Spatial Analyses and Susceptibility Modeling of Thermokarst Lakes in Permafrost Landscapes along the Qinghai–Tibet Engineering Corridor

Guoan Yin 1,*, Jing Luo 1, Fujun Niu 1, Fujun Zhou 2, Xianglian Meng 2, Zhanju Lin 1 and Minghao Liu 1

1 State Key Laboratory of Frozen Soil Engineering, Northwest Institute of Eco-Environment and Resources, Chinese Academy of Sciences, Lanzhou 730000, China; yinguoan@lzb.ac.cn (G.Y.); niufujun@lzb.ac.cn (F.N.); zhanjulin@lzb.ac.cn (Z.L.); liuminghao@lzb.ac.cn (M.L.)
2 China Railway First Survey and Design Institute Group Co., LTD., Xi’an 710043, China; dlzfj@fsdi.com.cn (F.Z.); zgmbmxl@fsdi.com.cn (X.M.)
* Correspondence: luojing@lzb.ac.cn

Abstract: Thermokarst lakes (TLs) caused by the thaw of massive ground ice in ice-rich permafrost landscapes are increasing and have strong impacts on the hydro–ecological environment and human infrastructure on the Qinghai–Tibet Plateau (QTP), however, its spatial distribution characteristics and environmental controls have not been underrepresented at the local scale. Here, we analyzed the spatial distribution of small TLs along the Qinghai–Tibet Engineering Corridor (QTEC) based on high-resolution (up to 2.0 m) satellite images. The TLs gathered in the plains and upland plateau and covered 8.3% of the QTEC land. We deployed a random-frost method to investigate the suitable environmental conditions for TLs. Climate including summer rainfall and the air temperature was the most important factor controlling the TL distribution, followed by topography and soil characteristics that affected the ground ice content. TL susceptibility was mapped based on the combinations of climate, soil, and topography grid data. On average, around 20% of the QTEC area was in a high to very-high-susceptibility zone that is likely to develop TLs in response to climate change. This study improved the understanding of controlling factors for TL development but also provided insights into the conditions of massive ground ice and was helpful to assess the impacts of climate change on ecosystem processes and engineering design.

Keywords: geomorphology; environment; remote sensing; machine learning; susceptibility map

1. Introduction

Approximately 15% (ca. 13.6 × 10^6 km^2) of the northern land surface is covered by permafrost, which is defined as ground that remains frozen for two or more consecutive years [1]. Field observations of the ground temperature provide evidence that the permafrost is degrading (warming and thawing) at the global scale in response to climate change [2]. This degradation is expected to continue in the future decades following the predicted climate warming in the high arctic and at high altitudes [3,4]. Permafrost thawing is often accompanied by impacts on water and land surface energy balances [5], changes in landscapes, e.g., [6], and release of old organic carbon [7].

The Qinghai–Tibet Plateau (QTP), referred to by scientists as the Third Pole [8], has a vast expanse (ca. 1.1 × 10^6 km^2) of permafrost [9]. In recent decades, climate warming in most QTP has resulted in the noticeable warming and thawing of permafrost [10–12]. Such permafrost degradation in terms of ground ice thawing and active-layer thickening has been accelerated along the Qinghai–Tibet Highway (QTH) due to the disturbance of engineering activities [13,14]. Permafrost degradation is of particular concern because the thawing of ice-rich permafrost can lead to the development of thermokarst terrain, which is defined as surface subsidence caused by the melting of massive ice [15,16]. Thermokarst significantly impacts the local landscapes, global biogeochemistry, and climate cycle [15,17,18].
but also influences the stability of the local infrastructure. Thermokarst lakes are one common thermokarst landform [16] and are widespread in the permafrost region across the QTP [19,20]. Thermokarst lake can be initiated when the ground subsidence following the thaw of ice-rich permafrost or melt of massive ground ice [21,22]. After a thermokarst lake has formed, the mean annual temperature at the lake bottom is generally higher than 0 °C, which could warm the surrounding permafrost and promotes talik formation [22–24]. In addition, as a major heat source to the surrounding ground, the development of the thermokarst lake may also have implications for the foundation stability of nearby infrastructures [25,26].

The upland plains along the QTH are commonly underlain by ice-rich permafrost [27], which make the ground vulnerable to thermal disturbance and thermokarst development. Permafrost thickness varied from 0 m to more than 200 m across the hilly landscapes and mountain areas [28]. The depth of thermokarst lakes usually ranges from 0 m to 3.0 m [20,23]. Rapid climate warming and human activities make the QTH an interesting and topical area to study the processes of thermokarst lake development. Ongoing climate warming has already led to numerous thermokarst lakes distributed along the QTH [20,24] and their number showed an obvious increase over the last four decades due to permafrost warming and human activities [19]. The developments of such thermokarst lakes have significant effects on the existing engineering facilities and local environment (e.g., ecology, geomorphology, hydrology). Warming over the coming century is expected to continue over the QTP [29]. It is necessary to assess the impacts of climate change on the evolution of permafrost landscapes, carbon release, and engineering constructions.

Recent studies of modern thermokarst lakes on the QTP mainly focused on lake thermal development at a site-specific scale [19,22,23,26]. The information on the current and potential future distribution of the thermokarst lakes and their susceptibility to climate and local environmental factors at a high spatial resolution on the QTP is sparse. To illustrate the distribution characteristics and susceptibility mapping of thermokarst lakes, occurrence can not only provide an important basis for the management of water resources on the QTP but also provide guidelines for the route selection of the planned Qinghai–Tibet express highway or other types of infrastructures in the near future. In this paper, the permafrost region along the QTH from Xidatan to Anduo is selected as the study area. Here, the objectives of this study are: (1) to illustrate the thermokarst lake distribution along the QTEC using the high-resolution satellite images; (2) to quantify, for the first time, the contributions of the environmental variable to the occurrence of thermokarst lakes; (3) to produce the susceptibility map of thermokarst lakes in the study area using a machine learning method.

