Using geographically weighted regression to predict the spatial distribution of frozen ground temperature: a case in the Qinghai–Tibet Plateau

Rui Zhao, Mingxing Yao, Linchuan Yang, Hua Qi, Xianglian Meng and Fujun Zhou

1 Southwest Jiaotong University, Chengdu, People’s Republic of China
2 China Railway First Survey and Design Institute Group Co. Ltd, Xi’an, People’s Republic of China

E-mail: yanglc0125@swjtu.edu.cn

Keywords: geographically weighted regression, principal component analysis, frozen ground, ground temperature, permafrost, spatial heterogeneity, Qinghai–Tibet Plateau, Tibetan Plateau

Abstract
This paper combines the use of principal component analysis (PCA) and the geographically weighted regression (GWR) model to predict the spatial distribution of frozen ground temperature. PCA is used to reduce the multicollinearity among covariates, while the GWR model is used to address the spatially non-stationary relationship between frozen ground temperature and its predictors, such as air temperature, latitude, longitude, and vegetation cover. Our approach is applied in a typical permafrost area in the Qinghai–Tibet Plateau, Western China. The result demonstrates the applicability of our approach in the spatial distribution of frozen ground temperature and shows that the approach can be used for analysis and prediction. This study offers insight into temperature monitoring of frozen ground, which helps prevent regional geological disasters.

1. Introduction

With pronounced global warming, frozen ground, which is a major part of the cryosphere and includes permafrost and seasonally frozen ground, is facing degradation (James et al 2013, Vasiliiev et al 2020). For example, the Qinghai–Tibet Plateau (alternatively known as the Tibetan Plateau) has recently undergone a severe permafrost degradation problem. From 1975 to 2017, the area of the permafrost in the plateau has considerably decreased by 440 000 km$^2$ (Zou et al 2017). Moreover, the change in the active layer thickness in the plateau has accelerated. Evidently, the degradation of frozen ground may cause numerous geological disasters, including but not limited to collapse, landslides, mudslides, and mercury release (Li et al 2019, Mu et al 2020). As a large number of studies have indicated, frozen ground is highly sensitive to changes in ground temperature and thus becomes an indicator of climate change (Sheng and Yao 2009, Cheng et al 2013). It is widely acknowledged that understanding the temperature distributions of frozen ground is of paramount importance in many facets (Wang et al 2020). For example, it provides a valuable basis for the regional prevention of geological disasters.

The most common method used to explore temperature distributions of frozen ground is numerical modeling (or computational simulation). Numerous researchers have made the effort along this line. Yu et al (2014) coupled the finite element method with a heat transfer and seepage model to analyze the effect of underground oil heating pipes on the temperature of frozen ground. Mu et al (2016) conducted a heat transfer–based finite element analysis to simulate temperature distributions of thermosyphons and foundation soils around the frozen ground along the Qinghai–Tibet transmission line. Garayshin et al (2019) established a 2D nonlinear heat equation with the phase change to explore temperature across the ice wedge. Liu et al (2020) developed a two-dimensional non-steady-state heat conduction finite element model to predict the distribution of permafrost temperature by taking into account factors such as wind speed, wind direction, evaporation, and solar radiation. To facilitate the preparation of input parameters and the visualization of simulated results, Marchenko et al (2008) developed the GIPL
(Geophysical Institute Permafrost Laboratory) 2.0 model by combining the GIPL 1.0 model (which numerically solves 1D nonlinear heat equation with phase change) and ArcGIS. The GIPL 2.0 model is determined to well simulate the dynamic evolution of frozen ground temperature in Alaska, the United States. Jafarov et al (2012) adopted the GIPL2-MPI numerical transient model with various factors (e.g. altitude, slope, aspect, and precipitation) to simulate the temperature distribution of the Alaska permafrost region. Qin et al (2017) and Yin et al (2018) used the GIPL2 model to predict the active layer variation of the permafrost in the Qinghai–Tibet Plateau and the annual average temperature distribution in the permafrost region of the Qinghai–Tibet Plateau Engineering Corridor, respectively.

However, existing studies mainly focused on predicting frozen ground temperature at the vertical level under the analytical framework of numerical modeling. By contrast, they seldom investigated the temperature at the horizontal level. More importantly, they have rarely considered the spatially non-stationary relationship\(^3\) between frozen ground temperature and its influencing factors. Put it in another way, a spatially invariant (or fixed) relationship between frozen ground temperature and its correlates was often assumed. However, as Fotheringham et al (2002) indicated, it is often not the case in real life. In light of this, based on in-situ borehole observation data, this study applies geostatistical models (more specifically, geographically weighted regression (GWR) models) to analyze the variation in frozen ground temperature and to predict the spatial–temporal distribution of the temperature in June–September, 2001, in a typical permafrost area of the Qinghai–Tibet Plateau. In addition, in stark contrast to global regression models, the GWR model is a local regression technique that captures the spatial non-stationarity of associations between the dependent and independent variables (Fang et al 2015). On the other hand, in the application of GWR (and other regression models), it is easily trapped in multicollinearity among independent variables, thus giving rise to uncertainty in prediction results (Nguyen and Ng 2020). This is highly applicable to our temperature modeling study. In such a case, this study uses principal component analysis (PCA) to eliminate multicollinearity among contributory factors to frozen ground temperature.

The contributions of this study include (a) combining PCA and the GWR model to explore the spatially varying relationship between frozen ground temperature and its correlates; (b) predicting the spatial–temporal distribution of the temperature of frozen ground in the study area; and (c) revealing the potential of using multi-source data to predict the spatial–temporal distribution of frozen ground temperature and providing a new angle to this issue.

