Comments on egusphere-2026-2202

General Comments

This study focuses on the lagged responses of seasonally frozen ground to extreme heat events over the Qinghai-Tibet Plateau, which is inherently an important topic. However, there are several critical issues regarding the study design, data processing, methodological details, and the interpretation of physical mechanisms. These issues make the manuscript very difficult to review. A possible reason is that the paper piles up numerous results, but the storyline is unclear (or perhaps due to my own lack of understanding). I believe that, in its current state, the manuscript has too many issues need to be substantially revised (may not be done within a short timeframe). Please refer to my specific comments below.

Major Comments

- As issues raised by Dr. Zhu, the performance of the model itself is a fatal flaw. A bias of nearly 40 cm is quite an exaggerated deviation, and an $R = 0.77$ implies that the model can only explain about 60% of the variance in the target variable ($R^2 \approx 0.6$). I am not sure if this predictive accuracy is sufficient for analysis. Could this be an issue with the performance of the LightGBM model itself? Were other machine learning models tested? As is well known, these machine learning models do not inherently account for temporal correlations, can they be directly linked to the preceding DLNM model? These methodological issues make it difficult for me to comprehend this study. Furthermore, there might be issues with variable selection in the machine learning model. For example, there is a deterministic relationship among the contents of Sand, Silt, and Clay. Does the precipitation variable encompass snow water equivalent?

- Unclear methods.

  - In Section 2.2.1, the authors mention that when the freezing front extends below the deepest observed layer (2.0 m), a linear extrapolation is performed using the temperature gradient between the deepest layers, allowing extrapolation up to a maximum of 5 m. This is a highly risky approach because soil temperature gradients are typically non-linear. Additionally, how were missing values in the observational data handled? The authors mention using "depth-corrected linear interpolation" and "controlled linear extrapolation" methods, but do not explain how these are specifically implemented, despite their crucial importance. Moreover, the soil temperature observations from the China Meteorological Administration can reach depths of 3.2 m; I completely fail to understand how the authors processed this data here.
  - The identification of extreme heat events in Section 2.2.2 is also confusing. Line 155 states that the 90th percentile (P90) is used to identify heat events, but the subsequent paragraph move to "typical" high-temperature values of 10, 15, ..., 30°C. Are these thresholds applied universally across all regions? Are these thresholds reasonable for the seasonally frozen ground on the Qinghai-Tibet Plateau? These issues make it difficult to properly evaluate the methodology.
  - In Section 2.2.3, when introducing the LightGBM+SHAP framework, the authors specify the input variables (number of extreme heat days, soil moisture, etc.) and the output variable (maximum lag time). The DLNM model outputs a three-dimensional lag-response. For each spatial pixel on the Qinghai-Tibet Plateau, how is a single maximum lag time scalar quantitatively extracted from the continuous lag curve of the DLNM to serve as the output label for the LightGBM model? While my understanding might not be entirely accurate, these details are, at the very least, highly opaque to the reader.

- The entire analysis in Section 3 reads more like a report than a research paper. It sequentially introduces spatiotemporal changes, freeze-thaw processes, and model uncertainties, and only touches upon the core scientific questions in Section 3.4, which unfortunately lacks depth. The presentation of the results also faces issues of poor clarity. It appears the authors constructed an extremely complex analytical framework only to derive an overly simplified set of results. The text is filled with simple quantitative statements (e.g., "the TMD is advanced by an average of 3.28 days, whereas the FSD is delayed by approximately 1.14 days"). How were these numbers derived? Are they

averages across the entire seasonally frozen ground region of the Qinghai-Tibet Plateau? Regarding the statement "each additional extreme heat day reduces MFD by approximately 1.12 cm on average," what is the intensity of the extreme heat events? There are numerous similar ambiguities throughout the manuscript.

Specific Comments

- **L390**: "This behavior is likely associated with the thermal saturation mechanism of SFG (Gouttevin et al., 2012): moderate warming tends to enhance land–atmosphere thermal gradients and prolong lagged responses, whereas excessively strong warming may promote more rapid subsurface thermal adjustment, leading to shorter lag duration." Could explain a bit more clearly?
- **Table 1:** Why were both CRA and ERA datasets utilized? Does the CRA dataset lack temperature, precipitation, and solar radiation? How was data with different spatial resolutions harmonized? What is the final spatial resolution used in this study?
- **Table 2:** This table appears highly unstandardized, displaying six indicators. I have never seen a table formatted this way.
- **Figure 6:** The figure looks like a simple patchwork. What is a "Linear predictor"? What are its units? Furthermore, the caption is completely uninformative and overly simplistic.
- *(There are many other issues not pointed out here.)*