2. Study Area

The study area, with an area of about 22,351 km², encompassed a 40 km-wide transect along the QTH that extends for about 550 km from Xidatan in the north to Anduo in the south (Figure 1). Due to the convenient transportation after the construction of the QTH, many linear projects, such as the Qinghai–Tibet Railway, Golmud–Lhasa pipeline, and electric transmission line, were constructed along with it. The mean annual air temperature (MAAT) in the study region was between −6.5 °C and −2.0 °C, while the extremely low air temperature was about −30 °C at the end of January and the highest air temperature was approximately 25 °C in July [30]. The annual total precipitation ranges from 250 mm to 450 mm and is concentrated between May to August in the study region [30]. Due to the strong wind and uneven distribution of precipitation, the majority of the study region had a free snow cover in winter [9,30,31]. The vegetation was characterized by alpine meadow and steppe with coverage ranging from 0.3 to 0.9 [32].
The topographic features varied significantly within the study area from north to south. Most parts of the terrain were located above 4500 m asl, with the alternating distribution of mountains, valleys, and basins. Based on different topographic features and surficial geology, we divided the whole study area into six sub-regions (Table 1). These were the Qinghusihe–Chumaerhe (QC) upland plain, the Wudaoliang–Beiluhe (WB) basin, TuoTuohe (TTH) region, Kaixinling (KXL) upland plain, Tanggula (TGL) mountain areas, and Anduo (AD) plain (Figure 1b). All of these regions are underlain by continuous permafrost, which had a mean annual ground temperature (MAGT) ranging from $-3.0 \degree C$ to $0 \degree C$ [11–13].

Figure 1. Map of the study area: (a) location of the study area across the permafrost region in the northern hemisphere. The permafrost extent data are from Obu et al. [1]; (b) topographic map of the whole study area based on a 90 m DEM data, including the locations of sample site in the subregions. The background is based on true color satellite imagery of the Landsat-8.

The topographic features varied significantly within the study area from north to south. Most parts of the terrain were located above 4500 m asl, with the alternating distribution of mountains, valleys, and basins. Based on different topographic features and surficial geology, we divided the whole study area into six sub-regions (Table 1). These were the Qinghusihe–Chumaerhe (QC) upland plain, the Wudaoliang–Beiluhe (WB) basin, TuoTuohe (TTH) region, Kaixinling (KXL) upland plain, Tanggula (TGL) mountain areas, and Anduo (AD) plain (Figure 1b). All of these regions are underlain by continuous permafrost, which had a mean annual ground temperature (MAGT) ranging from $-3.0 \degree C$ to $0 \degree C$ [11–13].
Table 1. Summaries of the characteristics of climate, landform, vegetation, surficial geology, and permafrost conditions in six subregions.

| Section                | MAAT (°C)/Rainfall (mm) | Landform                     | Vegetation               | Surficial Geology          | MAGT (°C) | Ground Ice Condition                | Lake Number |
|------------------------|-------------------------|------------------------------|--------------------------|---------------------------|-----------|-------------------------------------|-------------|
| Qinghusihe–Chumaerhe (QC) | −4.1/527                | Upland alluvial plain        | Alpine meadow            | Silt loam and clay        | −1.0−0.7  | Ice rich (>50%) and ice saturated   | 11,423      |
| Wudaoliang–Beiluhe (WB) | −4.5/312                 | Upland denudation platform and basin | Alpine meadow and Sparse grassland | Silt loam and clay | −2.0−2.4  | Ice-saturated and ice layers with soils | 4074        |
| TuoTuoho (TTH)         | −3.5/302                 | Upland pluvial plain         | Alpine steppe            | Silt loam                 | −0.1−0.1  | Less ice and talik                  | 1520        |
| Kaixinling (KXL)       | −3.2/370                 | Upland alluvial plain        | Alpine meadow            | Loamy sand                | −0.8−0.7  | Ice-saturated and ice layers with soils | 1533        |
| Tanggula (TGL)         | −4.7/352                 | Upland till plain            | Alpine meadow            | Loamy sand                | −0.8−1.2  | Ice rich (>50%)                     | 5392        |
| Anduo (AD)             | −2.2/459                 | Upland alluvial and pluvial plain | Alpine meadow            | Silty clay                | 0−0.1     | Ice rich (>50%)                     | 4746        |

3. Methods and Data

3.1. Satellite Data and Processing

We used waterbody maps from the study by Niu et al. [32] for the analysis of waterbody size distribution and shape features. Detailed methods in Niu et al. [32] are summarized herein. Gaofen-1 (GF; Gaofen = high-resolution) satellite imagery was the basis for mapping the surface water bodies. Forty-five images with poor cloud were acquired from the China Centre for Resources Satellite Data and Application (CCRSDA) and preprocessed, including geometric and atmospheric correction and image fusion. These images had a spatial resolution of 2.0 m in panchromatic mode. To extract the water bodies automatically, an object-oriented classification method was applied to generate a map of water bodies based on the typical features in image color, geometrical shape, and uniform texture. Accuracy was assessed using the field surveys of lakes and ponds (Figure 1b; [33]) or visual interpretation. The accuracy of the waterbody maps was more than 90% [32]. Detailed descriptions of the image properties and the classification procedures used can be found in Niu et al. [32]. In this study, the water body classifications were converted from raster layers to vector polygons. The rivers and streams were manually removed from the datasets. Waterbodies smaller than 16 m², corresponding to 4 pixels at the lowest resolution of 2.0 m, were removed. Waterbodies > 1.0 km² were also excluded because they were not considered thermokarst lakes in this study. Meanwhile, waterbodies > 1 km² contributed only 1.2% to the total number of waterbodies along the QTEC. In the following, all water bodies were referred to as thermokarst lakes (TLs) even though smaller water bodies may be actually ponds.

3.2. Spatial Analyses

To analyze the shape features, five morphometric variables [34]—area, perimeter, circularity index (CI), elongation index (EI), and the orientation of major axis (OMA)—were calculated for all TLs. The CI (12.56 × area/perimeter²) was used to measure how strongly a TL deviates from a perfect circle (CI = 1). The EI (major axis/minor axis; 1 = equal axes) and OMA were calculated based on a best-approximated ellipse. These variables were calculated using a Geographical Information System (GIS) software—ArcGIS [35] and its spatial data analysis toolbox.