The remainder of the paper is organized as follows. Section 2 succinctly presents the data. Section 3 introduces the methodologies of PCA and the GWR model. Section 4 shows the results. Section 5 concludes the paper, discusses research limitations, and points out avenues for future research.

2. Data

2.1. Study area

Our study area is a typical permafrost area of the Qinghai–Tibet Plateau (figure 1). North to the Kunlun Mountains and south to the Chiqu Valley, this study area takes Wudaolang as the center and includes the Chumar River High Plain, the Hoh Xil Mountain area, and the Beilu River Basin. It is the most densely monitored area of frozen ground during the construction and operation of the Qinghai–Tibet Railway (figure 1). The average altitude of this area is 4900 m. This area has a plateau continental climate. It is characterized by an arid climate on the glacial margin. In the study area, the evaporation is much larger than the rainfall, and the vegetation is sparse. The average annual air temperature of the areas is from \(-2\,^{\circ}\text{C}\) to \(-6.9\,^{\circ}\text{C}\). The temperature is highest in July (with an average of 6.5–8.1\,^{\circ}\text{C}\) and lowest in January (with an average of \(-14.5\,^{\circ}\text{C}\) to \(-17.4\,^{\circ}\text{C}\)).

2.2. Frozen ground temperature data

The borehole temperature data were obtained by the First Survey and Design Institute of China Railway Group. The data utilized in this study were collected from a maximum of 231 boreholes from May to December, 2001, along the Qinghai–Tibet Railway. The depth for these boreholes ranges from 0.2 m to 15 m. Abnormal drilling data were eliminated before subsequent statistical analysis.

Through the statistical analysis of the temperature data measured at the boreholes (figure 2), we find that the average temperature of the study area is not much different from the median, indicating that the distribution of the ground temperature may be relatively concentrated with low dispersion. Moreover, the ground temperature in July and August was highest, indicating that permafrost’s active layer thickness may change significantly.

This study only focuses on frozen ground temperature distribution at many depths (ranging from 0.2 m to 15 m) from June to September because of limited observations in other months.

---

\(^3\)‘Spatially non-stationary relationship’ means that the association the dependent variable and covariates varies across space; or that the marginal responses of covariates are not spatially fixed. Its reasons are diverse, including model misspecification, sampling variation, intrinsically varying local spatial behavior, and omitted variables bias (Fotheringham et al 2002; Yang et al 2020a).
2.3. Selection of independent variables

The temperature of frozen ground is expected to be affected by elevation, longitude, latitude, and various local factors, including but not limited to rock lithology, topography, vegetation, snow, and water content (Osterkamp 2007, Wang et al 2020). Obviously, these factors have temperature effects that vary across space and time (or season). For example, vegetation and snow cover have varying temperature effects in cold and warm seasons. In addition, related studies have identified that meteorological factors have significant impacts on frozen ground temperature (Zhang and Stamnes 1998, Throop 2012, Liu et al 2013). To this end, a total of 13 independent variables are selected for statistical modeling, including elevation, longitude, latitude, air temperature, air temperature amplitude, precipitation, wind speed, net radiation budget (difference between short-wave radiation and long-wave radiation), rock lithology, slope, aspect, fractional vegetation cover (FVC), and snow.

Different lithological rocks may have various physical, surface heat transfer, water-holding, and permeability properties, resulting in different effects on frozen ground temperature (Smith and Riseborough 1996). Various categories of rocks should be considered, such as sandstone, conglomerate rock, tuff, siltstone, marlstone, and grit.

Terrain and geomorphology influences ground temperature mainly via slope and aspect (Kenner and Magnusson 2017). Areas with different slopes and aspects receive different levels of solar radiation, so they vary in evaporation and vegetation and thus frozen ground temperature (Cheng 2004).

FVC has cooling effects in summer but warming effects in winter on frozen ground temperature. Therefore, this study selects FVC as one of the independent variables (Yi et al 2011, Wang et al 2012).
Table 1. Pearson’s correlation coefficients between the dependent and independent variables in August.

