Paleotopography continues to drive surface to deep-layer interactions in a subtropical Critical Zone Observatory

Xiao-Dong Song [a], Hua-Yong Wu [a], Paul D. Hallett [b], Xi-Cai Pan [a], Xue-Feng Hu [c], Qi Cao [d], Xiao-Rui Zhao [a], Gan-Lin Zhang [a, e *]

[a] *State Key Laboratory of Soil and Sustainable Agriculture, Institute of Soil Science, Chinese Academy of Sciences, Nanjing 210008, China*

[b] *School of Biological Sciences, University of Aberdeen, Aberdeen AB24 3UU, United Kingdom*

[c] *Department of Environmental Science and Engineering, School of Environmental and Chemical Engineering, Shanghai University, Shanghai 200444, China*

[d] *Environmental Protection Agency of Jiujiang, Jiujiang 332000, China*

[e] *University of Chinese Academy of Sciences, Beijing 100049, China*

* Corresponding author: Gan-Lin Zhang (glzhang@issas.ac.cn)

Competing interests

The authors declare no competing interests.

Abstract: Subsurface critical zone structures (SCZ_S) refer to the spatial variation in the interactive layers underground. Although SCZ_S greatly affect terrestrial biogeochemical and hydrological cycles, underpinning mechanisms are poorly documented. Herein, we characterized the SCZ_S of a typical red soil in subtropical China, a type of soil with vast global distribution. The thickness information of three layers was derived from hand augers, boreholes and ground-penetrating radar (GPR) radargrams and incorporated into geographically weighted regression (GWR) models for the reconstruction of paleotopography (Cretaceous sandstone). The interpreted GPR results in terms of thicknesses and interfaces for the three layers were consistent with the borehole logs. The trained GWR models accounted for 43%-77% of the spatial variations in the three layers. The paleotopographic elevations were highly correlated with those of the current land surface (r=0.85). Spatial analysis showed that the rougher paleotopography was inherited by the current landform. The SCZ_S evolution involving mainly the mantling covered by Quaternary red clay (QRC) was primarily driven by terrain attributes. These findings may enhance our understanding of the interaction between the paleoclimate and paleoenvironment. The combination of geophysical techniques, geochemical indicators and spatial prediction techniques provides an effective tool for understanding QRC landform evolution.

Keywords: Critical zone; Paleotopography; Ground-penetrating radar; Red Soil Critical Zone Observatory; Landscape evolution

1. Introduction

Paleotopography has been linked to landscape evolution for processes ranging from the rapid colluviation of lacustrine sediments by landslides (Bièvre *et al*., 2011) to the formation of loess soils by surface deposition and erosion (Xiong *et al*., 2016a). In ancient paleosols, where the soils have been protected from chemical and physical weathering, links to paleotopography are strong (Torres and Gaines, 2011). In locations where the paleotopography has been established more recently, such as the post-glacial landscapes of the late Pleistocene, the impacts on soil evolution are evident (Schaetzl *et al*., 2000). In tropical landscapes, the effects of paleotopography on modern landscapes are expected to be less evident. Tropical monsoonal rains and greater soil temperatures drive physical and chemical weathering that transform and move minerals, producing highly weathered environments that have deep and erosion-prone soils. Although the paleotopography in tropical regions has been linked to historical geological events (Wichura *et al*., 2010) and the geochemistry of lateritic paleosols (Cecil *et al*., 2006), it has not been associated with shallower subsurface layers.

Red soils cover approximately 2.18×10^6 km^2 in China, which accounts for 22.7% of the country's territory (Zhao *et al*., 2013). They are ultisols according to the USDA Soil Taxonomy (Soil Survey Staff, 2010) and are dominated by red clay that can be sedimentary red clay or may be derived from underlying parent rocks (Hu *et al*., 2010). The sedimentary red clay was formed during the Quaternary period (Zhao, 1992) and, thus, has been named Quaternary red clay (QRC), which is also widely distributed in other parts of the world (Muggler *et al*., 2001; Tanner and Lucas, 2018).

QRC forms from the interactions between desilicification-allitization and biological enrichment processes (Hu *et al*., 2014). Extensive studies have been conducted on the formation and evolutionary processes of QRC (Li *et al*., 2013; Xiong *et al*., 2002; Zhu, 1988) and the paleoclimatic implications of QRC (Hu *et al*., 2010, 2014). Generally, three layers can occur in QRC: yellow-brown earth, uniform red clay (URC) and reticulate red clay (RRC) (Hu *et al*., 2010). Cretaceous sandstone is the underlying paleotopography of QRC (Tang *et al*., 2008; Wu *et al*., 2019). The spatial distribution of each layer's thickness can be used to interpret the possible paleoclimatic records. From the perspective of geomorphological inheritance (Coventry, 1982; Olyphant *et al*., 2016; Xiong *et al*., 2016a), the evolution of the QRC landform may be greatly affected by the paleotopography. However, to the best of our knowledge, little focus has been given to identifying this control due to the lack of detailed paleotopographic information.

There may be three scenarios of geomorphological evolution related to landform inheritance (Fig. 1). The current landform may be subparallel (Fig. 1a), more rugged (Fig. 1b) or smoother (Fig. 1c) than the topography of the underlying strata (Xiong *et al*., 2016a). The weathered bedrock could be thick at the upslope due to the control of the underlying fresh bedrock, wherein the uppermost elevation of undrained fresh bedrock mainly results in a boundary of unweathered bedrock (Rempe and Dietrich, 2014). A weathered zone or soil mantle could thicken downslope due to soil erosion and wind erosion, such as the gully erosion in the Chinese Loess Plateau (Liu *et al*., 2018). If paleotopography still influences QRC landscapes, the spatial variation in subsurface layers could be larger than expected. This discrepancy would have a large impact on processes occurring in the critical zone, defined as the part of the Earth's surface spanning from the vegetation canopy to the bottom of the groundwater (National Research Council, 2001; Richter and Mobley, 2009). Subsurface critical zone structures (SCZ_S) include the composition and distribution of soil and saprolite (Holbrook *et al*., 2014; Wilford *et al*., 2016; Xu and Liu, 2017) and refers to the depth variability in different layers. SCZ_S greatly affect the physical/chemical processes of the critical zone. These impacts have been widely studied (Orlando *et al*., 2016; Scarpone *et al*., 2016; Xu and Liu, 2017). Identifying controls on SCZ_S enhances the understanding of the complicated coupling of geobiological, geochemical and hydrological processes occurring within the critical zone (Orlando *et al*., 2016).

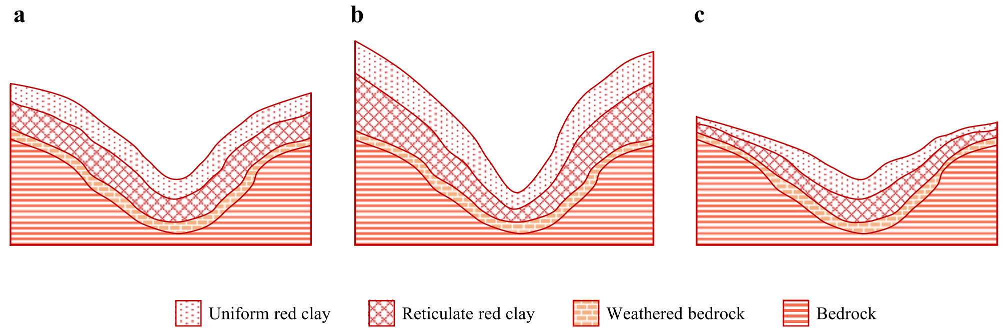


Fig. 1. A schematic diagram of the boundaries between URC, RRC, weathered bedrock and bedrock: current landform is subparallel to the topography of the underlying strata (a), current landform is more rugged than the topography of the underlying strata (b) and current landform is smoother than the topography of the underlying strata (c).

Using non-invasive detection technology, geophysical methods can be performed near the surface and are widely used to image spatial information beneath landscapes (Guo and Lin, 2016; Parsekian *et al*., 2015; Tian *et al*., 2019). Ground-penetrating radar (GPR) may be one of the most robust geophysical tools that can image the structures of subsurface strata in critical zone science (Kaufmann *et al*., 2018; Parsekian *et al*., 2015). Therefore, GPR has been widely utilized for the retrieval of regolith thickness, the bedrock-regolith interface and soil depth (Orlando *et al*., 2016; Simeoni *et al*., 2009; Tian *et al*., 2019). In a recent study, a 100 MHz antenna was used to image peat thicknesses exceeding 5 m at a centimetre spatial resolution (Comas *et al*., 2015). GPR with a low frequency (i.e., 50 MHz) was utilized to image the spatial distribution of subsurface materials at 15-20 m depth (Orlando *et al*., 2016).

Using a combination of drilling and geophysical techniques, the objective of this study was to explore the link between paleotopography and the subsurface layers of a lateritic soil in the Red Soil Critical Zone Observatory (Red Soil CZO), China. Drillings were used to characterize the SCZ_S, and geophysical techniques (GPR) were used to image the subsurface structure along transects of different land use types. The Cretaceous paleotopography was reconstructed based on the estimated thicknesses of different layers across the landscape, and the effect of the paleotopography on the SCZ_S was identified by regression analysis. Fresh bedrock was considered to be the boundary of the terrestrial biological, chemical and physical processes.

2. Materials and Methods

2.1. Study area

The Red Soil CZO, often referred to as the Sunjia CZO, is in Yujiang County, China, which has an area of 51 ha (28°14′ N, 116°53′ E) (Fig. 2). The elevation ranges from 41 to 55 m, and the slope varies from 0 to 5.5°. The study area has a subtropical monsoon climate with a mean annual air temperature of 17.8 °C. There are approximately 272 frost-free days. The mean annual precipitation is 1795 mm, 48% of which is observed during spring and summer (April to July), and the annual evaporation amount is 1229 mm (Gao *et al*., 2016). Due to drought stress during the summer/autumn dry season (Peng *et al*., 2016), flood irrigation sourced from the Baita River provides the agricultural water supply. Due to the fragmented ownership of land property rights, the area of farmlands varies from approximately 0.05 to 1.5 ha, and therefore, various land use types are distributed in the area (Fig. 3). Approximately 37.5% of this area is covered by upland (rainfed cropland), followed by old paddy fields (cultivation duration greater than 100 years) at 18.1% (Wu *et al*., 2019).