We used the thermal borehole records and field investigations (Figure 1b; [33]) to clarify the TL characteristics in relation to permafrost thermal conditions, surficial geology, and vegetation cover. The climate, permafrost thermal state, and ground ice conditions along the highway were obtained from Wu et al. [13] and Zhao et al. [12,36]. These field stations coincide directly within our study areas. The vegetation cover, surficial geology,
and landforms were based on our field investigations and the studies by Jin et al. [27] and Niu et al. [32].

3.3. Machine Learning-Based Susceptibility Analysis

3.3.1. Machine Learning Model

We used geospatial data to investigate the influences of the climate, terrain, and soil variables with a known physical relationship to TL activities in ice-rich permafrost regions (Figure 2). A machine learning method—random forest (RF)—was used to estimate the probabilities of TL occurrence based on the intrinsic properties of environmental variables. The result was termed TL susceptibility analysis.

RF is an ensemble modeling approach that generates a large number of uncorrelated regression trees to form a single model with improved prediction accuracy [37]. RF is a widely used approach in regression and classification tasks [38] and is successfully applied in geomorphic distribution modeling and mapping [39–41]. In this study, the modeling process was implemented using R program [42] with the “Biomod2” package [43,44]. The maximum number of trees was set to 500, which was the default value in Biomod2, and widely used in geomorphic modeling [3,39–41,45]. Each tree was built by randomly selecting a training dataset from the observations (i.e., bootstrap sample with replacement).

To examine the contribution of each environmental variable to the TL occurrence probability, a variable importance score was calculated based on modeling technique-independent runs within Biomod2 [44]. A univariate probability response curve was created to show the relationship between TL occurrence and the corresponding environmental variable [44]. The response curves for each variable were created using the “response.plot2” function in Biomod2 [44].

Finally, the validated RF model was applied to the regional probabilities of TL occurrence by relating the compiled TLs to spatial environmental conditions.

3.3.2. Environmental Variables

Two climate variables including thawing degree days (TDD, °C * days) and rainfall (mm) were calculated during the period 1979–2014, which covered the acquisition time of the satellite images. TDD was the effective temperature sum of mean monthly temperatures above 0 °C [46]. TDD and rainfall were keys factors influencing the permafrost thermal conditions [45]. The climate variables (i.e., TDD and rainfall) were calculated separately using the China Meteorological Forcing Dataset (CMFD, https://data.tpdc.ac.cn/en/,
This database includes the monthly mean air temperature and precipitation sums with a spatial resolution of 0.1 degree [47]. In this study, monthly air temperature and annual precipitation were statistically downscaled to 100 m² resolution by computing atmospheric lapse rates (ALR) and the elevation differences between the CMFD orography surface and the elevation of a 100 × 100 m² grid cell (Figure S1). The downscaling algorithm is similar to the process described in Fiddes and Gruber [48]. Monthly ALR values for air temperature were set based on Zhang et al. (Table S1) [49]. Annual ALR values for precipitation were based on a regional-mean precipitation increase of 18.3 mm per 100 m elevation [50]. The Shuttle Radar Topography Mission (SRTM) digital elevation data at 90 × 90 m² resolution [51] was employed for the elevation difference between the CMFD orography surface and the elevation of a 100 × 100 m² grid cell, while the CMFD orography surface was interpolated to the center of each 100 × 100 m² pixel (S1).

Four terrain properties: slope, aspect, potential annual incoming solar radiation (PISR, MJ cm⁻² year⁻¹), and topographic wetness index (TWI) were derived from the 90 m SRTM digital elevation model (DEM). Surficial sediment plays an important role in the accumulation of ground ice in permafrost soils [52,53]. Fine sediments (sum of clay and silt content < 50 µm) within 2.0 m depth from the surface were averaged from the SoilGrids database [54]. The spatial distribution of each environmental variable for our TLS modeling were shown in Figure 3.

To perform the thermokarst lake susceptibility (TLS) analysis, all the aforementioned datasets were resampled to 100 × 100 m² to match the model resolution. In addition, the TL polygons were converted to points, while an equal number of undisturbed points were randomly selected for evaluating landscapes where TLs are absent. Finally, the training and testing dataset was produced based on a binary variable (1 = presence, 0 = absence), which were transformed from the points.

3.3.3. Model Performance

Model performance was assessed with a 10-fold cross-validation (CV) scheme (Figure 2). At each CV run, the model was fit using random 70% of the observation data and evaluated against the remaining 30%. To evaluate the overall performance results for each run, the area under receiver-operating curves (AUC), Cohen’s kappa Index (Kappa), and true skill statistic (TSS) were implemented as further quality metrics [55] (S2).
3.4. Producing TLS Maps

To distinguish and predict the locations that were more hazardous in the permafrost region. A thermokarst lake susceptibility (TLS) map was produced. For this study, the calibrated machine learning model (i.e., RF) was applied to the entire QTEC. The probability map was classified into five susceptibility zones based on threshold values of 20, 40, 60,
and 80 percentiles. These values represented very low (<20%), low (20–40%), moderate (40–60%), high (60–80%), and very high (>80%) susceptibilities to future disturbance.

3.5. Other Statistical Analysis

The Kolmogorov–Smirnov test was first used to test the non-normal distribution for all variables. A non-parametric Mann–Whitney-U test was run to confirm the morphometric differences among the TLs located in different subregions. Pearson’s correlation (r) was applied to assess the correlations between environmental variables. Complete independence between two datasets produced a Pearson’s r correlation value of 0. A bivariate descriptor, variance inflation factors (VIF), was used to detect and quantify multicollinearity among environmental variables.

4. Results

4.1. Spatial Distribution of TLs

Figure 4a showed the distribution of all thermokarst lakes (TLs) in the study area. Finally, 34,915 water bodies were classified as TLs, which had an area ranging from 17.6 m² to 944,600 m². These TLs had a total area of 187.6 km² that covered 8.3% of the total QTEC landscape. The TL size distributions followed a power-law distribution (Figure 4b). Frequency distributions of TL area values showed strong skewness towards lower values because over 90% of TLs were smaller than 1 ha (10,000 m²; Figure 4b). TLs tended to cluster in the open flat zones (e.g., Chumaerhe and Xieshuihe regions, Figure 4c), upland basins (e.g., Beiluhe basin), and upland plains (e.g., Tanggula Mountain areas). TL area densities reached particularly high (0.1–0.3 km² km⁻²) in these regions (Figure 4c). Table 1 summarized the climate, landform, and permafrost-related conditions in these sub-regions. These regions usually featured a flat topography with a vegetation cover of alpine meadow. Permafrost was generally warm (MAGT > −3.0 °C) with the ice-rich or ice-saturated fine-sediments.