| Variable                  | 0.2 m  | 0.5 m  | 1.0 m  | 1.5 m  | 2.0 m  | 2.5 m  | 3.0 m  | 3.5 m  | 4.0 m  | 4.5 m  | 5.0 m  | 10.0 m | 15.0 m |
|---------------------------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|
| Longitude                 | -0.378 | -0.345 | -0.303 | -0.249 | -0.298 | -0.358 | -0.394 | -0.401 | -0.403 | -0.404 | -0.403 | -0.406 | -0.410 |
| Latitude                  | -0.386 | -0.381 | -0.362 | -0.311 | -0.356 | -0.412 | -0.442 | -0.448 | -0.447 | -0.446 | -0.445 | -0.440 | -0.449 |
| Elevation                 | -0.442 | -0.434 | -0.342 | -0.203 | -0.192 | -0.277 | -0.400 | -0.485 | -0.527 | -0.549 | -0.568 | -0.606 | -0.614 |
| Slope                     | -0.177 | -0.105 | -0.124 | -0.124 | -0.136 | -0.154 | -0.169 | -0.197 | -0.224 | -0.231 | -0.235 | -0.225 | -0.221 |
| Aspect                    | -0.137 | -0.152 | -0.211 | -0.148 | -0.121 | -0.171 | -0.187 | -0.208 | -0.226 | -0.227 | -0.220 | -0.205 | -0.203 |
| Lithology                 | 0.082  | 0.155  | 0.321  | 0.453  | 0.534  | 0.448  | 0.312  | 0.217  | 0.154  | 0.129  | 0.128  | 0.115  | 0.133  |
| FVC                       | -0.188 | -0.191 | -0.203 | -0.185 | -0.168 | -0.16  | -0.196 | -0.231 | -0.250 | -0.265 | -0.291 | -0.316 | -0.322 |
| Air temperature           | 0.554  | 0.502  | 0.419  | 0.323  | 0.352  | 0.449  | 0.540  | 0.584  | 0.604  | 0.612  | 0.617  | 0.624  | 0.635  |
| Precipitation rate        | -0.348 | -0.294 | -0.237 | -0.184 | -0.232 | -0.293 | -0.333 | -0.342 | -0.347 | -0.348 | -0.347 | -0.355 | -0.357 |
| Wind speed                | 0.369  | 0.396  | 0.414  | 0.394  | 0.439  | 0.464  | 0.466  | 0.466  | 0.452  | 0.447  | 0.449  | 0.437  | 0.449  |
| Net radiation budget      | 0.136  | 0.221  | 0.295  | 0.308  | 0.326  | 0.309  | 0.274  | 0.257  | 0.242  | 0.234  | 0.235  | 0.215  | 0.227  |
| Air temperature amplitude | 0.523  | 0.555  | 0.504  | 0.398  | 0.392  | 0.469  | 0.568  | 0.630  | 0.655  | 0.663  | 0.673  | 0.668  | 0.682  |

* Significant at the 1% level (two-tailed).
* Significant at the 5% level (two-tailed).
Snow cover impacts the temperature of frozen ground in many ways. ‘Snow has high albedo and emissivity that cool the snow surface, high absorptivity that tends to warm the snow surface, low thermal conductivity so that a snow layer acts as an insulator, and high latent heat due to snowmelt that is a heat sink’ (Zhang 2005, p 1). The overall effect of snow cover on frozen ground temperature is dictated by various factors such as timing, duration, accumulation, density, structure, thickness of seasonal snow cover, and its interactions with other variables (e.g. micrometeorological conditions, local microrelief, and vegetation) (Zhang 2005, Park et al 2014, Zhang et al 2018). Hence, there should be numerous snow cover variables (e.g. snow depth and snow cover in the same period and the snow cover in the early winter and spring) for the prediction of frozen ground temperature.

Meteorological conditions, such as ground precipitation rate, wind speed, downward short-wave radiation, downward long-wave radiation, and air temperature amplitude, significantly influence frozen ground temperature (Zhang and Stamnes 1998, Throop et al 2012, Liu et al 2013, Qin et al 2013). It is known from climatology that the underlying surface is warmed by short-wave radiation and the ambient air is warmed by the underlying surface. Notably, this study uses net radiation budget, namely the difference between short-wave radiation and long-wave radiation as a predictor (Chou et al 2012, Xie et al 2015).

2.4. Measurement of independent variables

2.4.1. Latitude and longitude
It is measured by the authors.

2.4.2. Elevation, slope, and aspect
The study uses ArcGIS 10.2 to calculate the slope and aspect by the digital elevation model (DEM) (accuracy of 30 m) downloaded from the Geospatial Data Cloud website (www.gscloud.cn/).

2.4.3. Rock lithology
Its data are cut from the 1:1 million geological maps in 2002 and vectorized to obtain 13 basic rock units. By categorizing the borehole data distributed in different rock strata, the lithology character variable is quantified according to the inflection points of the borehole temperature in different months.

2.4.4. Vegetation cover
FVC is calculated by using the pixel dichotomy model with the input of the normalized difference vegetation index (NDVI). The NDVI data are obtained from the Terra satellite MOD13 Q1 global index (resolution = 250 m).

| Variable                  | VIF  |
|---------------------------|------|
| Longitude                 | >1000|
| Latitude                  | 476.573|
| Elevation                 | 5.824 |
| Slope                     | 1.374 |
| Aspect                    | 1.164 |
| Lithology                 | 2.254 |
| FVC                       | 2.984 |
| Air temperature           | 77.364|
| Precipitation rate        | >1000 |
| Wind speed                | 364.719|
| Net radiation budget      | 396.105|
| Air temperature amplitude | 15.489 |

2.4.5. Snow cover
Its data are derived from the 8 d synthetic product regarding the Terra satellite MOD10A2 global 500 m resolution snow cover index. During the study period (June–September, warm season), all the borehole points were not covered by snow, so we exclude snow cover variables in the subsequent analysis because of the lack of variability in the variables.

2.4.6. Meteorological conditions
Its data are provided by National Tibetan Plateau Data Center, including air temperature, ground precipitation rate, wind speed, downward short-wave radiation, and downward long-wave radiation (Yang et al 2010, Chen et al 2011). As the measurement unit of the original data is 0.1° × 0.1°, this study uses the Kriging interpolation method to transform the low-resolution data to high-resolution data (1 km × 1 km).

2.4.7. Air temperature amplitude
Its data are obtained from 1 km monthly minimum and maximum temperature datasets for China (1901–2017) (Peng et al 2017, 2018, 2019, Du and Peng 2020).

Pearson’s correlation analyses are conducted for the dependent and independent variables. Table 1 offers the results for August (N = 128). Since there was no snow cover in the borehole points in August, the number of independent variables becomes 12. It indicates that the selected 12 independent variables are more or less related to frozen ground temperature at different depths, justifying the selection of predictors.

Figures A1 and A2 show the distribution of time-constant and time-varying variables, respectively.