The SCZ_S in the Red Soil CZO are divided into three layers (Tang *et al*., 2008; Wu *et al*., 2019): URC, RRC and weathered bedrock. The parent material beneath is Cretaceous sandstone. Yellow-brown earth is lacking in this area. As the uppermost layer, the URC is mainly composed of soil A (topsoil strongly affected by vegetation) and B (weathered subsoil) horizons, wherein the original bedrock structure has been thoroughly broken down by pedogenesis. The RRC is also referred to as the ancient weathering crust of red soils, which was formed by weathering and sedimentation during the Quaternary. The thickness of URC is generally less than 2 m, whereas the RRC is approximately 2-10 m thick (Hu *et al*., 2010; Xiong *et al*., 2002). The URC and RRC are mainly clay loam to clay in texture (Wu *et al*., 2019). Clay minerals are dominated by kaolinite, followed by vermiculite and hydromica (Tang *et al*., 2008). The URC is characterized by granular aggregates, and the RRC has a blocky structure and red mottles and nodules. Different from the URC, the RRC is colourful and consists of speckled, worm-like and irregular striped patterns (Fig. 1d).

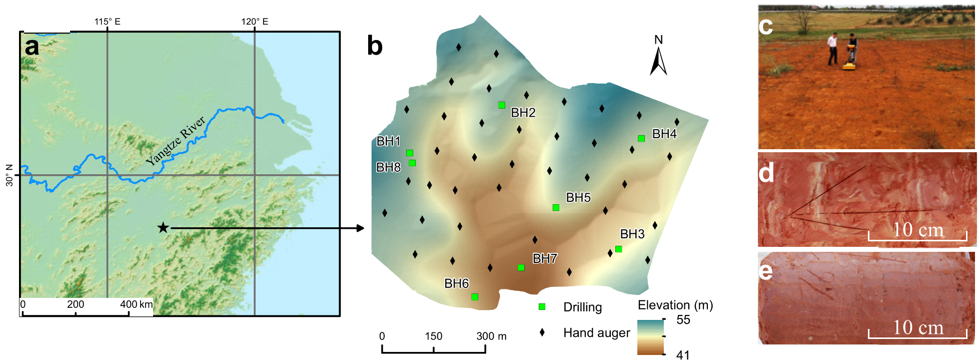


Fig. 2. Study area of the Red Soil CZO and core samples: regional landform (a), elevation and drilling sites in the study area (b), upland landscape of the red soil (c), a typical RRC sample (d) and a weathered layer (e). Note that the mud was scraped from the borehole samples (d-e).

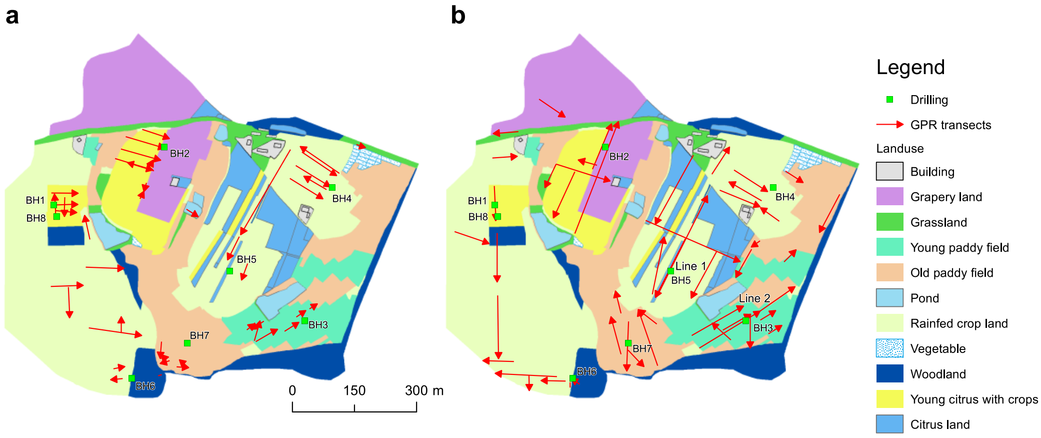


Fig. 3. Spatial distribution of the ground penetrating radar profiles surveyed by the US Radar with a 250 MHz shielded antenna in November 2016 (a) and by the Geoscanner/AKULA-9000C with 60, 120 and 200 MHz unshielded antennae in January 2018 (b). Note that old paddy fields indicate cultivation durations of greater than 100 years.

2.2. Drilling

Hand augering and drilling were conducted at different depths. For hand augering, the samples were taken on a 100 m by 100 m grid to capture the soil variation (Zhu *et al.*, 2010) (Fig. 2). In this area, farmland ridges (approximately 20–40 cm high) were built for irrigation and walking, which might greatly affect the observation of RRC. Although our survey was conducted outside the paddy growing season, some paddy fields were still under waterlogged conditions where the flooded mud soils prohibited the soil sampling. Therefore, some points were inaccessible, and thus five different points were sampled at sites having the same land use and slope position. Finally, 39 sites were visited (Fig. 2b). At each site, three auger borings were conducted to 1 m depth within an area of 4 m^2, and composited soil cores were taken at each depth increment. The topsoil was densely sampled (0-0.05 m, 0.05-0.15 m and 0.15-0.3 m) to account for the strong vertical soil variation (Gao *et al.*, 2015), and the subsoil was sampled at depth increments of 0.3-0.6 m and 0.6-1 m. RRC was observed at 16 sites, and the depths of RRC for these sites were recorded in the field. These observations were meaningful for the spatial prediction of the SCZ_S due to the limited number of drillings.

A hydraulic rotary drill rig was used at 8 sites in April and November 2016 to the depth of fresh bedrock (intact rock) without weathered material (Fig. 3). The drilling locations were selected through a subjective sampling strategy (Zhu *et al*., 2010), in which the catenary sequence, land use and cost-effectiveness were considered. The drill was equipped with a wireline core barrel with a 0.13 m diameter. A road passed through the grassland in the northern part (Fig. 3), and the grassland field had been greatly affected by human activity. Therefore, we did not make a borehole in the grassland field. These boreholes were named sequentially according to the sampling time. After the drilling of BH7, we found that the SCZ_S greatly varied in space. Because limited boreholes were conducted, the horizontal distances between boreholes were greater than 150 m. Thus, to investigate the variation in SCZ_S at a local scale, BH8 was sampled at the site approaching BH1 (about 30 m) (Fig. 2b). BH1 and BH8 had the same elevation, slope position and land use. The drilling numbers from upland (rainfed cropland), paddy field and woodland were 5, 2 and 1, respectively (Fig. 3). One pit was dug to 1 m depth close to every borehole, and samples were collected vertically every 0.1 m, as samples within the top 1 m of soil were easily fractured by drilling. Drill core samples were collected vertically in thickness increments of 0.2 m. The outer 1 cm layer of the drill core samples was scraped off to avoid potential contamination caused by the drill bit. Since the excavated sandstone disintegrated to mud at normal air temperature, the samples were temporarily stored in an ice-filled cooler in the field. More detailed information on the drillings can be found in Wu *et al*. (2019).

Soil samples were air-dried and sieved through a 2 mm mesh after sample pounding, and some samples were cryopreserved by freezing to -20 °C. The contents of major oxides (Al_2O_3, Na_2O, K_2O and CaO) were analysed with an ICP-AES (Optima 8000, PerkinElmer, Waltham, USA). The detailed procedure for sample dissolution can be found in Chen *et al*. (2011). The chemical weathering intensity of the drill core samples was determined by the chemical index of alteration (CIA) (Nesbitt and Young, 1982) as follows:

$$\mathrm{CIA} = \frac{100 \times \mathrm{Al_2O_3}}{\mathrm{Al_2O_3} + \mathrm{Na_2O} + \mathrm{K_2O} + \mathrm{CaO^*}} \tag{1}$$

where all variables are the molecular proportions of the oxides and CaO* represents only the fraction of CaO in silicates.

2.3. Paleotopographic reconstruction

The spatial prediction of each layer's thickness involved two phases (Fig. 4). The first step was imaging the SCZ_S based on GPR surveys. Descriptive information on the drill core samples was the auxiliary input for the interpretation of radargram signals. In the second step, the thickness of each layer was selected from the interpreted radargrams. Then, geographically weighted regression (GWR) models were trained to predict the three layers' thicknesses at the unvisited sites, and the paleotopography was reconstructed based on these thickness maps.

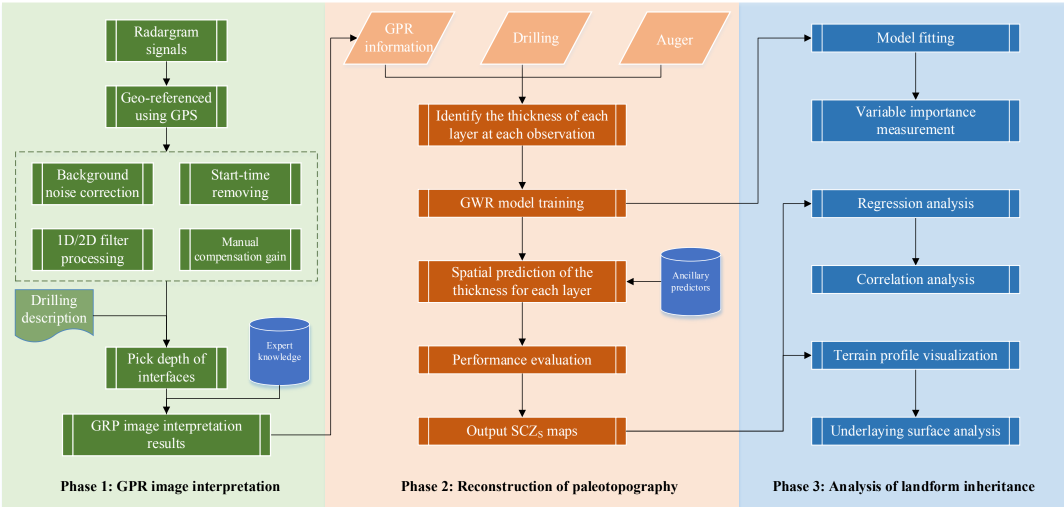


Fig. 4. Workflow of the paleotopographic reconstruction. GPR: ground-penetrating radar; GPS: global positioning system; GWR: geographically weighted regression; SCZ$_S$: subsurface critical zone structure.

2.3.1. Ground-penetrating radar

In this study, the subsurface structure of the QRC (i.e., URC and RRC), rather than the groundwater thickness, was retrieved by GPR because the vertical thickness of the aquifer was not available and varied seasonally (Gao *et al*., 2015). The GPR signals are electromagnetic waves with frequencies varying from 10 to 1000 MHz. The electromagnetic waves are emitted by a transmitting antenna and propagated through a conductive material (Kaufmann *et al*., 2018). The propagation velocity of the electromagnetic waves (v) (m/ns) can be calculated as follows:

$$v = \frac{c}{\varepsilon^{1/2}} \tag{2}$$

where ε is the relative permittivity (dimensionless) and c is the speed of light in free space (0.2998 m/ns). Different parts of the QRC are characterized by different magnetic properties. Thus, the signals will be partially reflected and received by a receiving antenna, and the two-way travel time is usually described in the y-direction. Different amplitudes of reflected signals can be used to characterize the subsurface structure (Tian *et al*., 2019). A high frequency leads to a fine resolution with a shallow penetration depth, and vice versa. The propagation velocities vary in different layers because of the changes in electrical conductivity, especially in the deep critical zone.

