold $F = 0.5$ is physically meaningful because it corresponds to the case where the maximum thawing depth equals the maximum freezing depth. When $F > 0.5$, the thaw depth is smaller than the freeze depth, indicating permafrost; when $F \leq 0.5$, the thaw depth exceeds the freeze depth, indicating SFG."

Reference added:

Hu, J., Zhao, S., Nan, Z., Wu, X., Sun, X., & Cheng, G. (2020). An effective approach for mapping permafrost in a large area using subregion maps and satellite data. Permafrost and Periglacial Processes, 31(4), 548-560. https://doi.org/10.1002/ppp.2068

181-184: As for hypothesis of E determination, soil moisture would be changed over the decadal scale, and statistical relation between E and environmental drivers might alter.

Response: We thank the reviewer for raising this fundamental point. We fully agree that soil moisture conditions can and do change over a decade, and that may in principle alter the statistical relationship between E and its environmental drivers.

The space-for-time substitution strategy was adopted precisely because we lack concurrent 2020 field surveys to recalibrate E directly, a limitation common to many regional-scale modeling studies. Our core hypothesis is that, while the absolute value of E at any given location may shift with changing moisture and surface conditions, the underlying statistical relationship between E and a suite of environmental covariates (topography, soil texture, and dynamic surface conditions) remains sufficiently stable over a decadal interval to allow meaningful extrapolation.

To implement this, we trained the Random Forest model on the high-quality 2010 E field (derived from extensive field surveys) and then drove the model with the actual observed 2016–2020 values of the dynamic predictors (precipitation as a proxy for regional soil moisture, NDVI, and fractional snow cover). In this way, the predicted 2020-period E field is allowed to respond to the decadal changes in moisture and surface conditions that actually occurred, rather than simply copying the

2010 pattern.

We acknowledge, however, that the assumption of a stationary driver–*E* relationship is an inherent limitation of the space-for-time approach. This assumption is more likely to hold over relatively short decadal timescales in relatively undisturbed alpine environments, but it may weaken or break down under abrupt disturbances (e.g., thermokarst development or intensive engineering activity) or when extrapolated over longer time periods or to strongly different climatic regimes. We have therefore taken several measures to reduce and transparently discuss this uncertainty:

- Use of representative spatial sampling (*k*-means clustering) across diverse environmental gradients to improve robustness, since random sampling tends to oversample the plateau's homogeneous interior at the expense of rare but distinct environmental settings where the driver-*E* relationship is most likely sensitive to decadal change;
- Inclusion of dynamic predictors so that *E* can adjust to observed decadal changes;
- Examination of the physical driver behind this shift in *E*: the decadal change in *E* tracks precipitation (used as a proxy for soil moisture), the variable most directly tied to ground thermal inertia, drier regions show a rising *E* and wetter regions a falling *E* (Fig. 10), supporting that the model is capturing a genuine moisture-driven signal rather than an extrapolation artifact;
- Testing of model generalization within the 2010 period itself, where ground-truth *E* values are available. The RF-based model achieved high accuracy and outperformed a more advanced MLP model, and also remained robust across multiple independent training-data sampling strategies. These results indicate that the learned relationship generalizes well.
- Explicit discussion in the revised manuscript of the conditions under which the space-for-time assumption is expected to be reliable versus where it may fail.

We believe these steps, together with the independent borehole validation and inter-comparison with existing 2020-period maps, support the reliability of the 2020-period permafrost map.

184: a space-for-time substitution strategy, needs for more detail explanation. It is hard to understand only from Fig.2(c).

Response: We sincerely thank the reviewer for this valuable comment. We have expanded the textual description of the space-for-time substitution strategy in Section 3.1 and added a clear reference to the detailed explanation in Section 3.3.

In the revised manuscript we have added the following text:

"To overcome this limitation, this study introduces a practical adaptation. We hypothesize that while the specific value of *E* at a given location may shift due to changes in moisture or surface cover, the fundamental statistical relationship between *E* and its environmental drivers (such as topography, vegetation, soil texture, soil moisture) remains applicable over the decadal scale. Consequently, we employ a space-for-time substitution strategy to predict the 2020 E field based on the relationships learned from the high-quality 2010 dataset (Fig. 2). More specifically, we used the high-quality 2010 E distribution map produced by Cao et al. (2023) and applied two transfer learning strategies, a neural network approach and a random forest approach, to establish the relationships between *E* and multiple environmental drivers. By updating the environmental drivers to their 2020 values, we derived the predicted *E* values for 2020 based on the learned statistical relationships. Further details on the estimation of the soil parameter E for the 2020 period are provided in Section 3.3."