3. Methodology
Figure 3 shows the workflow of this study. Due to the uneven distribution of borehole monitoring stations in the study area, a 3 km × 3 km fishnet was established for the subsequent analysis in ArcGIS.
eliminate multicollinearity among some independent variables, the study converts them into independent principal components by PCA and extracts the components with characteristic roots above one as the key variables for GWR modeling. The leave-one-out cross-validation (LOOCV) method is employed to cross-validate the modeling results.

### 3.1. PCA
PCA, an eigenvector-based exploratory data analysis technique, uses an orthogonal transformation to convert a large number of potentially correlated variables into several components that are linearly uncorrelated and contain most, though not all, of the information involved in the original variables.
Table 4. Indicators of cross-validation for the GWR model.

| Model performance statistics | $R^2$ | MAE | RMSE |
|------------------------------|-------|-----|------|
| Value                        | 0.8581| 0.644 | 0.924 |

The transformation reduces the collinearity in the data. Before performing the PCA, the original data are usually normalized. The relationship between the original variables and components is expressed as follows (UI-Saafie et al. 2013):

$$PC_i = l_{i1}X_1 + l_{i2}X_2 + \cdots + l_{in}X_n, \quad i, j = 1, 2, \cdots, n$$

where $PC_i$ is the $i$th component, $X_j$ is the $j$th variable, and $l_{ij}$ is the coefficient (or weight) of $X_j$.

3.2. GWR

The traditional ordinary linear squares (OLS) regression model assumes that the relationship between a dependent variable and its covariates is fixed across space and provides a global modeling approach. By contrast, the GWR model embeds spatial location information in the data and extends the traditional OLS regression framework (Brundon et al. 1996, Fotheringham et al. 1998, 2002). It assumes that the association between the dependent variable and covariates can be spatially variant and offers a set of equations to estimate coefficients at any given location. It has been extensively employed in existing literature (Xu and Huang 2015, Yang et al. 2017, 2020a, 2020b, 2020c, Xu and Yang 2019, Zhao et al. 2020). The introduction of the GWR model is heavily borrowed from Yang et al. (2020a, 2020b).

The GWR model is expressed as follows:

$$y_i = \beta_0(u_i, v_i) + \sum_{k=1}^{p} \beta_k(u_i, v_i)x_{ik} + \varepsilon_i$$

where $y_i$ is the dependent variable in the $i$th regression point, $(u_i, v_i)$ is the coordinate of the $i$th point, $x_{ik}$ is the $k$th explanatory variable of the $i$th regression point, $\beta_k(u_i, v_i)$ is the coefficient of $x_{ik}$, and $\varepsilon_i$ is the error term ($\varepsilon_i \sim N(0, \sigma^2), \text{Cov}(\varepsilon_i, \varepsilon_j) = 0$ ($i \neq j$)).

Let $\beta(u_i, v_i) = [\beta_0(u_i, v_i), \beta_1(u_i, v_i), \cdots, \beta_p(u_i, v_i)]$, the regression coefficient $\hat{\beta}(u_i, v_i)$ can be estimated by the weighted least squares method:

$$\hat{\beta}(u_i, v_i) = (X^T W(u_i, v_i) X)^{-1}X^T W(u_i, v_i) y,$$
Figure 6. Ground temperature distribution at different depths in (a) June, (b) July, (c) August, and (d) September.
Figure 7. Observed ground temperature variation. Note: in each sub-figure, one curve represents temperature data from one borehole.

Figure 8. Distribution of the borehole measured points and prediction points.

where $W(u_i, v_i) = \text{diag}(w_{i1}, w_{i2}, \cdots, w_{in})$ is the weight matrix at $(u_i, v_i)$.

The estimated value $\hat{y}_i$ of the observation value $y_i$ is:

$$\hat{y}_i = x_i^T \beta(u_i, v_i) = x_i^T (X^T W(u_i, v_i) X)^{-1} X^T W(u_i, v_i) y.$$  

For any given observation, GWR adopts a kernel function to estimate the weights of geographically close observations and sets a rule that closer observations have a higher effect on the estimation of the coefficients than observations that are more geographically distant. The element of the weight matrix $W(u_i, v_i)$ is measured by means of a weight kernel function. The choice of the weight kernel function has significant influences on the estimation of the parameters of the GWR model. Typically, four types of spatial kernel functions are used, including the
4. Results

4.1. Selection of covariates

Diagnostics are needed to assess the presence of collinearity (or non-independence of covariates) in the data. Common collinearity diagnostics include the fixed Gaussian, fixed bi-square, adaptive Gaussian, and adaptive bi-square kernel functions:

Fixed Gaussian kernel: \( w_{ij} = \exp\left(-\frac{d_{ij}^2}{\theta^2}\right) \)

Adaptive Gaussian kernel: \( w_{ij} = \exp\left(-\frac{d_{ij}^2}{\theta_{i(k)}^2}\right) \)

Fixed bi-square kernel: \( w_{ij} = \begin{cases} (1 - \frac{d_{ij}^2}{\theta^2})^2 & \text{if } d_{ij} < \theta \\ 0 & \text{otherwise} \end{cases} \)

Adaptive bi-square kernel: \( w_{ij} = \begin{cases} (1 - \frac{d_{ij}^2}{\theta_{i(k)}^2})^2 & \text{if } d_{ij} < \theta_{i(k)} \\ 0 & \text{otherwise} \end{cases} \)

where \( w_{ij} \) is the weight of point \( j \) in the estimation of the local regression equation of point \( i \), \( d_{ij} \) is the Euclidean (straight-line, crow-fly) distance between points \( i \) and \( j \), \( \theta \) is a fixed bandwidth, and \( \theta_{i(k)} \) is an adaptive bandwidth defined by the \( k \)th nearest neighbor distance for estimating the local equation of point \( i \).