Consequently, GPR measurements were obtained at different frequencies and were surveyed under two modes. A common offset mode survey was conducted to image the subsurface structure, where the transmitting antenna and receiving antenna simultaneously moved towards the same direction. The common midpoint mode was utilized to estimate the propagation velocity, v, as a one-dimensional (1D) model, where the transmitting antenna and receiving antenna shifted in opposite directions from the midpoint (Orlando *et al*., 2016). The GPR survey was performed under common midpoint mode near every drilling. The offset was approximately 10 m, and the steps of the transmitting antenna and receiving antenna moving around the midpoint were 0.05 m. Since the propagation velocity was very sensitive to the soil moisture, the GPR survey was performed during dry seasons while the groundwater level was stable. A 1D model of the velocities was developed, and the mean velocity was utilized for radargram interpretation.

We used a US Radar with a 250 MHz shielded antenna in November 2016 and a Geoscanner/AKULA-9000C with 60, 120 and 200 MHz unshielded antennae in January 2018. The transmitting and receiving antennae were integrated together in these GPRs. Thus, two integrated 60 MHz antennae were prepared and separately acted as transmitters and receivers for the GPR survey under common midpoint mode. All of the antennae with different frequencies were measured at the eight drilling locations except BH7 with 250 MHz. Testing with different frequencies was used to verify whether the electromagnetic waves were naturally attenuated or influenced by the dielectric properties of the underlying material or texture. A total of 105 profiles with lengths of 7.0 km were collected and georeferenced by a handheld global positioning system (Fig. 3). The post-processing of the radargrams was performed in Reflexw 8.5 software and included background noise correction, start-time removal, 1D-Filter based on subtract-DC-shift, 2D-Filter based on running average and manual gain (Fig. 4).

2.3.2. Selection of points with thickness information from the radargram

The topography of unweathered sandstone bedrock (Cretaceous sandstone) was defined as the paleotopography in this study. The paleotopographic elevation could be calculated by subtracting the thicknesses of URC, RRC and weathered bedrock from the current measured elevation. Fig. 5 illustrates the interfaces between different underground layers which can be derived from a GPR common offset profile. The thickness values of different layers at any site could be easily obtained along these curves.

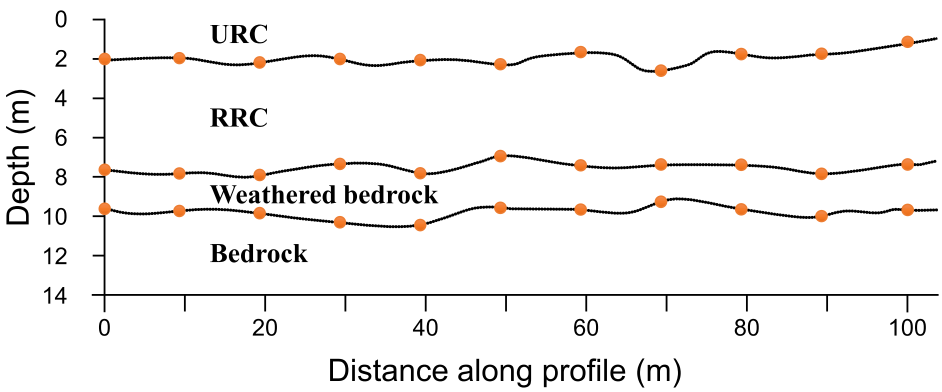


Fig. 5. A schematic diagram of the boundaries between URC, RRC, weathered bedrock and bedrock and manually selected points. Note that these boundaries (dotted lines) are interpreted from a GPR common offset profile.

After radargram post-processing, we manually selected points every 10 m along GPR profiles (Fig. 5). Each point had the thickness of the URC, RRC and weathered bedrock layers. The geographic positions of the start and end points of the radar profiles were recorded in the field. Thus, these selected points could be easily georeferenced and integrated into a geographic information system. These points were employed for the spatial prediction of each layer's thickness. A total of 420 points were selected and randomly divided into a calibration dataset (70%, N=294) for predictive model training and a validation dataset (30%, N=126) for performance assessment. In addition to the validation based on points interpreted by the radargram, the observed thickness data collected by hand augering (N=16) and drilling (N=8) were used to independently evaluate the fitted predictive models. It was noted that only the URC prediction model could be validated by the observations based on hand augering, as RRC was the lower boundary of URC and the thickness of URC can be confirmed only if the RRC was observed for 1-m hand auger samples. A one-way analysis of variance (ANOVA) with a confidence level of $p<0.05$ was conducted to test the significance of the land use type effects on URC thicknesses.

2.3.3. Covariates

In total, 16 independent variables were collected as covariates for modelling various maps. A contour terrain map (1:5000 scale) was utilized to create a digital elevation model (DEM) with a 5 m resolution. Ten terrain attributes were produced in System for Automated Geoscientific Analyses software (Conrad *et al*., 2015), including the elevation, slope, cosine of the aspect, plan curvature, profile curvature, relative slope position (ReSlpPosi), distance to the channel (DisToChan), convergence index, topographic wetness index (TWI) and potential insolation (incoming solar radiation). The TWI was calculated based on a modified catchment area algorithm (Conrad *et al*., 2015), which resulted in a high-potential and realistic soil wetness for the sites that were closer to a channel. Due to the small area, the climatic variables, such as annual air temperature and annual precipitation, could be deemed as homogeneous. Therefore, we adopted the potential insolation to account for the microclimatic conditions, which was mainly affected by the elevation, slope and aspect. The potential insolation expressed in kW m^{-2} was derived from the solar radiation function with ArcGIS 10.2 (ESRI, 2014).

Landsat 8 imagery was employed to represent the vegetation density in this area, which was acquired in June 2016 with little cloud cover (<10%). Five variables were derived from Landsat 8 images: the normalized difference vegetation index (NDVI), band 2, band 3, band 4 and band 8. NDVI has been widely used to indicate vegetation growth characteristics (Scarpone *et al*., 2016; Song *et al*., 2016). Band 2 (0.45-0.51 μm), band 3 (0.53-0.59 μm), band 4 (0.64-0.67 μm) and band 8 (0.50-0.68 μm) can be used to distinguish soil from vegetation, emphasize peak vegetation, discriminate vegetation types and combine visible colours into one channel, respectively.

Aerial photography was obtained using an unmanned aerial vehicle (DJI, Phantom 4 Pro, Shenzhen, China) that flew 500 m above the terrain. Each photograph covered approximately 16 ha, at least 20% of which overlapped with adjacent images to produce the mosaic. Following georeferencing, land use map polygons were digitized with ArcGIS 10.2, and the land use type was rectified through a field survey. The land use types were classified into five types according to the size of the area: citrus, upland (rainfed cropland), paddy field, vineyard and others (Fig. 3).

Predictors were resampled to a 5 m spatial resolution based on the bilinear interpolation method. Continuous predictors were normalized, and the average and standard deviation were 0 and 1, respectively. Categorical predictors were transformed into dummy variables during predictive model fitting. The best sets of predictors were selected by stepwise regression in both directions. The predictors that reduced the Akaike Information Criterion the most were selected, and the interaction between covariates was not considered to pursue a simple model. Notably, all of the environmental variables were considered for the prediction of URC, and only terrain attributes were considered for the other layers.

2.3.4. Predictive techniques

The spatial distributions of the thicknesses of URC, RRC and weathered bedrock (dependent variables) described in Sect. 2.3.2 (Fig. 5) were predicted by incorporating the covariates (independent variables) introduced in Sect.2.3.3 into the GWR models (Fig. 4). Given the geographical positions of the dependent variables, the values of covariates were derived from a geographic information system, as covariates covered the whole study area. The regolith thickness could be obtained by aggregating the thicknesses of the URC, RRC and weathered bedrock. However, prediction uncertainty might be involved in each prediction scenario, and its spatial pattern should be accounted for (Scarpone *et al*., 2016; Wilford *et al*., 2016; Xu and Liu, 2017; Zhu *et al*., 2010). Therefore, a model for regolith thickness was also trained to highlight the difference between the regolith thickness and the sum of three layer thicknesses in space (i.e., URC, RRC and weathered bedrock).

Thus, four GWR models were trained for the spatial prediction. The dependent variables would be transformed by a natural logarithm if the null hypothesis of normality was rejected through the Kolmogorov-Smirnov test (p<0.05). GWR could be referred to a local regression technique (Brunsdon *et al*., 1996) as follows:

$$Y_i = \beta_{i0} + \sum_k \beta_{ik} x_{ik} \quad (3)$$

where Y_i, x_{ik}, β_{i0} and β_{ik} are the dependent variable, the value for the kth independent variable, the estimated intercept and regression coefficients at location i, respectively. The regression coefficients are weighted by the observations around the predicted location i. A large weight will be fitted if a point approaches location *i*. Based on the calibration dataset (*N*=294), the GWR models were trained to quantify the spatially varying relationships between dependent variables and covariates at the visited sites (Eq. 3). Furthermore, the thickness values of each layer at the unvisited sites could be achieved by running the trained GWR models based on covariates. Predictions were also conducted for the sites of the validation dataset (*N*=126) and the observations of hand augering (*N*=16) and drilling (*N*=8). The model performance was evaluated by comparing the observations and predictions in terms of the mean error (ME), root mean square error (RMSE) and coefficient of determination (R^2).

Another important parameter for the weighting process is the kernel bandwidth. Unlike a fixed kernel with a constant bandwidth, an adaptive kernel will set a small bandwidth if the samples are densely distributed in space, and vice versa. We used an adaptive kernel bandwidth in which the optimal bandwidth was computed using the Akaike Information Criterion.

As predictors were scaled at the same magnitude, the absolute values of coefficients within the GWR models were used to indicate the relative importance of the predictors (Song *et al*., 2016). A variance inflation factor (VIF) was calculated to diagnose the collinearity. The GWR model performed a local regression for each point of the calibration dataset, in which the VIF for one predictor could be obtained. Thus, for one GWR model, there were 294 VIFs for one predictor. Here, the mean values were adopted for comparison. A VIF value of one predictor greater than 10 indicates a collinearity problem as a general rule. We visualized the maps using ArcGIS 10.2 and predicted the distribution of thickness using R (version 3.3.1, http://cran.r-project.org/) with the package “spgwr”.