Mann–Whitney-U test indicated the significant differences among the TL subgroups referring to morphometric features. The morphometric characteristics of TLs varied in different sub-regions. As shown in Figure 5a, TLs distributed in the Wudaoliang–Beiluhe (WB) basin were generally bigger than those in other sub-regions. TLs in Anduo (AD) had the smallest surfaces compared with TLs in other regions. The shape of TLs in WB was characterized by an ellipse which was more elongated than the TLs in the other five sub-regions (Figure 5b,c). Distributions of major axis orientation indicated that most TLs in the WB basin were in the northern direction (Figure 5d). TLs in the other five regions exhibited a major NNE orientation.
Figure 4. Spatial distribution of thermokarst lakes (TLs) on the territory: (a) overview of the TLs distribution in the QTEC; (b) histogram of the area of all TLs in the study area; and (c) section of the study area with the lake area density per 1 km² and the elevation. Background hillshade is from the elevation model (DEM).
Figure 5. The morphometric characteristics of TLs in six subregions. QC= Qinghusihe–Chumaerhe; WB = Wudaoliang–Beiluhe; TTH = TuoTuohe; KXL = Kaixinling; TGL = Tanggula; AD = Anduo. (a) area distribution; (b) circularity index (EI); (c) elongation index (EI); (d) major axis orientation of TLs. 0° = east, 90° = north, 180° = west.

4.2. Model Performance

In this study, r correlations between environmental variables were less than 0.7 (Figure 6), which indicated a low level of linearity between variables. The highest VIF value is 4.80 indicating no multicollinearity among the seven environmental variables in this study (Table 2).

Based on the split-sample approach (i.e., 10-fold cross-validation), the model performance was evaluated with the values of TSS, AUC, and Kappa, respectively. The mean value of AUC was 0.93 (±0.03), and the corresponding mean values of TSS and Kappa were 0.73 (±0.11) and 0.72 (±0.12), respectively. These evaluation metrics indicated that the RF approach well-predicated the TL occurrences and validation was satisfactory in the permafrost region.
4.3. Environmental Variable Importance

According to the analyses of variable importance (Figure 7), rainfall was the most important variable contributing to the occurrence of TLs, while the slope angle and fine soil foremostly affected the TL occurrence potential. The TDD was ranked as the fourth important factor delineating suitable conditions for TL. In addition, the PISR and TWI played notable roles in constraining environmental suitability for the TL. The aspect had the lowest impact on the occurrences of TL (Figures 7 and 8a). The response shapes indicated that the optimal conditions for TL were characterized by moderately high rainfall ($\geq 260$ mm/a; Figure 8b) with flat topography (slope angle $< 5^\circ$; Figure 8c), and fine soil sediment (Figure 8d). In addition, TL occurrence needed the TDD between 630 $^\circ$C * days and 1000 $^\circ$C * days (Figure 8e), and PISR between 0.58 MJ cm$^{-2}$ a$^{-1}$ and 0.59 MJ cm$^{-2}$ a$^{-1}$ (Figure 8f), while TWI also had a notable contribution with the value of at least 3.7 (Figure 8g).
4.4. Thermokarst Lake Susceptibility Map

Finally, a thermokarst lake susceptibility (TLS) map was produced based on the probability of the modeled TL occurrence and was classified into five classes: very low...
susceptibility; low susceptibility; moderate susceptibility; high susceptibility; and very high susceptibility (Figure 9a). Most study area (65%) was classified as very-low- and low-risk for TL (i.e., probability < 40%; Figure 9b). Moderate- to high-susceptibility zones occupied 24% of the study area. The spatial extent of very-high zones was mainly located in valley plains, such as Xieshuihe and Chumaerhe, and upland plains (e.g., Tanggula Mountain areas). The model indicated that 4668 km$^2$ areas or 11% of the QTEC land were located in a very-high-susceptibility zone.

5. Discussion

5.1. Spatial Distribution and Development of TLs

Thermokarst lake mapping is limited by the low spatial resolution of large-scale imageries across the QTP. Therefore, previous studies attempted to estimate water bodies > 1 km$^2$ on the QTP, usually using Landsat series imagery [56,57]. Here, the application of the model indicated that...
5. Discussion
5.1. Spatial Distribution and Development of TLs

Thermokarst lake mapping is limited by the low spatial resolution of large-scale imagery across the QTP. Therefore, previous studies attempted to estimate water bodies > 1 km$^2$ on the QTP, usually using Landsat series imagery [56,57]. Here, the application of the high-resolution GF-1 image provided the opportunity to estimate the number of small water bodies (<1 km$^2$) on the QTP. In this study, TLs covered 8.3% of the study area, while the small TLs significantly contributed to the lake coverage. This coverage is relatively high compared to the 1.5% in the previous study [20], because of the application of high-resolution satellite images in this study. Our coverage is low compared to other arctic regions, for example, ca. 20% lake coverage in Alaska [58,59] and 30% in Canada [60], but close to the coverage in Siberia [34,61]. Given that the modeled high- to very-high-susceptibility zones, the TL-affected regions can be 20% of the total area in our study.

We found that TLs in each subregion had different morphometric characteristics (Figure 5), especially the lakes in Wudaoliang–Beiluhe (WH) basin. Lake size and shape have a close relationship with ground-ice content, local relief, and surface age [59]. Sediments with high ice contents, and low surface relief are beneficial to the lateral and vertical development of large lakes. The soil layer with rich ice is usually located in the depth of 2.0 m–5.0 m [62], which limits the significant thermokarst subsidence. Therefore, the lake depths are usually shallow (less than 3.0 m) [20] and the surface area is small (less than 1 km$^2$). The expansion of the lake lakeshore was usually at a rate of 1.8 m/year, which was controlled by the thermal erosion of lake water during the summer period [19,23]. Meanwhile, a lake with a water depth of more than 0.7 m will develop a talik several tens of meters deep in ice-rich permafrost over several hundred years [22,23,26] on the QTP. Our study did not investigate the spatial evolution of lake size and depth over time due to the lack of high-resolution images. Further work is needed to understand the impacts of ground ice on the evolution of lakes (including size, shape, and thermal conditions) in different permafrost landscapes in space and time.