209: Any other environmental factors for multilinear regression? Any correlation between NDVI and latitude?

Response: Many thanks for this comment. Prior to finalizing the correction model, we tested additional predictors, including DEM, albedo, and longitude, and compared multiple modeling approaches, including nonlinear regression and machine-learning methods (e.g., Random Forest). The combination of LST-derived DDT, NDVI, and latitude already explained the majority of the variance in the observed ground surface thawing indices (DDT). Adding further variables or increasing model complexity resulted in only marginal improvements.

Physically, LST-derived DDT serves as the primary thermal baseline and implicitly incorporates macro-scale temperature gradients associated with elevation. The remaining offset between satellite LST and 0 cm GST is largely governed by local surface energy balance processes. Within this framework, NDVI captures vegetation-mediated effects on albedo and latent heat partitioning, while latitude serves as a reliable proxy for regional variations in solar incidence angles and radiation budgets across the Plateau.

To evaluate potential multicollinearity between NDVI and latitude we calculated the Variance Inflation Factor (VIF = 1.26), which is well below the conservative threshold of 5.0, confirming that the two variables can be included simultaneously without introducing substantial collinearity.

We have added the following clarification in the Discussion section:

"The freezing and thawing indices likewise have to be adapted to the local climate. In our case, LST-derived DDT was corrected to actual DDT with a multiple linear regression on LST-derived DDT, NDVI and latitude. LST-derived DDT serves as the primary thermal indicator and implicitly carries large-scale temperature gradients associated with elevation and regional climate. Before settling on this model, we also tested additional predictors (DEM, albedo, and longitude) and nonlinear and machine-learning alternatives (e.g., random forests); DDT_LST, NDVI and latitude already explained most of the variance, and adding variables or model complexity did not meaningfully improve accuracy. The applicability of this correction approach to other regions should be verified through independent validation. Finally, calculating the DDF directly from satellite LST is justified on the QTP, where winter snow cover is thin and short-lived; in Arctic or Russian permafrost regions, however, the insulating effect of thick, persistent snow would need to be parameterized explicitly in the DDF estimate."

322: Is usage of the term 'risk' appropriate?

Response: We sincerely thank the reviewer for this rigorous reminder. According to the IPCC definition, risk depends not only on the hazard itself but also on exposure and vulnerability. In this study, an increase in parameter E primarily reflects enhanced conditions conducive to permafrost degradation; therefore, the term "risk" lacks scientific rigor in this context. We have replaced "degradation risk" with "degradation potential" or "degradation trend" throughout the revised manuscript.

Fig.5: Can you say that these boxplots differ between Negative and Positive? Any statistics?

Response: We thank the reviewer for this valuable comment. To quantitatively assess the differences between the Negative and Positive response groups, we applied Welch's two-sample t-test, which is suitable for large sample sizes without assuming equal variances together with Cohen's d to evaluate effect size. The difference in parameter E between the two groups is highly significant ($P < 0.01$, and Cohen's $d = -3.83$), far exceeding the conventional threshold (0.2) for a large effect. For NDVI and FSC, the P-values are both less than 0.01, but the effect sizes are negligible (Cohen's d 0.010 and 0.005, respectively). For precipitation, the effect size (Cohen's $d = 0.02$) is several times larger than for NDVI or FSC (Figure R1), indicating that precipitation exhibits a relatively stronger influence on the divergent trajectories of the two groups under the background of general warming and wetting.

We have updated the text and the figure caption (Figure 10 in the revised manuscript) accordingly.