2.4. Geomorphological evolution analysis

To investigate the effects of paleotopography on the geomorphological evolution of the QRC, a regression analysis was performed (Fig. 4). The topography of the RRC layer was also compared with the paleotopography, as the RRC was obviously different from the URC. If the topography of the paleotopography greatly shaped the current landform or the terrain of the RRC layer surface, the landform inheritance could be proven (Xiong *et al*., 2016a). Correlations between the topographies of different layers can be analysed by regression as follows:

$$y = c + Kx \tag{4}$$

where x denotes the elevation of the paleotopography, y denotes the elevation of current land surface or RRC layer and K is the regression coefficient. The correlation between y and x can be represented by K. If the value of K is 1, the different layers are parallel (Fig. 1a). K values greater than 1 or smaller than 1 can be used to demonstrate that the current land surface or RRC layer is more rugged (Fig. 1b) and smoother (Fig. 1c) than that of the paleotopography, respectively. In addition, the terrain profiles were extracted from the elevation maps of different layers, and the Pearson correlation coefficients of elevation values were used to analyse the relationships. The slope of the underlying strata was computed and compared with that of the current land surface to verify the terrain roughness changes.

3. Results

3.1. Characterizing the SCZ_S

The observed regolith thickness in the Red Soil CZO ranged from 3.49 to 9 m (Table 1). The thickest regolith was found in BH6 at the woodland while the thinnest regolith was observed at the ridge top (BH2). The thickness of the RRC ranged from 1.68 to 6.25 m and accounted for 40%-70% of the regolith.

The ANOVA results show that the thickness of the URC significantly differed with land use type (p<0.05).

The SCZ_S were semi-quantitatively verified by the vertical pattern of the weathering index based on the core samples. The decline pattern of CIA matched well with the observed SCZ_S (Fig. 6). Due to the dissolution and leaching of mobile elements, the CIA values ranged from 33% to 91% and sharply decreased at the vertical transition to weathered bedrock, except in BH6 located at a lower elevation in the CZO (Fig. 6). The CIA values were greater than 89% and 81% within the URC and RRC layers, respectively. It could be inferred that downwards through the subsurface layers, highly soluble elements, such as K and Na, were completely leached.

Table 1. Summarized information of eight boreholes.

Borehole	URC [a)] (m)	RRC (m)	Weathered bedrock (m)	Regolith (m)	Aspect	Slope (degree)	Land use
BH1	0.70	5.50	1.80	8.00	77.47	3.27	Upland
BH2	0.12	1.68	1.69	3.49	125.19	1.15	Upland
BH3	1.20	4.00	2.10	7.30	305.42	2.28	Young paddy field
BH4	0.60	3.90	1.45	5.95	152.70	1.15	Upland
BH5	1.40	4.70	1.00	7.10	181.39	2.29	Upland
BH6	1.95	6.25	0.80	9.00	59.72	2.49	Woodland
BH7	0.90	2.00	2.10	5.00	142.48	0.88	Old paddy field
BH8	0.70	5.30	1.60	7.60	77.18	3.42	Upland

[a)] URC: uniform red clay; RRC: reticulate red clay.

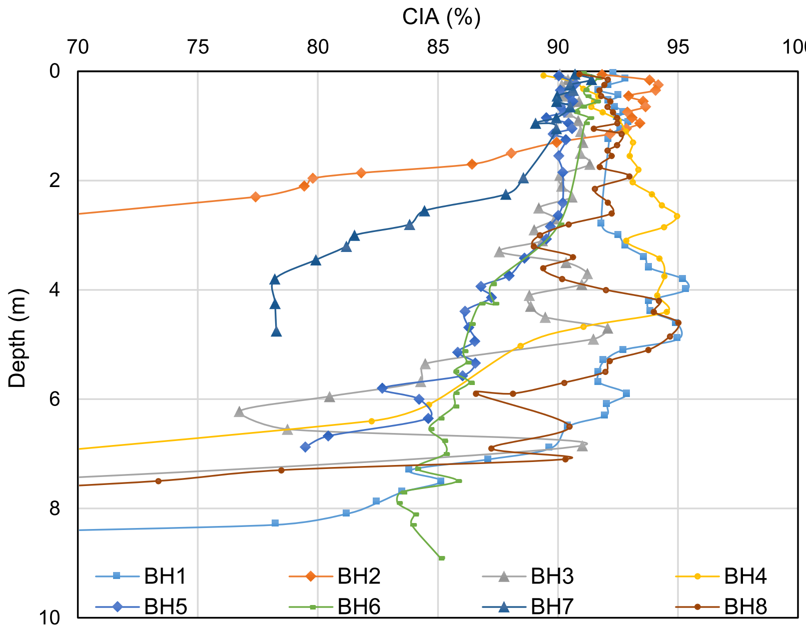


Fig. 6. Scatterplot of the chemical index of alteration (CIA) values for each borehole.

3.2. GPR surveys

GPR surveys were performed at different frequencies to retrieve the variations in SCZ_S at the transect scale. Based on the 1D model of velocities (Fig. 7), a mean velocity ($\bar{v}$) of 0.07 m/ns was utilized. The propagated velocity decreased with depth. Overall, the strata interpreted from the radargrams were subparallel to the terrain surface and showed few discrete hyperbolas, which implies the absence of faults, fractures and sag structures (Figs. 8-9).

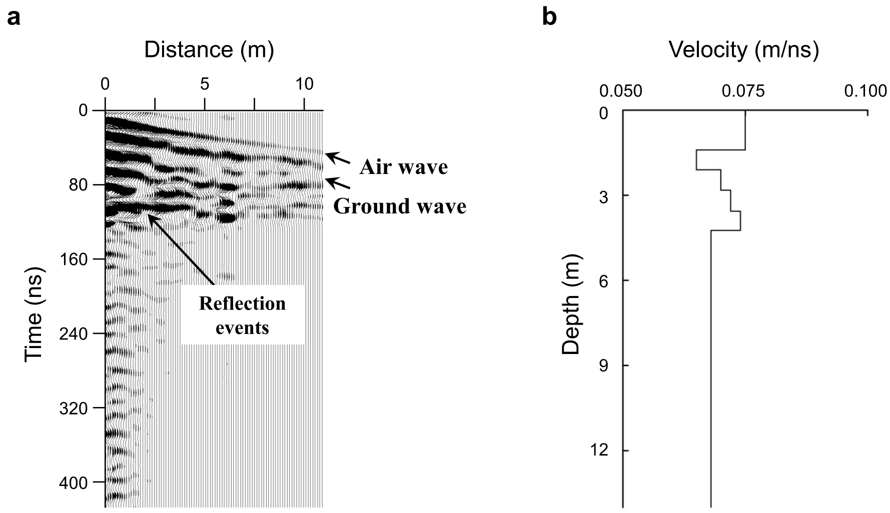


Fig. 7. Propagation velocity of electromagnetic waves interpreted by the common midpoint mode survey: GPR profile near borehole BH5 with a frequency of 60 MHz (a) and the interpreted 1D velocity model (b).

Fig. 8 shows the GPR profiles for the upland (Line 1 in Fig. 3b), which were southwest-facing and 150 m long and passed through BH5. Three antennae produced strong reflections at the interface between URC and RRC. The derived depth of the URC-RRC interface was below 1 m, which was consistent with the borehole logs. Continuous reflections of weathered bedrock were evidently distinguished by the 60 MHz antenna (Fig. 8c). The 60 MHz reflection became weak below 400 ns, which shows the homogeneity of the red sandstone. In contrast, reflections in the RRC and weathered bedrock became poor for the 200 and 120 MHz antennae due to the attenuation of electromagnetic waves in the wet strata and clay-rich material. The thicknesses of the URC, RRC and weathered bedrock layers ranged from 1.4-1.5 m, 4.3-5.1 m and 0.8-1.3 m, respectively.

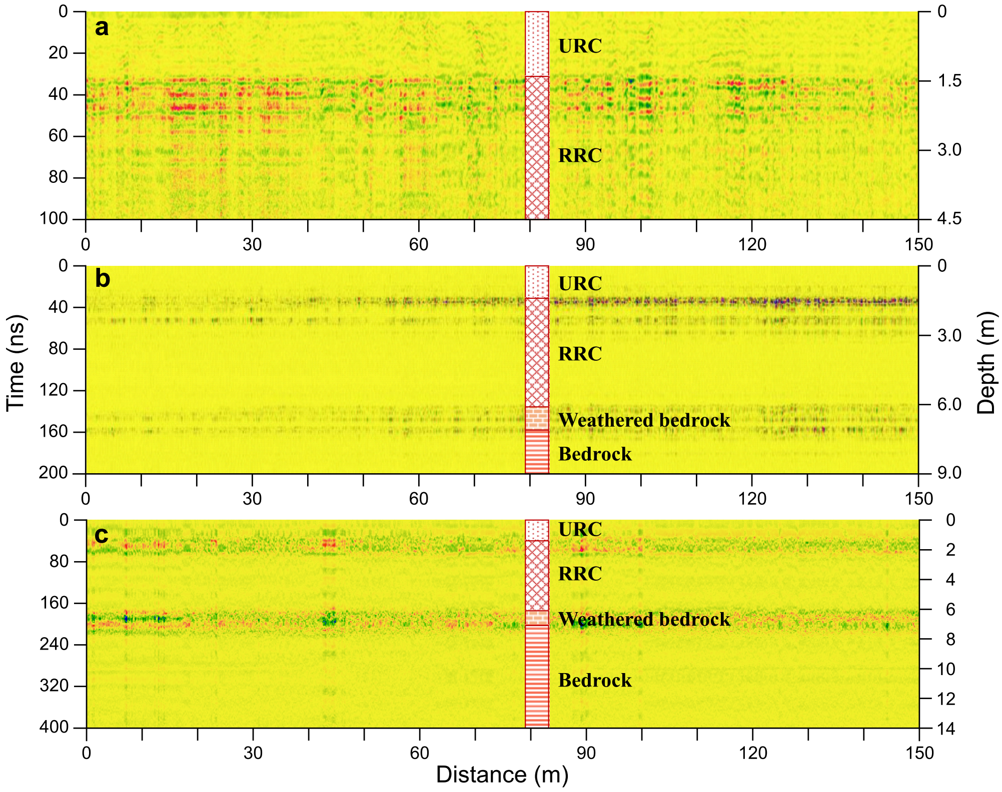


Fig. 8. GPR radargrams of Line 1 near borehole BH5 with frequencies of 200 MHz (a), 120 MHz (b) and 60 MHz (c). Note that the ground direct waves were removed, and the mean velocity was 0.07 m/ns. The land use and regolith of BH5 are upland and 7.10 m, respectively.

Antennae with various frequencies were also used on the paddy field (Line 2 in Fig. 3b) (Fig. 9). The paddy field below 3 m was saturated because of irrigation, and the strength of the perpendicular electromagnetic waves quickly decreased, resulting in a shallower penetration depth than that of upland. Consequently, the 60 MHz antenna was widely used in the paddy fields. The parameters of a gain function were amplified to achieve a clear image of the underlying substrate. This northeast-oriented radargram showed a 7.4 m depth of the regolith (Fig. 9), where the attenuation was greatly affected by the underlying texture rather than natural attenuation. The regolith thickness at the radargram between 0 and 30 m was greater than at other depths. The electromagnetic waves were intensively reflected at depths between 0.7-0.9 m, which suggests an interface between the URC and RRC layers. A similar reflection was exhibited at depths between 0.9-5.5 m, and the propagation velocity did not obviously change due to the same dielectric permittivity; thus, an RRC layer could be inferred. After the signal gain, an abrupt reflection change was observed from 5.5 to 7.5 m, which should have resulted from textural differences between the RRC and weathered bedrock layers.