5.2. Environmental Control Factors

To predict the responses of TL terrain to ongoing climatic warming, we need to better understand what environmental factors control TL development there. Usually, the ground ice volume and regional drainage gradient were the basic controls over the initiation of TLs [21,58]. Our machine learning method provided support for the combined impacts of climate and local environmental conditions on the distribution of TLs in permafrost landscapes. Overall, the climate, including the summer rainfall and TDD, turned out to be significant factors that determined the TL occurrence and distribution. Climate is the main spatial factor that affects the permafrost thermal conditions at a regional scale [45]. The effects of rainfall and TDD were attributed to the heat flux transferred into the permafrost. Climate changes (e.g., wetting and warming) increase heat conduction into the ground, as well as the water volume. Subsequent permafrost thermal degradation in terms of excess-ground-ice thaw can cause surface subsidence. In this study, most TLs were found in the upland plains and basins (Figure 4), where the thermokarst susceptibility is very high (Figure 9). These areas typically featured ice-rich permafrost and are relatively flat with fine-grained deposits [27]. Importantly, permafrost temperatures for these areas are usually warm (mean annual ground temperature > −3.0 °C) and sensitive to climate change [12]. Our results implied that soil properties hold notable roles in the distribution of TLs. Fine soil sediments (e.g., clay and silt) have a good capacity for holding excess ice [52,63]. The spatial distribution of ground ice as well as its amount and structure significantly influence thermokarst developments [22,58,64]. Flat topography (slope < 5° in this study) is beneficial to the sedimentation of fine soil and controls the surface drainage. In combination, permafrost thermal state, underlying surficial geology, and climate conditions contribute to repeated cycles of development of TLs at a regional scale.
Our modeling did not consider the lake shape and size. The main orientations of TLs were generally in the NNE direction, which is uniform with the main summer wind directions [20]. The wave-induced erosion could have an impact on the lateral growth of TLs [20, 34], however, the physics of the lateral erosion, including the wind direction and speed, are not fully clear and need further study in future work.

5.3. Implications of the TLS Modeling

Importantly, nonlinear relationships between climate and TL occurrence (Figure 8b, e) indicated that areas with low to moderate susceptibility can convert to the high-risk zone if the current climate changes to the sensitive value range in the future. In addition, our findings emphasize that similar underlying controls may exist across other regions on the QTP. Developing TLs are expected to have strong impacts on landscapes [6, 22] and hydrological environment processes [65], as well as carbon cycle [66]. On the QTP, TLs that have a water depth > 0.7 m cannot freeze back to the bottom in winter allow the development of a talik [25], which strongly changes the thermal and hydrological environment of underlying or surrounding permafrost [67]. Therefore, we infer that one or more of the small TLs gathering in the very-high-risk zone is likely to result in coalescence in the future.

Although we investigated the suitable environmental conditions for TL occurrences in recent climates, the physical processes (e.g., changes in ground-ice, thermal condition, and landscape) behind the TL development were not addressed. However, TLs are proxies for massive ice, and our TLS map (Figure 9) provided novel information on high ground ice content and permafrost thermal vulnerability across permafrost landscapes at a fine scale. Our approach yielded the feasibility in predicting the future responses of permafrost landscapes to climate warming in other regions, but also the assessment of the water resources across the QTP. In addition, the results of this study will be helpful to assess the impacts of thermokarst lake development on the hydrological environment, as well as the infrastructure integrity and engineering design. In addition, the proposed modeling approach in this study has implications for other permafrost landscapes across the whole QTP.

6. Conclusions

In this study, we used high-resolution satellite imagery to investigate the spatial distribution characteristics of thermokarst lakes (TLs) with sizes smaller than 1 km$^2$ in the Qinghai–Tibet Engineering Corridor (QTEC). Meanwhile, we presented a machine learning-based method for the thermokarst lake susceptibility (TLS) modeling using climate and terrain data as inputs across the permafrost region at a spatial resolution of 100 m$^2$. The following conclusions could be drawn:

(1) The TLs were mainly distributed in the upland plains and plateau that featured ice-rich permafrost and flat terrain with fine-grained deposits. In addition, 8.3% of the study area was covered by small TLs which had an area range from 17.6 m$^2$ to 944,600 m$^2$. These TLs usually featured a northwest direction and an ellipse shape.

(2) The suitable environmental conditions for TLs are closely related to climate, including the summer rainfall and air temperature (thawing degree days). Climate change induced the TL developments, provided that other geomorphic and soil conditions were suitable.

(3) The machine learning-based TL susceptibility map showed that more than 20% of the QTEC area was in high- to very-high-susceptibility zones. These areas were more vulnerable to climate change in the future, given the high and warm ground ice.

The results of this study improve our understanding of thermokarst development and its local control facts, as well as future geomorphological responses to climate change in permafrost landscapes underlain by different surficial geologies on the QTP. The potential implications should be considered when assessing the impacts of climate change on the local environment and engineering design.
Supplementary Materials: The following are available online at https://www.mdpi.com/article/10.3390/rs13101974/s1, Figure S1: Schematic of the downscaling scheme, Table S1: Monthly air temperature lapse rate (TLR, °C/100 m) on the QTP, Table S2: An error matrix used to evaluate the predictive accuracy of the presence–absence model.

Author Contributions: G.Y. wrote the paper, analyzed the field data as well as derived the methods. J.L. analyzed the remote sensing data, overviewed the whole process reviewed the paper, and provided comments. E.N. overviewed the whole process, reviewed the paper, and provided comments. F.Z. analyzed the field data. X.M. reviewed the paper, provided comments, and provided the field data. Z.L. conducted the field work and data collecting. M.L. finished the data collecting. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the National Key Research and development of China (Grant No. 2017YFA0603101), the National Natural Science Foundation of China (Grant No. 41801037 and 42071097), and CAS “Light of West China” Program.