GWR is sensitive to the bandwidth. If the bandwidth is too large, parameter estimates will be similar to global estimates. If the bandwidth is too small, parameter estimates will be highly localized, and their variance may be large (Windle et al 2010).
absolute value of correlation coefficient, condition number, variance-decomposition proportions, and variance inflation factor (VIF) (Dormann et al 2013). In this study, VIFs are used to evaluate the collinearity in the data for all the months, including June–September. Table 2 provides the VIF values of all the 12 independent variables in August and reveals the non-independence of many variables. (VIFs for other months are available from the authors upon reasonable request.)

To retain as many independent variables as possible, this study does not conduct PCA for all the 12 variables. Instead, it keeps five linearly uncorrelated variables (i.e. Elevation, Lithology, Slope, Aspect, and FVC) and transforms seven variables that are subject to collinearity. Two out of seven principal components (PC1 and PC2) have an eigenvalue greater than 1, and they collectively capture 90.06% of the variation in the data. Therefore, the two principal components are included in the GWR model. Table 3 shows the VIF values of the five linearly uncorrelated variables and the two PCs. Observe that VIF values of all the seven variables are far less than 10. This observation illustrates the absence of collinearity between the variables (Dormann et al 2013) and suggests that the seven variables can be used to predict frozen ground temperature in the GWR modeling framework.

4.2. Prediction of the spatial distribution of frozen ground temperature

Temperature data from hundreds of boreholes along the Qinghai–Tibet Railway and its covariates are used to predict the region-wide frozen ground temperature at many depths. Figure 4 provides predicted temperature fields in all the 4 months.

We initially interpret the temperature fields at the horizontal level, and frozen ground temperature at a specific depth (0.5 m) is used for illustration. Figure 5 shows the spatial distribution of predicted temperature at 0.5 m below the frozen ground from June and September. Such distributions seem highly reasonable. The ground temperature is relatively low
in high-altitude areas, including the Kunlun Mountains, the Hoh Xil Mountain, and the Fenghuo Mountain. Comparatively, the temperature is high in high plains and basins. In other words, the temperature is predicted to decrease with advancing altitude. In addition, the ground temperature is higher in areas adjacent to lakes and rivers than in other areas. Moreover, the change in ground temperature is bigger in July than in August. Furthermore, 0.5 m frozen ground temperature ranges from 0 °C to 10 °C in July–September, but it can reach −4 °C in June. Besides, from June to September, frozen ground temperature first increases and then decreases, and it peaks in July.

Figure 6 shows the spatial distribution of frozen ground temperature at different depths for each month. Observe that the ground temperature decreases with increasing depth. The Chumar, Beilu, and Tongtian River Basins demonstrate a similar characteristic: the temperature is higher in the downstream area than in the upstream. In the depth range of 0–5 m, the ground temperature varies greatly. By contrast, when the depth is above 5 m, the ground temperature changed a little.

By comparing the spatial distributions in different months, we find that except for the 0.5 m frozen ground, the temperature of the frozen ground at all the other depths increases from June to September, and it reaches the peak in September. This finding is due to the time lag caused by the downward temperature diffusion.

Figure 7 shows the variation in ground temperature based on the observed borehole data. It indicates that the ground temperature fluctuates between −4 °C and 16 °C. Notably, the ground temperature varies significantly within a depth range of 0–5 m, and its maximum difference can reach 16 °C. Within 5–15 m, the ground temperature remains relatively stable. Such results provide insightful evidence to support the spatial distribution of ground temperature regarding the study area.

4.3. Result of model validation
4.3.1. Performance statistics
Table 4 shows the cross-validation results. The GWR model has a high degree of goodness of fit, and its R² reaches 85.81%. In addition, the mean absolute prediction error (MAE) and root mean square error (RMSE) values are small, implying that the GWR model is robust.

4.3.2. Results of predictions for nearby locations
To further validate the GWR model and gain confidence in the proposed methodology, we decide to compare the ground temperature data from some randomly selected borehole points to the prediction results of nearby points (figure 8(a)).

On the one hand, points with similar topography and geomorphology are compared. The borehole points in the Kunlun Mountains and the Hoh Xil Mountain area are selected to compare with those located within 10 km (figures 8(b) and (c)). On the other hand, points with similar rock lithology are compared. The study randomly selects the borehole points and the corresponding prediction points in the Fenghuo Mountain area (figure 8(d)) to make the comparison.

Figure 9 shows the comparison of observed temperature at the borehole points and predicted temperature at nearby points. Ground temperatures at the borehole points and their nearby points are relatively consistent. This further illustrates that the GWR model can be used to simulate the ground temperature in the study area effectively.

4.3.3. Results of predictions for terrain-specific locations
To gain further confidence in model performance, we decide to investigate the relationship between terrain types and frozen ground temperature, thus categorizing the study area into three groups: mountainous area, high plain, and basin. Ten geographic locations (points) are randomly selected in each area for temperature prediction. Figure 10 displays the distribution of the prediction points.

Figure 11 offers frozen ground temperature curves at the prediction points. It demonstrates that temperature is lower in high-altitude mountainous areas than in low-altitude plains and basins (see figure 11(a)); the Kunlun Mountains area has the lowest frozen ground temperature, which is stable between −3 °C and −4 °C; frozen ground temperature in the Hoh Xil Mountain and the Fenghuo Mountain areas is −1 °C to −3 °C; and the temperature in other areas is relatively high (see figure 11(b)).