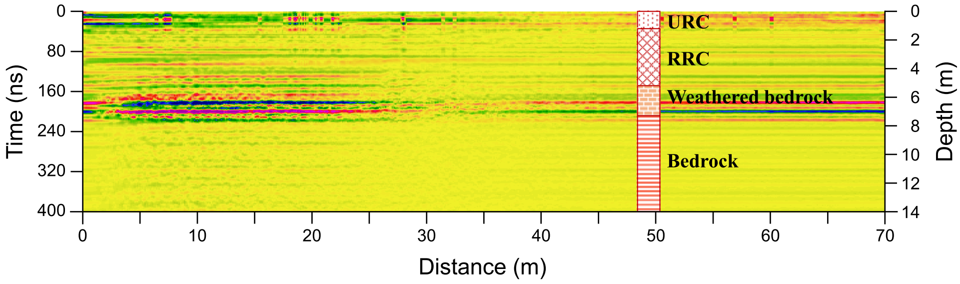


Fig. 9. GPR radargram of Line 2 near borehole BH3 where the land use was paddy field and regolith depth was 7.30 m.

3.3. Paleotopography reconstruction

Based on the thickness information interpreted from radargrams (420 points), the derived thickness of the URC layer was positively correlated with that of the RRC layer (r=0.54), and the thickness of the weathered bedrock layer was negatively correlated with those of the URC (r=-0.40) and RRC layers (r=-0.57). Furthermore, the paleotopography was reconstructed by integrating this information with the spatial predictions. The VIFs in different GWR models were less than 10, indicating that there was no collinearity problem (Table 2).

Table 2. The VIFs of the geographically weighted regression (GWR) models for the spatial prediction. Note that the mean VIF was used for each predictor of the GWR model (N=294).

URC [a)]		RRC		Weathered bedrock		Regolith	
Predictors	VIF [b)]	Predictors	VIF	Predictors	VIF	Predictors	VIF
Band 4	2.25	DisToChan [c)]	8.27	Elevation	9.25	Convergence index	1.31
Elevation	3.05	Elevation	7.14	Plan curvature	1.22	DisToChan	2.30
Land use	2.89	Plan curvature	1.69	Slope	1.25	Plan curvature	1.69
Profile curvature	1.58	Profile curvature	1.55	ReSlpPosi	9.25	Profile curvature	1.58
Solar radiation	2.19	Solar radiation	1.50			Solar radiation	1.43
TWI	2.07	TWI	2.17			TWI	2.34

[a)] URC: uniform red clay; RRC: reticulate red clay.

[b)] VIF: variance inflation factor.

[c)] DisToChan: distance to channel; ReSlpPosi: relative slope position; TWI: topographic wetness index.

The relative variable importance was measured for each model (Fig. 10). The variable ranks in terms of coefficients varied across different prediction cases (e.g., URC versus RRC). Remote sensing and land use data were employed for the prediction of the uppermost layer (URC). For the prediction of URC thickness, land use was the most important predictor (Fig. 10), followed by elevation and solar radiation. Remote sensing predictors (band 4) did not greatly benefit the overall performance. The two most important predictors were elevation and DisToChan for RRC and ReSlpPosi and elevation for weathered bedrock.

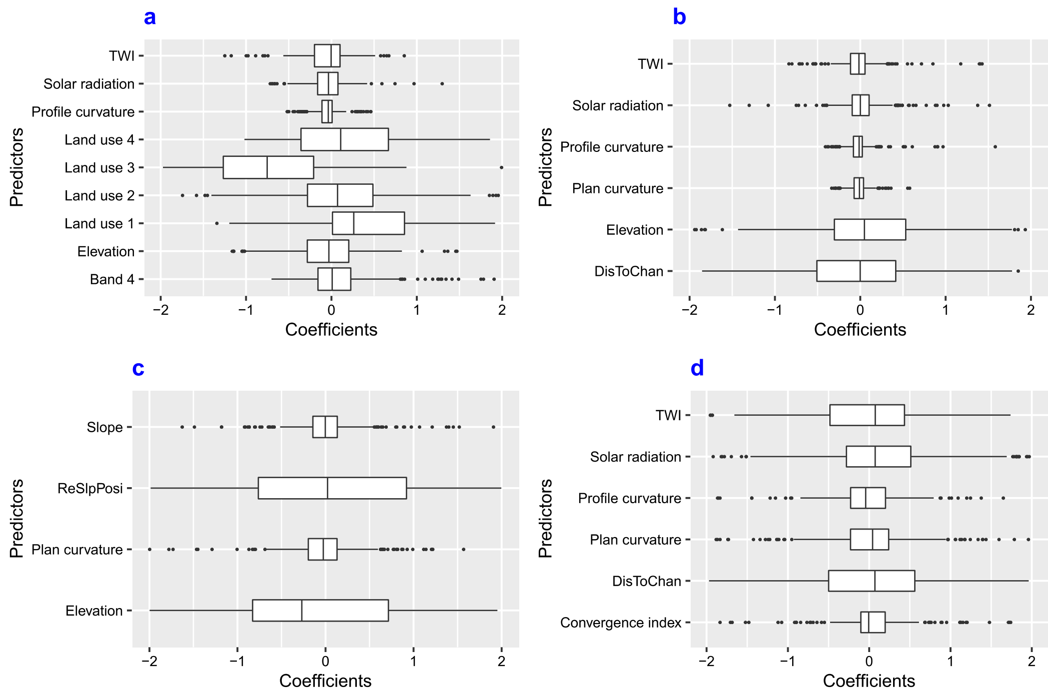


Fig. 10. The variable importance measured in GWR models in terms of regression coefficients: URC (a), RRC (b), weathered bedrock (c) and regolith (d). Land use 1, land use 2, land use 3 and land use 4 are dummy variables. DisToChan: distance to channel; ReSlpPosi: relative slope position; TWI: topographic wetness index.

The performance of the fitted models was separately evaluated by points selected from radargrams and observations (i.e., augering and borehole logs) (Table 3). Two validation datasets showed similar prediction accuracies. The R^2 values were generally greater than 0.5 (Table 3). The RMSEs for the prediction of URC, RRC and weathered bedrock ranged from 0.29 to 0.34 m, 0.34 to 0.38 m and 0.22 to 0.37 m, respectively. Few errors were involved in the prediction procedure, and the paleotopographic information was adequately robust. The paleotopography elevation map (Fig. 11f) was derived by subtracting the thicknesses of the URC (Fig. 11a), RRC (Fig. 11b) and weathered bedrock (Fig. 11c) from the current elevation (Fig. 2b). The mean regolith thicknesses for citrus, upland, paddy field and vineyard were 5.46 m, 7.31 m, 5.82 m and 4.86 m, respectively.

To assess the prediction uncertainty, the regolith thickness was also interpolated, and the differences between regolith thickness and the sum of three layers' thicknesses (i.e., URC, RRC and weathered bedrock) were obtained (Fig. 11e). Approximately 88% and 98% of the differences were smaller than 1 m and 2 m, respectively. The mean value of these differences was 0.05 m, and the first and third quantiles were -0.21 m and 0.31 m, respectively (Fig. 11e and Table 4). Large differences were mainly found in the western part of this area.

Table 3. The validation results of different predictive models based on thickness data interpreted by GPR and observations.

	Validated by GPR points				Validated by observations			
	ME [a] (m)	RMSE (m)	R^2	N	ME (m)	RMSE (m)	R^2	N
URC [b]	-0.03	0.29	0.75	126	0.10	0.34	0.43	24
RRC	-0.05	0.38	0.74	126	-0.02	0.34	0.77	8
Weathered bedrock	0.01	0.22	0.51	126	-0.16	0.37	0.45	8
Regolith	-0.01	0.80	0.79	126	-0.10	0.35	0.68	8

[a] ME: mean error; N: number of samples; R^2: coefficient of determination; RMSE: root mean square error.

[b] URC: uniform red clay; RRC: reticulate red clay.

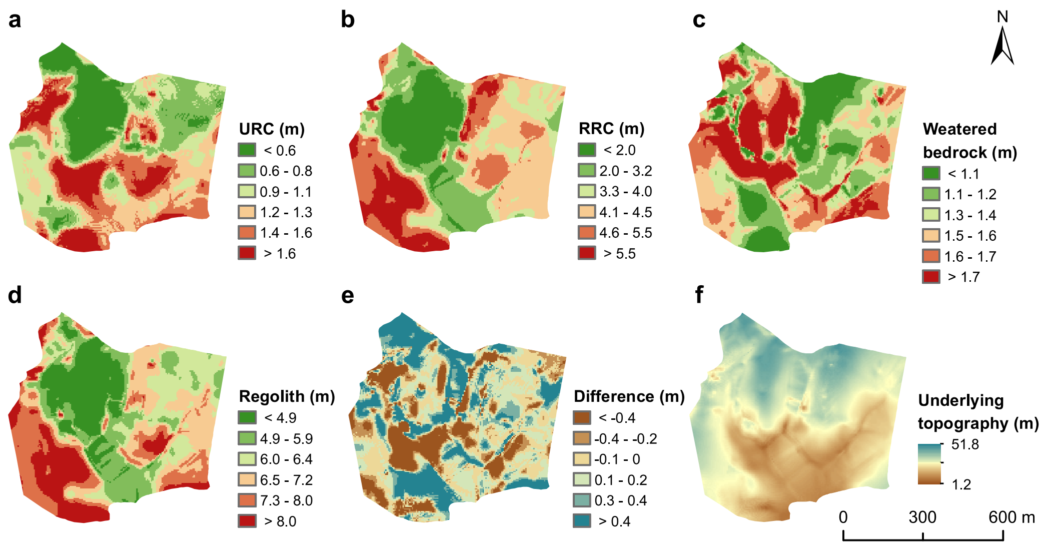


Fig. 11. Predicted thickness and elevation maps: URC layer (a), RRC layer (b), weathered bedrock layer (c), regolith thickness (d), the difference between the regolith thickness and the thickness sum of URC, RRC and weathered bedrock ($\hat{y}_{\text{Regolith}} - \hat{y}_{\text{URC}} - \hat{y}_{\text{RRC}} - \hat{y}_{\text{Weathered bedrock}}$) (e) and the elevation of the underlying paleotopography ($y_{\text{Elevation}} - \hat{y}_{\text{URC}} - \hat{y}_{\text{RRC}} - \hat{y}_{\text{Weathered bedrock}}$) (f).

Table 4. Statistics of the predicted thickness maps (m) (*N*=20,367).