Data Availability Statement: The model code and data that support the findings of this study are available upon reasonable request from the corresponding author.

Acknowledgments: The authors would like to thank the three anonymous reviewers for their constructive comments.

Conflicts of Interest: The authors declare no competing interests.

References
1. Obu, J.; Westermann, S.; Bartsch, A.; Berdnikov, N.; Christiansen, H.H.; Dashtseren, A.; Delaloye, R.; Elberling, B.; Etzelmüller, B.; Kholodov, A.; et al. Northern Hemisphere permafrost map based on TTOP modelling for 2000–2016 at 1 km² scale. Earth-Sci. Rev. 2019, 193, 299–316. [CrossRef]
2. Biskaborn, B.K.; Smith, S.L.; Noetzli, J.; Matthes, H.; Vieira, G.; Streletskiy, D.A.; Schoeneich, P.; Romanovsky, V.E.; Lewkowicz, A.G.; Abramov, A.; et al. Permafrost is warming at a global scale. Nat. Commun. 2019, 10, 264. [CrossRef] [PubMed]
3. Aalto, J.; Karjalainen, O.; Hjort, J.; Luoto, M. Statistical forecasting of current and future circum-Arctic ground temperatures and active layer thickness. Geophys. Res. Lett. 2018, 45, 4889–4898. [CrossRef]
4. Guo, D.; Wang, H. CMIP5 permafrost degradation projection: A comparison among different regions. J. Geophys. Res. Atmos. 2016, 121, 4499–4517. [CrossRef]
5. Osterkamp, T.E.; Jorgenson, M.T.; Schuur, E.A.G.; Shur, Y.L.; Kanevskiy, M.Z.; Vogel, J.G.; Tumskoy, V.E. Physical and ecological changes associated with warming permafrost and thermokarst in Interior Alaska. Permafrost. Periglac. Process. 2009, 20, 235–256. [CrossRef]
6. Nitzbon, J.; Westermann, S.; Langer, M.; Martin, L.C.P.; Strauss, J.; Laboor, S.; Boike, J. Fast response of cold ice-rich permafrost in northeast Siberia to a warming climate. Nat. Commun. 2020, 11, 2201. [CrossRef]
7. Grosse, G.; Harden, J.; Turetsky, M.; McGuire, A.D.; Camill, P.; Tarnocai, C.; Froliking, S.; Schuur, E.A.G.; Jorgenson, T.; Marchenko, S.; et al. Vulnerability of high-latitude soil organic carbon in North America to disturbance. J. Geophys. Res. Space Phys. 2011, 116. [CrossRef]
8. Qiu, J. The third pole. Nature 2008, 454, 24. [CrossRef]
9. Zou, D.; Zhao, L.; Sheng, Y.; Chen, J.; Hu, G.; Wu, T.; Wu, J.; Xie, C.; Wu, X.; Pang, Q.; et al. A new map of permafrost distribution on the Tibetan Plateau. Cryosphere 2017, 11, 2527–2542. [CrossRef]
10. Yang, M.; Wang, X.; Pang, G.; Wan, G.; Liu, Z. The Tibetan Plateau cryosphere: Observations and model simulations for current status and recent changes. Earth-Sci. Rev. 2019, 190, 353–369. [CrossRef]
11. Zhao, L.; Wu, Q.; Marchenko, S.; Sharkhuhu, N. Thermal state of permafrost and active layer in Central Asia during the international polar year. Permafro. Periglac. Process. 2010, 21, 198–207. [CrossRef]
12. Li, R.; Wu, Q.; Li, X.; Sheng, Y.; Hu, G.; Cheng, G.; Zhao, L.; Jin, H.; Zou, D.; Wu, X. Characteristic, changes and impacts of permafrost on Qinghai-Tibet Plateau. Chin. Sci. Bull. 2019, 64, 2783–2795. [CrossRef]
13. Wu, Q.; Zhang, T.; Liu, Y. Thermal state of the active layer and permafrost along the Qinghai-Xizang (Tibet) Railway from 2006 to 2010. Cryosphere 2012, 6, 607–612. [CrossRef]
14. Wu, Q.; Zhang, Z.; Gao, S.; Ma, W. Thermal impacts of engineering activities and vegetation layer on permafrost in different alpine ecosystems of the Qinghai-Tibet Plateau, China. Cryosphere 2016, 10, 1695–1706. [CrossRef]
15. Grosse, G.; Goetz, S.; McGuire, A.D.; Romanovsky, V.E.; Schuur, E.A.G. Changing permafrost in a warming world and feedbacks to the Earth system. Environ. Res. Lett. 2016, 11, 040201. [CrossRef]
16. Kokelj, S.V.; Jorgenson, M.T. Advances in Thermokarst Research. Permafro. Periglac. Process. 2013, 24, 108–119. [CrossRef]
17. Olefeldt, D.; Goswami, S.; Grosse, G.; Hayes, D.; Hugelius, G.; Kuhry, P.; McGuire, A.D.; Romanovsky, V.E.; Sannel, A.; Schuur, E.; et al. Circumpolar distribution and carbon storage of thermokarst landscapes. Nat. Commun. 2016, 7, 13043. [CrossRef] [PubMed]
18. Walter, K.M.; Zimov, S.A.; Chanton, J.P.; Verbyla, D.; Chapin, F.S. Methane bubbling from Siberian thaw lakes as a positive feedback to climate warming. Nat. Cell Biol. 2006, 443, 71–75. [CrossRef]

19. Luo, J.; Niu, F.; Lin, Z.; Liu, M.; Yin, G. Thermokarst lake changes between 1969 and 2010 in the Beilu River Basin, Qinghai–Tibet Plateau, China. Sci. Bull. 2015, 60, 556–564. [CrossRef]

20. Niu, F.; Luo, J.; Lin, Z.; Liu, M.; Yin, G. Morphological Characteristics of Thermokarst Lakes along the Qinghai-Tibet Engineering Corridor. Acta. Antarct. Alp. Res. 2014, 46, 963–974. [CrossRef]

21. Grosse, G.; Jones, B.; Arp, C. Thermokarst Lakes, Drainage, and Drained Basins. In Treatise on Geomorphology; Shroder, J.F., Ed.; Academic Press: San Diego, CA, USA, 2013; pp. 325–353.