Figure 12 shows frozen ground temperature curves at all the six prediction areas separately (see Figure 12 for geographic distribution). The temperature changes by depth in the Chumar River High Plain (figure 12(b)), the Hoh Xil Mountain (figure 12(c)), the Beilu River Basin (figure 12(d)), and the Chiqu Valley (figure 12(f)) are relatively small, which may be due to the homogeneous soil lithology in these areas. However, the ground temperature varies by depth significantly in the Kunlun Mountains (figure 12(a)) and the Fenghuo Mountain (figure 12(e)).

To recap, the above findings are reasonable and agree with reality, thereby supporting the plausibility of the proposed methodology (PCA + GWR).

5. Conclusions
In this study, we combine PCA and the GWR model to predict the spatial distribution of frozen ground temperature. The GWR model is used to address spatially non-stationary relationships between the dependent variable (frozen ground temperature) and
covariates (e.g., latitude, elevation, and air temperature), while PCA is used to relieve the collinearity (or correlation between covariates) problem.

The model provides reasonable results of frozen ground temperature. Moreover, through LOOCV and the comparison between the measured temperature data from several borehole points with temperature predictions of nearby points, this study shows that the GWR model performs reasonably well in predicting the spatial distribution of frozen ground temperature. This study provides the following findings: (a) frozen ground temperature in the study area is predicted to have a trend of decreasing with increasing altitude and increasing underground depth; (b) frozen ground temperature of mountainous areas is lower than that of high plains and basins; (c) in the same altitude, lakes and rivers' temperature is relatively high; (d) the temperature fluctuates greatly when the underground depth of the study area is within 5 m, and the maximum difference is 18 °C; and (e) when the depth is over 5 m, the ground temperature remains nearly unchanged.

The frozen ground temperature analysis provides insights into the determination of 0 °C isotherms and the monitoring of the change in ground thickness. In a departure from the existing literature that heavily relies on numerical modeling (or computational simulation) to analyze frozen ground temperature at the vertical level, this study mainly focuses on frozen ground temperature at the horizontal level and thus tries to use a statistical model that addresses spatial non-stationarity (the GWR model) to explore the spatial distribution of the temperature. Admittedly, this study is not so ambitious to provide a definite conclusion on the issue. Instead, it intends to offer a new angle or perspective and arouse attention to the studied problem from, inter alia, researchers and decision makers. The methodology itself is, of course, open to debate. Moreover, this study has other limitations that highly deserve further investigation. First, mainly restricted by data availability, this study only chooses June–September, 2001, for the frozen ground temperature prediction. The generalizability of our approach can be challenged. We, therefore, hope that more studies can be conducted to test the applicability of our approach and reach a firmer conclusion. Second, this study selects 12 variables for frozen ground modeling. Therefore, more temperature-influencing variables can be adopted in future research to optimize the GWR modeling and improve the accuracy of prediction. Third, our borehole points are highly concentrated on a line (i.e. the Qinghai–Tibet Railway), which may modestly contaminate temperature predictions. Last, this study uses the traditional GWR framework. Some enhanced versions of GWR models, such as semi-parametric GWR (differentiating local and global variables) and multiscale GWR (allowing the covariate-specific bandwidth) (Fotheringham et al 2017, Yu et al 2020), have been proposed in recent years. They may enhance the prediction accuracy and therefore can be tested in future research.

Data availability statement

The data generated and/or analyzed during the current study are not publicly available for legal/ethical reasons.

Acknowledgments

This research was supported by the National Key Research and Development Program of China (2017YFA0603101), National Natural Science Foundation of China (No. 41571520), Sichuan Young Talent Scientific Funding (No. 2019JDJQ0020), Sichuan Province Cyclic Economy Research Center (XHJJ-2002; XHJJ-2005), and Sichuan Science and Technology Program (2020JDTD0003). The air temperature dataset is provided by the National Tibetan Plateau Data Center (http://data.tpdc.ac.cn).
Appendix

Figure A1. Distribution of time-constant variables.

Figure A2. Distribution of time-varying variables.
References

Brunsdon C, Fotheringham A S and Charlton M E 1996 Geographically weighted regression: a method for exploring spatial nonstationarity Geogr. Anal. 28 281–98
Chen Y, Yang K, He J, Qin J, Shi J, Du J and He Q 2011 Improving land surface temperature modeling for dry land of China J. Geophys. Res.: Atmos. 116 D20104
Cheng G 2004 Influences of local factors on permafrost occurrence and their implications for Qinghai–Xizang Railway design Sci. China Ser. D 47 704–9
Cheng G and Jin H 2013 Permafrost and groundwater on the Qinghai–Tibet Plateau and in northeast China Hydrogeol. J. 21 3–23
Chou Y, Sheng Y and Zhu Y 2012 Study on the relationship between the shallow ground temperature of embankment and solar radiation in permafrost regions on Qinghai–Tibet Plateau Cold Reg. Sci. Technol. 78 122–30
Ding Y and Peng S 2020 Spatiotemporal trends and attribution of drought across China from 1901–2100 Sustainability 12 477
Dormann C F, Elith J, Bacher S, Buchmann C, Carl G, Carr G, Ding Y and Peng S 2020 Spatiotemporal trends and attribution of drought across China from 1901–2100 Int. J. Climatol. 38 2250–64
Prat P and Huang H 2015 Modeling crash spatial heterogeneity: random parameter versus geographically weighting Accid. Anal. Prev. 75 16–25
Xu W A and Yang L 2019 Evaluating the urban land use plan with transit accessibility Sustain. Cities Soc. 45 474–85
Yang H, Lu X, Cherry C, Liu X and Li Y 2017 Spatial variations in active mode trip volume at intersections: a local analysis