	Minimum	25%	Mean	Median	75%	Maximum	Standard deviation	Skewness
URC [a)]	0.07	0.76	1.10	1.04	1.36	11.47	0.51	1.33
RRC	0.43	2.57	3.83	4.07	4.85	48.09	1.65	2.34
Weathered bedrock	0.05	1.15	1.40	1.42	1.61	4.08	0.50	0.65
Regolith thickness	0.06	5.38	6.37	6.44	7.51	16.66	1.55	-0.25
Difference	-4.48	-0.21	0.05	0.02	0.31	4.14	0.65	0.09

[a)] URC: uniform red clay; RRC: reticulate red clay; Difference: difference between the regolith thickness and the thickness sum of URC, RRC and weathered bedrock.

3.4. Quantitative assessment of landform inheritance

Approximately 98% of the paleotopographic elevations were greater than 36.27 m (Fig. 11f), with a range greater than that of the current land surface (Fig. 2b). There was an obvious difference between the URC and RRC layers, so the topography of the RRC layer was also compared with the paleotopography. The landform inheritance was investigated via linear regressions between the bedrock elevation and the current elevation (Fig. 12a) and between the bedrock elevation and the elevation of the RRC layer (Fig. 12b). The slope gradients of the fitted lines were less than 1, suggesting that the topographies of the current land surface and the surface of RRC layer were smoother than the paleotopographic terrain.

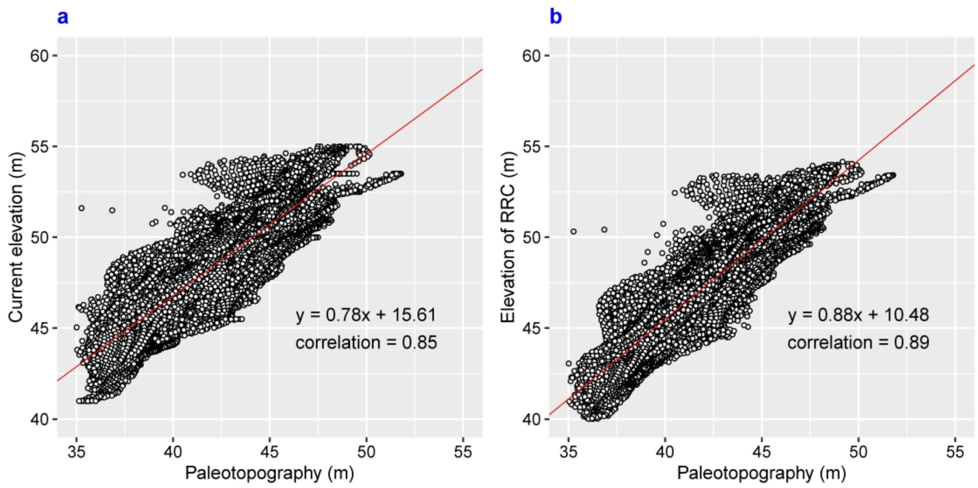


Fig. 12. Analysis of geomorphological inheritance: scatterplots of the paleotopographic elevation versus the current elevation (a) and scatterplots of the paleotopographic elevation versus the elevation of the RRC layer (b).

Two north-south and two east-west transects across the Red Soil CZO with equal horizontal intervals were also considered (Figs. 13-14). The elevation of each layer varied with the changing slope position. The regolith thickness decreased from the footslope to the summit. These curves suggest that the URC layer was thin at the summit and that the three layers were subparallel. For transect 1, the Pearson correlation coefficients between the current elevation and paleotopography was 0.55, and between the elevation of the RRC layer and paleotopography was 0.64. The correlation coefficients were greater than 0.86 for the other three transects. In conclusion, the topography of the underlying bedrock clearly shaped the landforms of the current surface and the surface of the RRC layer.

Slope maps of the current land surface, RRC layer and paleotopography were produced (Fig. 14) to visualize the surface roughness. High slope values were produced in similar places in all three maps, and the topographies of the different layers became more rugged with increasing elevation. The mean slopes and standard deviations of the land surface, RRC layer and paleotopography maps were 1.85°, 2.14° and 3.13°, and 0.88°, 1.23° and 3.18°, respectively. The mean paleotopographic slope was greater than those of the current land surface and RRC layer, which was in line with the conclusion drawn from the scatterplots (Fig. 12) and terrain profiles (Fig. 13).

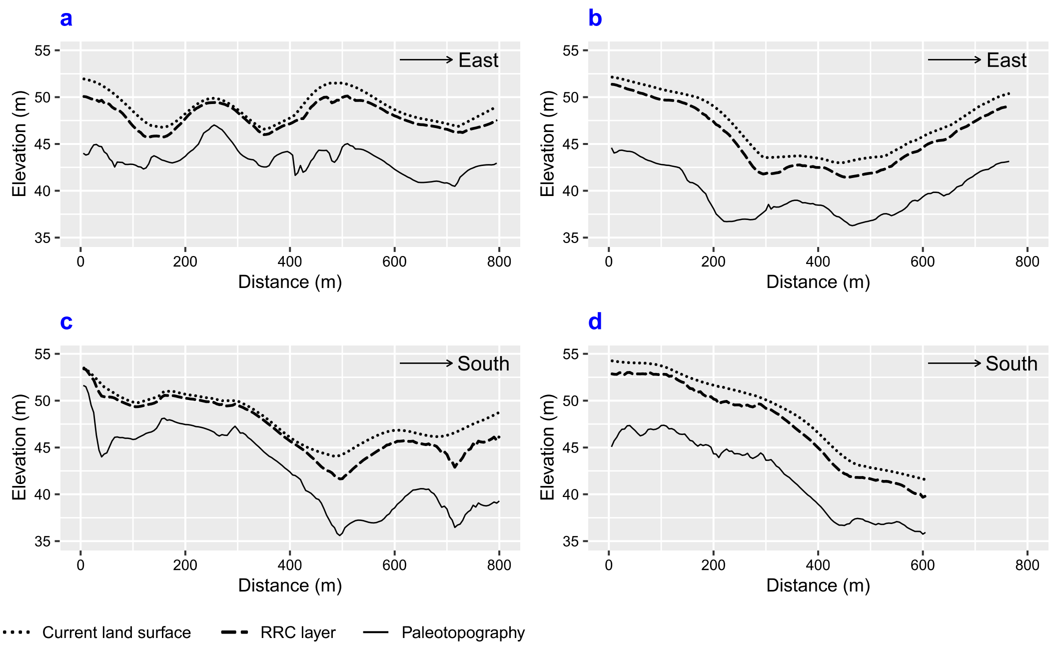


Fig. 13. The elevation of the current land surface, RRC layer and paleotopography at transect 1 (a), transect 2 (b), transect 3 (c) and transect 4 (d). The locations of these four transects are exhibited in Fig. 14a.

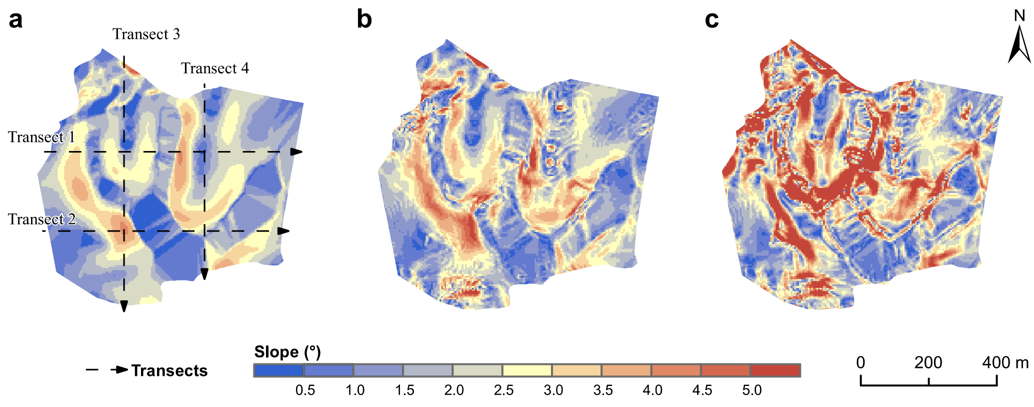


Fig. 14. The slope maps of current land surface (a), RRC layer (b) and paleotopography (c).

4. Discussion

4.1. Effect of paleotopography on geomorphological evolution

In this study, the interpreted radargrams illustrated that the boundaries of different layers were subparallel, and few discrete hyperbolas were found, which implies that paleotopography drives surface topography, even in a highly weathered subtropical landscape. The regression analysis and terrain profiles of the paleotopography and QRC layers (Figs. 12 and 13) suggested that the paleotopography clearly controls the evolution of the SCZ_S covered by the QRC in terms of terrain. This agrees with previous research that has reported significant correlations between SCZ_S and terrain attributes (Scarpone *et al.*, 2016; Wilford *et al.*, 2016). Geomorphological inheritance from the paleotopography played an important role in the QRC evolution, regardless of origin, e.g., aeolian deposits, eluvial deposits or fluvial deposits (Li *et al.*, 2013; Xiong *et al.*, 2002; Zhao *et al.*, 1992). This conclusion might not only be helpful in improving the understanding of paleoenvironmental change but also benefit geomorphological evolution simulations in future studies.

The reconstructed terrain of the underlying sandstone was rough (Fig. 14), which could be attributed to the long-term influence of paleoclimate. The seasonal changes in the paleoclimate would result in frequent changes in the paleoenvironment from warm/humid to cool/dry from the Middle Pleistocene to the Late Pleistocene (Kostić and Protić, 2000). The alteration between humid and dry conditions could directly cause groundwater table changes and, thus, enhance chemical weathering, while changes between warm and cool conditions could greatly promote the disintegration of rocks. In addition, extreme rain events massively accelerate soil erosion during summer monsoons (Cohen *et al*., 2013; Zhao *et al*., 1992). However, detailed information on the QRC evolution remains unclear. Simulations of landscape trajectories require many parameters to quantitatively represent the complex interactions between vegetation, soil, geology and climate (Hancock *et al*., 2016).

The impact of paleoclimate on pedogenic processes could be reflected by the RRC layer that exhibited net or worm-like reticulate patterns with red and white stripes (Fig. 2d). These form mainly from the integrated effect of desilicification-allitization and strong chemical eluviation that occurred in the hot and humid environment since the Pleistocene (Hu *et al*., 2014; Xiong *et al*., 2002). There are two debated reasons for the formation of reticulate clay. Some researchers believe that Fe and Al oxides were removed when rainwater infiltrated the microporosity, and soluble Fe^{3+} and Al^{3+} may be partly absorbed by plant roots (Xiong *et al*., 2000). Then, more weatherable minerals were retained (e.g., quartz). Another viewpoint is that changes in the groundwater table during rainy and dry seasons resulted in the frequent alternation of reducing and oxidizing conditions (Zhu, 1988). Free Fe would migrate along the macroporosity, and thus reticulate patterns were formed. These interpretations rely on the hypothesis that physical, chemical and biological weathering processes act together (Topal, 2002). The impact of paleoclimate on the SCZ_S, to a certain degree, can be semi-quantitatively reflected by the geochemical composition of mineralogical and mobile elements (Nesbitt and Young, 1982; Topal, 2002). Therefore, the SCZ_S in this study area were well represented by the vertical patterns of the CIA values (Fig. 6).