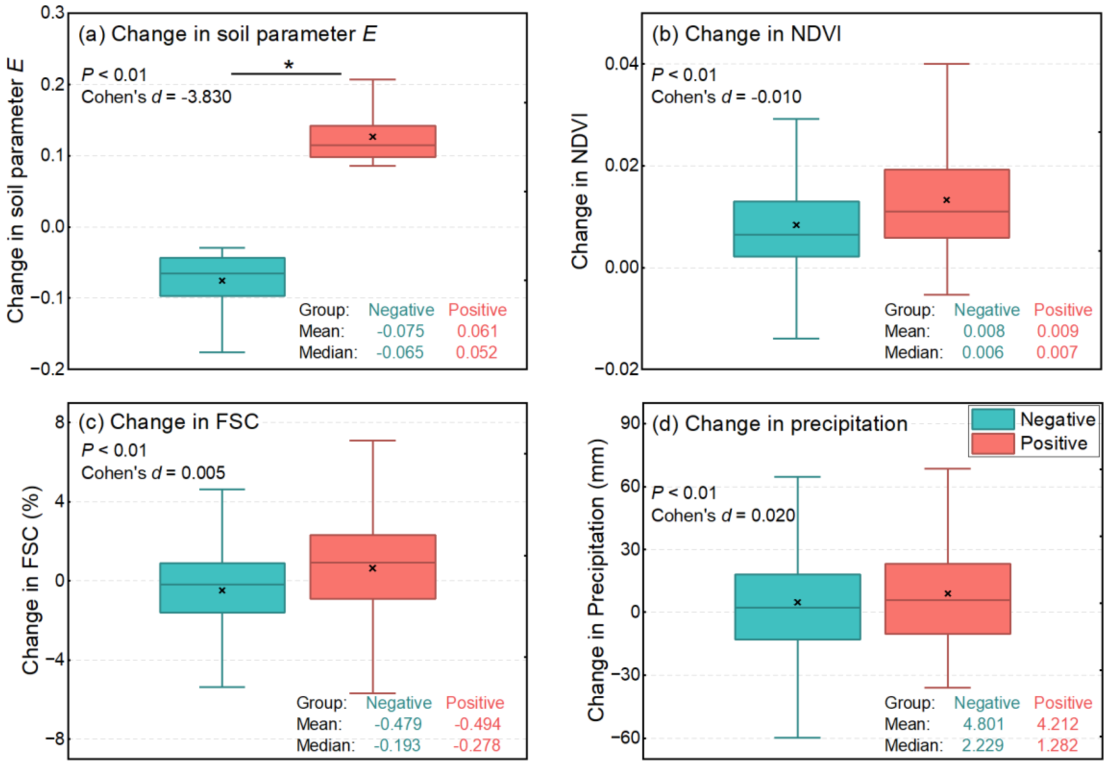

Figure R1 Distribution of decadal changes (2010-2020) in key environmental factors, categorized by the trend in soil parameter *E*. Boxplots compare changes in (a) soil parameter *E*, (b) NDVI, (c) FSC and (d) precipitation between regions of increased potential (positive group) and reduced potential (negative group). In each box plot, the center line shows the median, the box represents the lower and upper quartiles (25th-75th percentiles), and the whiskers extend to the furthest data points within 1.5 times the interquartile range. The 'x' marks the mean. Asterisks indicate significant differences between the two groups (Welch Two Sample *t*-test, $P < 0.01$ and $|\text{Cohen's } d| \geq 0.2$).

334-335: How we see Fig 5 to confirm this description?

Response:

We apologize for the error in the original manuscript. The FSC actually exhibited a decreasing trend rather than an increase. This has been corrected in the revised text and the figure has been renumbered (now Figure 10). The corrected sentence reads:

"To elucidate the physical mechanisms driving changes in soil parameter E, we analyzed the temporal variations of the environmental predictors (Fig. 10). While NDVI increased in both groups (Fig. 10b) and FSC generally decreased (Fig. 10c), precipitation emerged as the distinct driver of divergence (Fig. 10d). This is supported by the statistical analysis, which showed a stronger significance for precipitation ($P < 0.01$, Cohen's d = 0.020) than for NDVI and FSC. It is important to note that due to the lack of high-resolution soil moisture products for the region, mean annual precipitation is utilized here as a proxy for regional soil moisture conditions"

Table 2: Include citations for MAGT and TTOP-based maps. Can you show borehole location in any of the figures?