ORCID iD

Linchuan Yang https://orcid.org/0000-0001-6070-9044

of Northern Engineering University of Alaska Fairbanks) pp 1125–30
Mu C, Schuster P F, Abbott B W, Kang S, Guo J, Sun S and Zhang T 2020 Permafrost degradation enhances the risk of mercury release on Qinghai–Tibetan Plateau Sci. Total Environ. 708 135127
Mu Y, Li G, Yu Q, Ma W, Wang D and Wang F 2016 Numerical study of long-term cooling effects of thermosyphons around tower footings in permafrost regions along the Qinghai–Tibet Power Transmission Line Cold Reg. Sci. Technol. 121 237–49
Nguyen V C and Ng C T 2020 Variable selection under multicollinearity using modified log penalty J. Appl. Stat. 47 201–30
Osterkamp T E 2007 Characteristics of the recent warming of permafrost in Alaska J. Geophys. Res. Earth. 112 F02S05
Park H, Shershukov A B, Fedorov A N, Polyaakov I V and Walsh J E 2014 An observation-based assessment of the influences of air temperature and snow depth on soil temperature in Russia Environ. Res. Lett. 9 064026
Peng S, Ding Y, Wen Z, Chen Y, Gao Y and Ren J 2017 Spatiotemporal change and trend analysis of potential evapotranspiration over the Loess Plateau of China during 2011–2010 Agric. For. Meteorol. 233 183–94
Peng S, Gang C, Gao Y and Chen Y 2018 Assessment of climate change trends over the Loess Plateau in China from 1901 to 2100 Int. J. Climatol. 38 2250–64
Peng S, Ding Y, Liu W and Li Z 2019 1 km monthly temperature and precipitation dataset for China from 1901 to 2017 Earth Syst. Sci. Data 11 1931–46
Qin Y, Hiller J E, Jiang G and Bao T 2013 Sensitivity of temperature parameters affecting cold-region ground-temperature predictions Environ. Earth Sci. 68 1757–72
Qin Y, Wu T, Zhao L, Wu X, Li R, Xie C and Liu G 2017 Numerical modeling of the active layer thickness and permafrost thermal state across Qinghai–Tibetan Plateau J. Geophys. Res.: Atmos. 122 11604–20
Sheng Y and Yao T 2009 Integrated assessments of environmental change on the Tibetan Plateau Environ. Res. Lett. 4 045201
Smith M W and Riseborough D W 1996 Permafrost monitoring and detection of climate change Permafrost Periglac. Environ. 7 301–9
Thurow J, Lewkowicz A G and Smith S L 2012 Climate and ground temperature relations at sites across the continuous and discontinuous permafrost zones northern Canada Can. J. Earth Sci. 49 865–76
Ul-Saufie A Z, Yahaya A S, Ramlit N A, Rosaida N and Hamid H A 2013 Future daily PM10 concentrations prediction by combining regression models and feedforward backpropagation models with principle component analysis (PCA) Atmos. Environ. 77 621–30
Vasiliev A A, Drozdov D S, Gravis A G, Malkova G V, Nyland K E and Münkemüller T 2013 Collinearity: a review of methods and discontinuous permafrost zones northern Canada Can. J. Earth Sci. 49 865–76
Zhu A Z, Yahaya A S, Ramlit N A, Rosaida N and Hamid H A 2013 Future daily PM10 concentrations prediction by combining regression models and feedforward backpropagation models with principle component analysis (PCA) Atmos. Environ. 77 621–30
Vasiliev A A, Drozdov D S, Gravis A G, Malkova G V, Nyland K E and Münkemüller T 2013 Collinearity: a review of methods

NOTES

---

1. The references include a variety of scientific studies and reviews that focus on permafrost dynamics, climate change, and its impacts on environmental and engineering aspects in the Qinghai-Tibet Plateau region of China. The studies cover topics such as temperature modeling, permafrost degradation, and the effects of human activities on permafrost stability.

2. The references span from 1996 to 2017, indicating a significant body of research in this field.

3. The references are cited in a format that includes the title of the work, authors, publication details (journal, volume, issue, pages), and year of publication. This is a standard academic citation style.

---

1. Marchenko S, Romanovsky V, and Tipenko G (2008) Numerical modeling of spatial permafrost dynamics in Alaska. Proc. of the Ninth International Conf. on Permafrost vol 29 (Institute of Northern Engineering University of Alaska Fairbanks) pp 1125–30

2. Mu, C., Schuster, P. F., Abbott, B. W., Kang, S., Guo, J., Sun, S., and Zhang, T. (2020) Permafrost degradation enhances the risk of mercury release on Qinghai–Tibetan Plateau. Sci. Total Environ. 708, 135127.

3. Mu, Y., Li, G., Yu, Q., Ma, W., Wang, D., and Wang, F. (2016) Numerical study of long-term cooling effects of thermosyphons around tower footings in permafrost regions along the Qinghai–Tibet Power Transmission Line. Cold Reg. Sci. Technol. 121, 237–49.

4. Nguyen, V. C. and Ng, C. T. (2020) Variable selection under multicollinearity using modified log penalty. J. Appl. Stat. 47, 201–30.