The predictions were consistent with the understanding of soil evolution processes. For example, the regolith was thick in the footslope and thin in the summit (Fig. 11d). We observed that the thickness of the URC significantly differed with land use type ($p<0.05$). Presently, intensive cultivation in this region has accelerated soil erosion, especially during the monsoon season (Peng *et al*., 2016), as most cash crops are planted along the slope. The transported soil material would be deposited downslope in the paddy field. To make a stagnant environment, the boundaries of paddy fields are usually constructed by an embankment, which has the environmental function of reducing soil erosion and may directly affect the soil depth. Paddy fields are characterized by high groundwater tables and hence promote the chemical weathering of shallow bedrock, in which thick weathered bedrock layers are generally formed (Table 1). Additionally, the anthropogenic disturbance that occurred during the last century was not negligible, and it includes industry development and deforestation. Consequently, it could be inferred that through anthropogenic processes the soil mantle would thicken downslope in the future, and the land use type might directly affect runoff and soil erosion. In conclusion, the SCZ_S evolution was successively driven by the paleotopography and anthropogenic disturbance.

4.2. Paleotopography modelling

Paleotopography reconstruction has received wide attention in recent years (Infante-Paez *et al*., 2017; Xiong *et al*., 2016a, 2016b). Poor prediction performance may occur due to the spatial heterogeneity features of layer thickness (Bourennane *et al*., 2014). In contrast to the observed paleotopographic elevations (i.e. outcrops) (Xiong *et al*., 2016a) that could be directly retrieved from DEMs and geological maps, detailed paleotopographic information on the QRC requires costly traditional drilling (Scarpone *et al*., 2016), which usually prohibits dense field surveying. Therefore, drilling and geophysical detection data were used together to estimate the thickness of each layer by incorporating environmental variables. The proposed method in the current study could be used in similar areas to supply more information on SCZ_S evolution.

Terrain attributes were used as covariates for the modelling in this study. The variable importance measured by GWR was also compared with the mean decrease in accuracy with random forest models in which the same covariate set was employed for one dependent variable. Broadly, the variable ranks were similar across different predictive models. This result was in line with the understanding of red soil evolution in south-eastern China, and therefore, the terrain attributes were beneficial to the prediction of the SCZ_S formation and its impact on landscape processes (Olyphant *et al*., 2016; Xu and Liu, 2017). The toposequence may affect the soil evolution through soil erosion, soil water content and flow direction. If discussed at the regional scale, the terrain may influence the vegetation and climate (Wilson *et al*., 2012). For example, the TWI was suggested to be the main controller of the global distribution of critical zone thickness (Xu and Liu, 2017). Evidence of geophysical imaging also indicated that interaction between the topography and tectonic stresses significantly impacted the groundwater flow, bedrock disaggregation and chemical weathering (Clair *et al*., 2015). In the current study, the shape of the substrates was subparallel to the land surface (Fig. 8), which concurred with the GPR surveys (Orlando *et al*., 2016).

Notably, the high values of difference between the regolith thickness and the thickness sum of URC, RRC and weathered bedrock were mainly found in the western part of the study area (Fig. 11e). These large differences could be jointly ascribed to the low sampling density, few available predictors and limitations of the predictive models. Fewer GPR radargrams with a frequency of 60 MHz were collected in the western area than those in other areas (Fig. 3b), resulting in a low sampling density. Even if the considered predictors benefited the spatial prediction, they were not ideal, as these predictors may greatly affect the evolution of SCZ_S in the short term (Table 2). Some long-term environmental variables related to the geology and land surface processes, such as aeolian deposits, soil erosion and hydrological regimes in the soils, may have contributed significantly to the formation of red clay and red weathering mantles during the Quaternary (Coventry, 1982; Hu *et al*., 2010; Li *et al*., 2013; Xiong *et al*., 2000). Nevertheless, these factors were difficult to quantify and unavailable at the moment. In addition, even if the overall prediction accuracy was acceptable (Table 3), the GWR models might generate biased predictions in the areas where a spatial random effect was included (Song *et al*., 2016).

We also compared GWR with other prediction techniques: multiple linear regression and random forest. The GWR models generally achieved more accurate performance than the other two techniques. For the prediction of URC, the random forest model achieved a similar accuracy as GWR when using GPR points as validation. However, when we used the observations for performance validation, the GWR model outperformed the random forest model (R^2: 0.43 versus 0.38). The GWR also performed better than the other models in predictions of the RRC (R^2: 0.74 versus 0.69), weathered bedrock (R^2: 0.51 versus 0.35) and regolith (R^2: 0.79 versus 0.66). Notably, when using observations as the validation dataset, the prediction accuracy of the URC was slightly lower than those of the other layers (Table 3), which could be attributed to the complicated variation in topsoil that was affected by natural soil evolution and anthropogenic perturbations.

The regolith map, with a mean value of 6.37 m (Fig. 11d), was compared with a global regolith thickness product with a 1 km resolution (Pelletier *et al*., 2016), and the value of this area was approximately 5 m, which was consistent with our predictions. The modern groundwater thickness in this area was approximately 1.7-5.3 m (Gleeson *et al*., 2016), which referred to the vertical distance of the most recently recharged groundwater table. Even if the published datasets had a coarse resolution, the predicted thickness maps in this study could be independently validated.

4.3. Application of geophysical measurements

The challenges of GPR surveying in this study should be noted, even if many profiles were surveyed and interpreted by linking the borehole log and radargrams (Figs. 8-9). The GPR signals were very sensitive to soil moisture and clay materials. An accurate propagation velocity was the most important parameter for radargram interpretation. The weathered bedrock layer could be reflected by the retrieval of an aquitard or confining bed (sandstone rock). The spatial patterns of soil moisture were generally heterogeneous, and the depth of the aquifer in this area varied from 2 to 8 m (Gao *et al*., 2016). Therefore, the soil water content and thickness of the aquifer may be uncertain when conducting GPR measurements.

Reference items (e.g., iron bars, iron plates or rebar) were not buried at known depths before performing a GPR survey to estimate the velocity. This method may not be necessary in our study, as this technique provided only a velocity estimate of a very small part of the entire survey area (Aziz *et al*., 2016; Jacob and Urban, 2015). In addition, deep ploughing and other intensive agricultural management of the topsoil in this area may subsequently affect the accuracy of the velocity estimation.

The common midpoint mode can map the dielectric permittivity change along a profile where the velocity obtained from the topsoil is consistent with that measured from the calibration with the buried rebar (Aziz *et al*., 2016).

Some uncertainties in the GPR survey of this study should be improved in the future. For example, some surveys were performed at the boundary ridges between paddy fields due to saturated soil conditions. Other geophysical approaches might be required to provide additional information on the SCZ_S, such as electrical resistivity tomography and shallow seismic refraction (Holbrook *et al*., 2014; Kaufmann *et al*., 2018; Olyphant *et al*., 2016). Even if a 1D model of velocities was fitted (Fig. 7), data fusion of the GPR radargrams with different frequencies was still lacking.

5. Conclusions

The SCZ_S in the Red Soil CZO were investigated via a traditional borehole survey at the plot scale, a geophysical survey at the transect scale and a spatial prediction at the landscape scale. A spatial prediction workflow was proposed to integrate the information from borehole survey and the geophysical survey and to reconstruct the topography of the underlying layers. The interpreted GPR results in terms of thicknesses and interfaces for the three layers were consistent with the borehole logs. The trained GWR models explained 43%-77% of the spatial variation of the three layers. The linear regressions and terrain profiles of the paleotopography and QRC layers suggested that the current topography inherited the paleotopography. The reconstructed terrain of the underlying sandstone was rather rough when comparing the slope of paleotopography with that of the current landform. It was inferred that the underground structure of the critical zone covered by QRC was mainly driven by terrain. The proposed prediction based on thickness information, which was in turn based on radargrams, could easily be conducted in areas with similar underground layers to supply more information on the corresponding SCZ$_S$ and might benefit the simulation of geomorphological evolution.

Acknowledgements

This study was supported by the National Natural Science Foundation of China (grant No. 41571130051, No. 41771251 and No. 41977003), the National Key Research and Development Program of China (No. 2018YFE0107000) and the UK Natural Environmental Research Council (NE/N007611/1). We thank the individual authors of each study regarding critical zone research in the Red Soil CZO. We are grateful to Dong-Sheng Yu, Li-Gang Zhou, Shun-Hua Yang and Yue Zhao for their support in conducting the GPR survey. We are grateful to Qin-Bo Cheng for interpreting the radargram images.

Data Availability

All radargram data, soil data and environmental predictors used in this study are available from the corresponding author upon request.

References

Aziz, A. S., Stewart, R. R., Green, S. L., Flores, J. B., 2016. Locating and characterizing burials using 3D ground-penetrating radar (GPR) and terrestrial laser scanning (TLS) at the historic Mueschke Cemetery, Houston, Texas. Journal of Archaeological Science Reports 8, 392–405.

Bièvre, G., Kniess, U., Jongmans, D., Pathier, E., Schwartz, S., Westen, C.J., et al., 2011.

Paleotopographic control of landslides in lacustrine deposits (Trieves plateau, French western Alps). Geomorphology 125, 214–224.

Bourennane, H., Salvador-Blanes, S., Couturier, A., Chartin, C., Pasquier, C., Hinschberger, F., et al., 2014. Geostatistical approach for identifying scale-specific correlations between soil thickness and topographic attributes. Geomorphology 220(3), 58–67.

Brunsdon, C. F., Fotheringham, A. S., Charlton, M. E., 1996. Geographically weighted regression: A method for exploring spatial non-stationarity. Geographical Analysis 28, 281–298.

Cecil, M. R., Ducea, M. N., Reiners, P. W., Chase, C. G., 2006. Cenozoic exhumation of the northern Sierra Nevada, California, from (U-Th)/He thermochronology. Geological Society of America Bulletin 118(11), 1481–1488.

Chen, L. M., Zhang, G. L., Effland, W. R., 2011. Soil characteristics response times and pedogenic thresholds during the 1000-year evolution of a paddy soil chronosequence. Soil Science Society of America Journal 75(5), 1807–1820.

Clair, J. St., Moon, S., Holbrook, W. S., Perron, J. T., Riebe, C. S., Martel, S. J., et al., 2015. Geophysical imaging reveals topographic stress control of bedrock weathering. Science 350(6260), 534–538.

Cohen, S., Willgoose, G., Hancock, G., 2013. Soil–landscape response to mid and late Quaternary climate fluctuations based on numerical simulations. Quaternary Research 79(3), 452–457.

