Potential Risk Analysis and Susceptibility Modeling of Thermokarst Hazard in Permafrost Landscapes of the Arctic Based on Interpretable Ensemble Learning

Yuting Yang and Xi Mao (China, PR)

Key words: Remote sensing; Risk management; Arctic； thermokarst hazard; susceptibility evaluation; interpretability; potential risk analysis; ensemble learning

SUMMARY

Climate change is degrading Arctic permafrost, leading to thermokarst hazards and significant environmental impacts. Despite their critical ecological role, the distribution and risks associated with thermokarst hazards are not well understood due to the complexity of influencing factors. Our primary objective is to provide scientific support for accurately identifying areas prone to thermokarst hazards.In this study, we focus on the Arctic and utilize Landsat-5, Landsat-7, and Landsat-8 satellite imagery, along with additional spatial data, to investigate the causes and spatial distribution of thermokarst lakes. From this extensive dataset, we have selected ten key evaluation factors. We introduce a new interpretable ensemble learning method to enhance our understanding of thermokarst hazard susceptibility.The primary learning units in our model include the random forest (RF), extremely randomized trees (EXTs), extreme gradient boosting (XGBoost), and categorical boosting (CatBoost) algorithms. Additionally, gradient boosted decision trees (GBDTs) serve as the secondary learning unit. Using a stacking model, we assessed the susceptibility to thermokarst hazards and validated the model's accuracy through six evaluation indices.To ensure the model's interpretability, we employed three methods: accumulated local effects (ALE), local interpretable model-agnostic explanations (LIME), and Shapley additive explanations (SHAP). The results indicate that our ensemble learning stacking model not only performed exceptionally well but also achieved the highest prediction accuracy. It is estimated that high and very high susceptibility regions, covering about 10.4% of the area south of 60°N, amount to 1.8 × 10^6 km².Our findings highlight that slope, elevation, topographic wetness index (TWI), and precipitation are pivotal in assessing the susceptibility to thermokarst hazards. This comprehensive analysis reveals the broader impacts of thermokarst hazards, particularly in regions with significant railway and highway infrastructure, substantial soil organic carbon reserves, and extensive alpine grasslands. The high accuracy of our interpretable ensemble learning model offers significant practical value for project route selection, construction, and operation in the Arctic.