22. Lin, Z.; Luo, J.; Niu, F. Development of a thermokarst lake and its thermal effects on permafrost over nearly 10 yr in the Beiluhe Basin, Qinghai-Tibet Plateau. Geosphere 2016, 12, 632–643. [CrossRef]

23. Lin, Z.; Niu, F.; Xu, Z.; Xu, J.; Wang, P. Thermal regime of a thermokarst lake and its influence on permafrost, Beiluhe Basin, Qinghai-Tibet Plateau. Pernafrr. Periglac. Process. 2010, 21, 315–324. [CrossRef]

24. Niu, F.; Lin, Z.; Liu, H.; Lu, J. Characteristics of thermokarst lakes and their influence on permafrost in Qinghai–Tibet Plateau. Geomorphology 2011, 132, 222–233. [CrossRef]

25. Lin, Z.; Niu, F.; Fang, J.; Luo, J.; Yin, G. Interannual variations in the hydrothermal regime around a thermokarst lake in Beiluhe, Qinghai-Tibet Plateau. Geomorphology 2017, 276, 16–26. [CrossRef]

26. Ling, F.; Wu, Q.; Zhang, T.; Niu, F. Modelling Open-Talik Formation and Permafrost Lateral Thaw under a Thermokarst Lake, Beiluhe Basin, Qinghai-Tibet Plateau. Pernafrr. Periglac. Process. 2012, 23, 312–321. [CrossRef]

27. Jin, H.; Wei, Z.; Wang, S.; Yu, Q.; Lu, L.; Wu, Q.; Ji, Y. Assessment of frozen-ground conditions for engineering geology along the Qinghai–Tibet highway and railway, China. Eng. Geol. 2008, 101, 96–109. [CrossRef]

28. Zhao, L.; Zou, D.; Hu, G.; Du, E.; Pang, Q.; Xiao, Y.; Li, R.; Sheng, Y.; Wu, X.; Sun, Z.; et al. Changing climate and the permafrost environment on the Qinghai-Tibet (Xizang) plateau. Pernafrr. Periglac. Process. 2020, 31, 396–405. [CrossRef]

29. Zhou, T.; Zhang, W. Anthropogenic warming of Tibetan Plateau and constrained future projection. Environ. Res. Lett. 2021, 16, 044039. [CrossRef]

30. Wu, Q.; Zhang, T. Changes in active layer thickness over the Qinghai-Tibetan Plateau from 1995 to 2007. J. Geophys. Res. Space Phys. 2010, 115. [CrossRef]

31. Chen, H.; Zhu, Q.; Peng, C.; Wu, N.; Wang, Y.; Fang, X.; Gao, Y.; Zhu, D.; Yang, G.; Tian, J.; et al. The Impacts of Climate Change and Human Activities on Biogeochemical Cycles on the Qinghai-Tibetan Plateau. Glob. Chang. Biol. 2013, 19, 2940–2955. [CrossRef]

32. Niu, F.; Yin, G.; Luo, J.; Lin, Z.; Liu, M. Permafrost Distribution along the Qinghai-Tibet Engineering Corridor, China Using High-Resolution Statistical Mapping and Modeling Integrated with Remote Sensing and GIS. Remote Sens. 2018, 10, 215. [CrossRef]

33. Gao, Z.; Lin, Z.; Niu, F.; Luo, J.; Liu, M.; Yin, G. Hydrochemistry and controlling mechanism of lakes in permafrost regions along the Qinghai-Tibet Engineering Corridor, China. Geomorphology 2017, 297, 159–169. [CrossRef]

34. Morgenstern, A.; Grosse, G.; Günther, F.; Fedorova, I.; Schirrmacher, L. Spatial analyses of thermokarst lakes and basins in Yedoma landscapes of the Lena Delta. Cryosphere 2011, 5, 849–867. [CrossRef]

35. ESRI. ArcGIS Desktop: Release 10.2.2; Environmental Systems Research Institute: Redlands, CA, USA, 2014.

36. Zhao, L.; Zou, D.; Hu, G.; Wu, T.; Du, E.; Liu, G.; Xiao, Y.; Li, R.; Pang, Q.; Qiao, Y.; et al. A synthesis dataset of permafrost thermal state for the Qinghai-Xizang (Tibet) Plateau, China. Earth Syst. Sci. Data Discuss. 2021, 2021, 1–22.

37. Breiman, L. Random forests. Mach. Learn. 2001, 45, 5–32. [CrossRef]

38. Belgum, M.; Drăguț, L. Random forest in remote sensing: A review of applications and future directions. ISPRS J. Photogramm. Remote Sens. 2016, 114, 24–31. [CrossRef]

39. Karjalainen, O.; Luoto, P.M.; Aalto, J.; Etxelmuller, P.B.; Grosse, G.; Jones, B.M.; Lilleeeren, K.S.; Hjort, J. High potential for loss of permafrost landforms in a changing climate. Environ. Res. Lett. 2020, 15, 104065. [CrossRef]

40. Aalto, J.; Luoto, M. Integrating climate and local factors for geomorphological distribution models. Earth Surf. Process. Landf. 2014, 39, 1729–1740. [CrossRef]

41. Marmion, M.; Hjort, J.; Thuiller, W.; Luoto, M. A comparison of predictive methods in modelling the distribution of periglacial landforms in Finnish Lapland. Earth Surf. Process. Landf. 2008, 33, 2241–2254. [CrossRef]

42. R Core Team. R: A Language and Environment for Statistical Computing; R Foundation for Statistical Computing: Vienna, Austria, 2020; Available online: http://www.R-project.org/ (accessed on 1 March 2021).

43. Thuiller, W.; Georges, D.; Engler, R.; Breiner, F. Biomod2: Ensemble Platform for Species Distribution Modeling. R Package Version: 3.4.6. Available online: https://cran.r-project.org/web/packages/biomod2/ (accessed on 1 March 2021).

