This study by Jafarov et al. aims to show the capability of a calibration method to recover the synthetic parameter. The motivation is that the proposed method can reduce the labor and time of manual parameter calibration.

I have fundamental concerns regarding the manuscript.

1. The study calibrates parameters using synthetic data, which is generated from the same model being calibrated. Even the proposed method can successfully recover the set of parameters, I don't see the value in this analysis for advancing our understanding on permafrost ecosystem dynamics. If the model has structural errors that misrepresent a process, calibrating based on the biased or errored simulation would be meaningless. However, if the authors can demonstrate that the method can perform well on 'ground truth' data derived from in-situ measurements and remote sensing, I would think it makes more valuable. But according to the comparison between observed and calibrated results (Figure 5), the performance seems not ideal. Overall, I think this study is going to a wrong direction.
2. The authors employ a hierarchical approach to parameter calibration, but the rationale for the chosen order (e.g., calibrating vegetation parameters before soil parameters) is not clearly justified. Would reversing the calibration sequence affect the results? How sensitive is the method to the order of parameter calibration? Addressing these questions is critical for evaluating the robustness and generalizability of the proposed approach.
3. The manuscript acknowledges the challenge of equifinality but does not present a clear strategy to address it. The results indicate that the choice of initial parameter values has a greater impact on calibration outcomes when using observational data compared to synthetic data. This suggests that the proposed approach may be less reliable when applied to real-world ecosystems, where uncertainties and biases in observations are unavoidable. The study would benefit from a more rigorous discussion of how to mitigate equifinality and improve the method's performance on real-world datasets.

Considering all the above-mentioned aspects, I would not recommend this manuscript for publication in its current form.