5. Osterkamp, T. E. (2007) Characteristics of the recent warming of permafrost in Alaska. J. Geophys. Res. Earth. 112, F02S05.

6. Park, H., Shershukov, A. B., Fedorov, A. N., Polyaakov, I. V., and Walsh, J. E. (2014) An observation-based assessment of the influences of air temperature and snow depth on soil temperature in Russia. Environ. Res. Lett. 9, 064026.

7. Peng, S., Ding, Y., Wen, Z., Chen, Y., Gao, Y., and Ren, J. (2017) Spatiotemporal change and trend analysis of potential evapotranspiration over the Loess Plateau of China during 2011–2010. Agric. For. Meteorol. 233, 183–94.

8. Peng, S., Gang, C., Gao, Y., and Chen, Y. (2018) Assessment of climate change trends over the Loess Plateau in China from 1901 to 2100. Int. J. Climatol. 38, 2250–64.

9. Peng, S., Ding, Y., Liu, W., and Li, Z. (2019) 1 km monthly temperature and precipitation dataset for China from 1901 to 2017. Earth Syst. Sci. Data 11, 1931–46.

10. Qin, Y., Hiller, J. E., Jiang, G., and Bao, T. (2013) Sensitivity of temperature parameters affecting cold-region ground-temperature predictions. Environ. Earth Sci. 68, 1757–72.

11. Qin, Y., Wu, T., Zhao, L., Wu, X., Li, R., Xie, C., and Liu, G. (2017) Numerical modeling of the active layer thickness and permafrost thermal state across Qinghai–Tibetan Plateau. J. Geophys. Res.: Atmos. 122, 11604–20.

12. Sheng, Y. and Yao, T. (2009) Integrated assessments of environmental change on the Tibetan Plateau. Environ. Res. Lett. 4, 045201.

13. Smith, M. W. and Riseborough, D. W. (1996) Permafrost monitoring and detection of climate change. Permafrost Periglac. Environ. 7, 301–9.

14. Thurow, J., Lewkowicz, A. G., and Smith, S. L. (2012) Climate and ground temperature relations at sites across the continuous and discontinuous permafrost zones northern Canada. Can. J. Earth Sci. 49, 865–76.

15. Ul-Saufie, A. Z., Yahaya, A. S., Ramlit, N. A., Rosaida, N., and Hamid, H. A. (2013) Future daily PM10 concentrations prediction by combining regression models and feedforward backpropagation models with principle component analysis (PCA). Atmos. Environ. 77, 621–30.

16. Vasiliev, A. A., Drozdov, D. S., Gravis, A. G., Malkova, G. V., Nyland, K. E., and Münkemüller, T. (2013) Collinearity: a review of methods.

---

IOP Publishing

Environ. Res. Lett. 16 (2021) 024003

R Zhao et al.
utilizing geographically weighted regression. *J. Transp. Geogr.* 64 184–94.

Yang K, He J, Tang W, Qin J and Cheng C C 2010 On downward shortwave and longwave radiations over high altitude regions: observation and modeling in the Tibetan Plateau. *Agric. For. Meteorol.* 150 38–46.

Yang L, Chau K W, Szeto W Y, Cui X and Wang X 2020a Accessibility to transit by transit and property prices: spatially varying relationships. *Transp. Res. D* 85 102387.

Yang L, Chu X, Gou Z, Yang H, Lu Y and Huang W 2020b Accessibility and proximity effects of bus rapid transit on housing prices: heterogeneity across price quantiles and space. *J. Transp. Geogr.* 88 102850.

Yang H, Zhang Y, Zhong L, Zhang X and Ling Z 2020c Exploring spatial variation of bike sharing trip production and attraction: A study based on Chicago's Divvy system. *Appl. Geogr.* 115 102130.

Yi S, Zhou Z, Ren S, Xu M, Qin Y, Chen S and Ye B 2011 Effects of permafrost degradation on alpine grassland in a semi-arid basin on the Qinghai–Tibetan Plateau. *Environ. Res. Lett.* 6 045403.

Yin G, Zheng H, Niu F, Luo J, Lin Z and Liu M 2018 Numerical mapping and modeling permafrost thermal dynamics across the Qinghai–Tibet engineering corridor China integrated with remote sensing. *Remote Sens.* 10 2069.

Yu H, Fotheringham A S, Li Z, Oshan T, Kang W and Wolf L J 2020 Inference in multiscale geographically weighted regression. *Geogr. Anal.* 52 87–106.

Yu W B, Liu W B, Bai Y M, Chen L and Yi X 2014 Nonlinear analysis of coupled temperature-seepage problem of warm oil pipe in permafrost regions of Northeast China. *Appl. Therm. Eng.* 70 985–95.

Zhang T 2005 Influence of the seasonal snow cover on the ground thermal regime: an overview. *Rev. Geophys.* 43 RG4002.

Zhang T and Stammes K 1998 Impact of climatic factors on the active layer and permafrost at Barrow Alaska. *Perm. Perigl.* 9 229–46.

Zhang Y, Sherstiukov A B, Qian B, Kokelj S V and Lantz T C 2018 Impacts of snow on soil temperature observed across the circumpolar north. *Environ. Res. Lett.* 13 044012.

Zhao R, Zhan L, Yao M and Yang L 2020 A geographically weighted regression model augmented by Geodetector analysis and principal component analysis for the spatial distribution of PM2.5. *Sustain. Cities Soc.* 56 102106.

Zou D, Zhao L, Yu S, Chen J, Hu G, Wu T and Wang W 2017 A new map of permafrost distribution on the Tibetan Plateau. *Cryosphere* 11 2327.