44. Thuiller, W.; Lafourcade, B.; Engler, R.; Araújo, M.B. BIOMOD—A platform for ensemble forecasting of species distributions. Ecography 2009, 32, 369–373. [CrossRef]

45. Karjalainen, O.; Luoto, M.; Aalto, J.; Hjort, J. New insights into the environmental factors controlling the ground thermal regime across the Northern Hemisphere: A comparison between permafrost and non-permafrost areas. Cryosphere 2019, 13, 693–707. [CrossRef]

46. Frauenfeld, O.W.; Zhang, T.; McCreight, J.L. Northern Hemisphere freezing/thawing index variations over the twentieth century. Int. J. Clim. 2006, 26, 47–63. [CrossRef]
47. Yang, K.; He, J. China Meteorological Forcing Dataset (1979–2018); National Tibetan Plateau Data Center: Beijing, China, 2019.
48. Fiddes, J.; Gruber, S. TopoSCALE v.1.0: Downscaling gridded climate data in complex terrain. Geosci. Model Dev. 2014, 7, 387–405. [CrossRef]
49. Zhang, H.; Zhang, F.; Zhang, G.; Che, T.; Yan, W. How Accurately Can the Air Temperature Lapse Rate over the Tibetan Plateau Be Estimated from MODIS LSTs? J. Geophys. Res. Atmos. 2018, 123, 3943–3960. [CrossRef]
50. Guo, X.; Wang, L.; Tian, L. Spatio-temporal variability of vertical gradients of major meteorological observations around the Tibetan Plateau. Int. J. Clin. 2016, 36, 1901–1916. [CrossRef]
51. Jarvis, A.; Reuter, H.I.; Nelson, A.; Guevara, E. Hole-filled SRTM for the Globe Version 4, Available from the CGIAR-CSI SRTM 90m Database. 2008. Available online: http://srtm.csi.cgiar.org (accessed on 1 February 2021).
52. Cheng, G. The mechanism of repeated-segregation for the formation of thick layered ground ice. Cold Reg. Sci. Technol. 1983, 8, 57–66.
53. Kokelj, S.V.; Burn, C.R. Near-surface ground ice in sediments of the Mackenzie Delta, Northwest Territories, Canada. Permaf. Periglac. Process. 2005, 16, 291–303. [CrossRef]
54. Hengl, T.; De Jesus, J.M.; Heuvelink, G.B.M.; Gonzalez, M.R.; Kilibarda, M.; Blagotić, A.; Shangguan, W.; Wright, M.N.; Geng, X.; Bauer-Marschallinger, B.; et al. SoilGrids250m: Global gridded soil information based on machine learning. PLoS ONE 2017, 12, e0169748. [CrossRef]
55. Allouche, O.; Tsoar, A.; Kadmon, R. Assessing the accuracy of species distribution models: Prevalence, kappa and the true skill statistic (TSS). J. Appl. Ecol. 2006, 43, 1223–1232. [CrossRef]
56. Zhang, G.; Yao, T.; Shum, C.K.; Yi, S.; Yang, K.; Xie, H.; Feng, W.; Bolch, T.; Wang, L.; Behrangi, A.; et al. Lake volume and groundwater storage variations in Tibetan Plateau’s endorheic basin. Geophys. Res. Lett. 2017, 44, 252–260. [CrossRef]
57. Zhang, G.; Yao, T.; Shum, C.K.; Yi, S.; Yang, K.; Xie, H.; Feng, W.; Bolch, T.; Wang, L.; Behrangi, A.; et al. Extensive and drastically different alpine lake changes on Asia’s high plateaus during the past four decades. Geophys. Res. Lett. 2017, 44, 5550–5560. [CrossRef]
58. Farquharson, L.; Mann, D.; Grosse, G.; Jones, B.; Romanovsky, V. Spatial distribution of thermokarst terrain in Arctic Alaska. Geomorphology 2016, 273, 116–133. [CrossRef]
59. Hinkel, K.M.; Frohn, R.C.; Nelson, F.E.; Eisner, W.R.; Beck, R.A. Morphometric and spatial analysis of thaw lakes and drained thaw lake basins in the western Arctic Coastal Plain, Alaska. Permaf. Periglac. Process. 2005, 16, 327–341. [CrossRef]
60. Côté, M.M.; Burn, C.R. The oriented lakes of Tuktoyaktuk Peninsula, Western Arctic Coast, Canada: A GIS-based analysis. Permaf. Periglac. Process. 2002, 13, 61–70. [CrossRef]
61. Grosse, G.; Romanovsky, V.; Walter, K.; Morgenstern, A.; Lantuit, H.; Zimov, S. Distribution of Thermokarst Lakes and Ponds at Three Yedoma Sites in Siberia in International Conference on Permafrost; Institute of Northern Engineering, University of Alaska Fairbanks: Fairbanks, AK, USA, 2008.
62. Lin, Z.; Gao, Z.; Fan, X.; Niu, F.; Luo, J.; Yin, G.; Liu, M. Factors controlling near surface ground-ice characteristics in a region of warm permafrost, Beiluhe Basin, Qinghai-Tibet Plateau. Geoderma 2020, 376, 114540. [CrossRef]
63. French, H.M. The Periglacial Environment; John Wiley and Sons: Hoboken, NJ, USA, 2007.
64. French, H.; Shur, Y. The principles of cryostratigraphy. Earth-Sci. Rev. 2010, 101, 190–206. [CrossRef]
65. Gao, Z.; Niu, F.; Wang, Y.; Lin, Z.; Wang, W. Suprapermafrost groundwater flow and exchange around a thermokarst lake on the Qinghai–Tibet Plateau, China. J. Hydrol. 2021, 593, 125882. [CrossRef]
66. Mu, C.; Abbott, B.W.; Norris, A.J.; Mu, M.; Fan, C.; Chen, X.; Jia, L.; Yang, R.; Zhang, T.; Wang, K.; et al. The status and stability of permafrost carbon on the Tibetan Plateau. Earth-Sci. Rev. 2020, 211, 103433. [CrossRef]
67. Walvoord, M.A.; Kurtylyk, B.L. Hydrologic Impacts of Thawing Permafrost—A Review. Vadose Zone J. 2016, 15, 1–20. [CrossRef]