Responses: Thank you for this helpful comment. We have added the corresponding references for the MAGT-based and TTOP-based maps. In addition, the locations of all 109 validation boreholes (83 permafrost and 26 seasonally frozen ground sites) have been added to Figure 1, with the Three-River Headwater Region (TRHR) and surrounding headwater areas now clearly indicated.

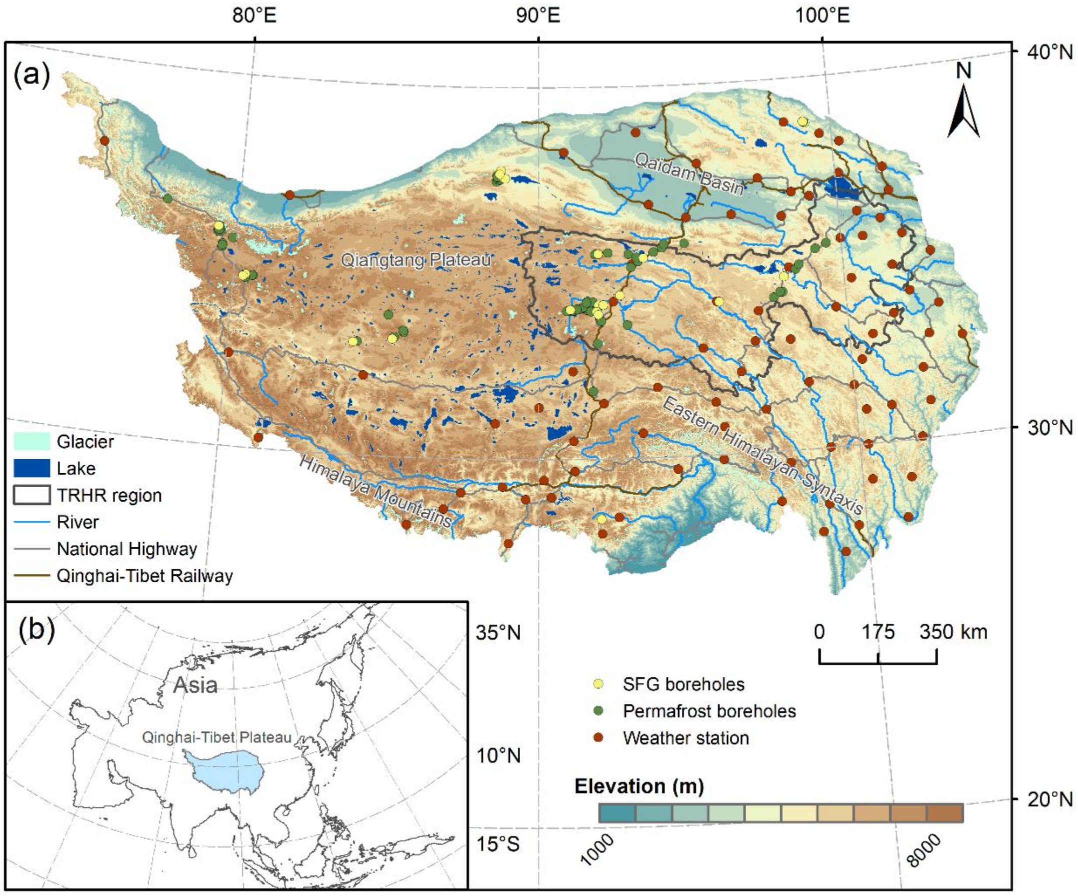

Figure R2 Topographic map and geographical context of the Qinghai-Tibet Plateau (QTP). (a) The topography of the region. The background color gradient represents elevation derived from the Shuttle Radar Topography Mission DEM. The 87 national meteorological stations (red dots) used for ground surface temperature correction, alongside major river systems and lakes are shown. Notably, the 83 permafrost boreholes and 26 seasonally frozen ground (SFG) boreholes utilized in this study are largely concentrated in the Three-River Headwater Region (TRHR) and surrounding headwater areas. (b) The location of the QTP in Asia.

Section 4.6: better to move to Discussions

Response: Many thanks for this structural suggestion. We have reorganized the manuscript: the original Section 4.6 (Mechanisms driving permafrost degradation) has been integrated and expanded within the Discussion section to improve overall flow.

333-340: Better to move to Discussions

Response: Accepted. The relevant analysis has been moved to the Discussion section as part of the broader reorganization described above.