Comas, X., Terry, N., Slater, L., Warren, M., Kolka, R., Kristiyono, A., et al., 2015. Imaging tropical peatlands in Indonesia using ground-penetrating radar (GPR) and electrical resistivity imaging (ERI): Implications for carbon stock estimates and peat soil characterization. Biogeosciences 12, 2995–3007.

Conrad, O., Bechtel, B., Bock, M., Dietrich, H., Fischer, E., Gerlitz, L., et al., 2015. System for

Automated Geoscientific Analyses (SAGA) v. 2.1.4. Geoscientific Model Development 8, 1991–2007.

Coventry, R. J., 1982. The distribution of red, yellow, and grey earths in the Torrens Creek area, central north Queensland. Australian Journal of Soil Research 20(1), 1–14.

ESRI, 2014. ArcGIS Desktop: Release 10.2. Environmental Systems Research Institute, Redlands, CA.

Gao, L., Lv, Y., Wang, D., Muhammad, T., Biswas, A., Peng, X., 2016. Soil water storage prediction at high space–time resolution along an agricultural hillslope. Agricultural Water Management 165, 122–130.

Gao, L., Lv, Y., Wang, D., Tahir, M., Peng, X., 2015. Can shallow-layer measurements at a single location be used to predict deep soil water storage at the slope scale? Journal of Hydrology 531, 534–542.

Gleeson, T., Befus, K. M., Jasechko, S., Luijendijk, E., Cardenas, M. B., 2016. The global volume and distribution of modern groundwater. Nature Geoscience 9(2). 161–167.

Guo, L., Lin, H., 2016. Critical zone research and observatories: Current status and future perspectives. Vadose Zone Journal 15(9).

Hancock, G. R., Lowry, J. B. C., Coulthard, T. J., 2016. Long-term landscape trajectory — Can we make predictions about landscape form and function for post-mining landforms? Geomorphology 266, 121–132.

Holbrook, W. S., Riebe, C. S., Elwaseif, M., Hayes, J. L., Basler-Reeder, K., Harry, D. L., et al., 2014. Geophysical constraints on deep weathering and water storage potential in the Southern Sierra Critical Zone Observatory. Earth Surface Processes & Landforms 39(3), 366–380.

Hu, X. F., Wei, J., Du, Y., Xu, L. F., Wang, H. B., Zhang, G. L., et al., 2010. Regional distribution of the

Quaternary Red Clay with aeolian dust characteristics in subtropical China and its paleoclimatic implications. Geoderma 159, 317–334.

Hu, X. F., Du, Y., Guan, C. L., Xue, Y., Zhang, G. L., 2014. Color variations of the Quaternary Red Clay in southern China and its paleoclimatic implications. Sedimentary Geology 303(6), 15–25.

Infante-Paez, L., Cardona, L. F., McCullough, B., Slatt, R., 2017. Seismic analysis of paleotopography and stratigraphic controls on total organic carbon: Rich sweet spot distribution in the Woodford Shale, Oklahoma, USA. Interpretation 5(1), 33–47.

Jacob, R. W., Urban, T. M., 2015. Ground-penetrating radar velocity determination and precision estimates using common-midpoint (CMP) collection with hand-picking, semblance analysis and cross-correlation analysis: a case study and tutorial for archaeologists. Archaeometry 57, 1–16.

Kaufmann, G., Romanov, D., Tippelt, T., Vienken, T., Werban, U., Dietrich, P., Mai, F., Börnerc, F., 2018. Mapping and modelling of collapse sinkholes in soluble rock: The Münsterdorf site, northern Germany. Journal of Applied Geophysics 154, 64–80.

Kostić, N., Protić, N., 2000. Pedology and mineralogy of loess profiles at Kapela-Batajnica and Stalac, Serbia. Catena 41(1), 217–227.

Li J. W., Ye, W., Zhang, G. L., Zhu, L. D., Jiang, Y. J., Gong, Z. T., 2013. Grain Size Evidence of Multiple Origins of Red Clays in the Jinhua-Quzhou Basin, South China. Pedosphere 23(5), 686–695.

Liu, K., Ding, H., Tang, G., Song, C., Liu, Y., Jiang, L., et al., 2018. Large-scale mapping of gully-affected areas: An approach integrating Google Earth images and terrain skeleton information. Geomorphology 314, 13–26.

Muggler, C. C., Van Loef, J. J., Buurman, P., van Doesburg, J. D. J., 2001. Mineralogical and (sub)microscopic aspects of iron oxides in polygenetic Oxisols from Minas Gerais, Brazil. Geoderma 100, 147–171.

National Research Council, 2001. Basic research opportunities in earth sciences. National Academies Press, Washington, DC.

Nesbitt, H. W., Young, G. M., 1982. Early Proterozoic climates and plate motions inferred from major element chemistry of lutites. Nature 299, 715–717.

Olyphant, J., Pelletier, J. D., Johnson, R., 2016. Topographic correlations with soil and regolith thickness from shallow-seismic refraction constraints across upland hillslopes in the Valles Caldera, New Mexico. Earth Surface Processes & Landforms 41(12), 1684–1696.

Orlando, J., Comas, X., Hynek, S. A., Buss, H. L., Brantley, S. L., 2016. Architecture of the deep critical zone in the Río Icacos watershed (Luquillo Critical Zone Observatory, Puerto Rico) inferred from drilling and ground penetrating radar (GPR). Earth Surface Processes & Landforms 41(13), 1826–1840.

Parsekian, A. D., Singha, K., Minsley, B. J., Holbrook, W. S., Slater, L., 2015. Multiscale geophysical imaging of the critical zone. Reviews of Geophysics 53(1), 1–26.

Pelletier, J. D., Broxton, P. D., Hazenberg, P., Zeng, X., Troch, P. A., Niu, G. Y., et al., 2016. A gridded global data set of soil, immobile regolith, and sedimentary deposit thicknesses for regional and global land surface modeling. Journal of Advances in Modeling Earth Systems 8(1), 41–65.

Peng, X., Zhu, Q. H., Xie, Z. B., Darboux, F., Holden, N. M., 2016. The impact of manure, straw and biochar amendments on aggregation and erosion in a hillslope Ultisol. Catena 138, 30–37.

Rempe, D. M., Dietrich, W. E., 2014. A bottom-up control on fresh-bedrock topography under landscapes. Proceedings of the National Academy of Sciences 111(18), 6576–6581.

Richter, D. D., Mobley, M. L., 2009. Monitoring earth's critical zone. Science 326(5956), 1067–1068.

Scarpone, C., Schmidt, M. G., Bulmer C. E., Knudby, A., 2016. Modelling soil thickness in the critical zone for Southern British Columbia. Geoderma 282, 59–69.

Schaetzl, R. J., Krist, F. J., Rindfleisch, P. R., Liebens, J., Williams, T. E., 2000. Postglacial landscape evolution of northeastern lower Michigan, interpreted from soils and sediments. Annals of the Association of American Geographers 90(3), 443–466.

Simeoni, M. A., Galloway, P. D., O'Neil, A. J., Gilkes, R. J., 2009. A procedure for mapping the depth to the texture contrast horizon of duplex soils in south-western Australia using ground penetrating radar, GPS and kriging. Australian Journal of Soil Research 47(6), 613–621.

Soil Survey Staff, 2010. Keys to Soil Taxonomy, 11th ed. USDA-Natural Resources Conservation Service, Washington, DC.

Song, X. D., Brus, D. J., Liu, F., Li, D. C., Zhao, Y. G., Yang, J. L., et al., 2016. Mapping soil organic carbon content by geographically weighted regression: A case study in the Heihe River Basin, China. Geoderma 261, 11–22.

Tang, J. L., Zhang, B., Gao, C., Zepp, H., 2008. Hydrological pathway and source area of nutrient losses identified by a multi-scale monitoring in an agricultural catchment. Catena 72, 374–385.

Tanner, L. H., Lucas, S. G., 2018. Pedogenic record of climate change across the Pennsylvanian-Permian boundary in red-bed strata of the Cutler Group, northern New Mexico, USA. Sedimentary Geology 373, 98–110.

Tian, M., Han, L., Meng, Q., Jin, Y., Meng L., 2019. In situ investigation of the excavation-loose zone in surrounding rocks from mining complex coal seams. Journal of Applied Geophysics 168, 90–100.

Topal, T., 2002. Quantification of weathering depths in slightly weathered tuffs. Environmental Geology 42, 632–641.

Torres, M., Gaines, R., 2011. Paleosol geochemistry of the late Paleocene Goler Formation of Southern California. Applied Geochemistry 26, S135–S138.

Wichura, H., Bousquet, R., Oberhansli, R., Strecker, M. R., Trauth, M. H., 2010. Evidence for middle Miocene uplift of the East African Plateau. Geology 38, 543–546.

Wilford, J. R., Searle, R., Thomas, M., Pagendam, D., Grundy, M. J., 2016. A regolith depth map of the Australian continent. Geoderma 266, 1–13.

Wilson, J. P., 2012. Digital terrain modeling. Geomorphology 137(1), 107–121.

Wu, H. Y., Song, X. D., Zhao, X. R., Peng, X. H., Zhou, H., Hallett, P. D., et al., 2019. Accumulation of nitrate and dissolved organic nitrogen at depth in a red soil Critical Zone. Geoderma 337, 1175–1185.

Xiong, L. Y., Tang, G. A., Strobl, J., Zhu, A. X., 2016a. Paleotopographic controls on loess deposition in the Loess Plateau of China. Earth Surface Processes and Landforms 41, 1155–1168.

Xiong, L. Y., Tang, G. A., Zhu, A. X., Li, J. L., Duan, J. Z., Qian, Y. Q., 2016b. Landform-derived placement of electrical resistivity prospecting for paleotopography reconstruction in the loess landforms of China. Journal of Applied Geophysics 131, 1–13.

Xiong, S., Ding, Z., Liu, D., 2000. The worm-shaped veins in the red earth of South China—Pedological evidence for root traces of past forest. Chinese Science Bulletin 45(19), 1800–1804.

Xiong, S., Sun, D., Ding, Z., 2002. Aeolian origin of the red earth in Southeast China. Journal of Quaternary Science 17(2), 181–191.

Xu, X., Liu, W., 2017. The global distribution of earth's critical zone and its controlling factors.

Geophysical Research Letters 44(7), 2967–3428.

Zhao, Q. G., 1992. A study on recent pedogenesis and its developing age of red soils in China. Quaternary Sciences 4, 344–351. (in Chinese with English abstract)

Zhao, Q. G., Huang, G. Q., Ma, Y. Q., 2013. The problems in red soil ecosystem in southern of China and its countermeasures. Acta Ecologica Sinica 33(24), 7615–7622. (in Chinese with English abstract)

Zhu, A. X., Yang, L., Li, B., Qin, C., Pei, T., Liu, B., 2010. Construction of membership functions for predictive soil mapping under fuzzy logic. Geoderma 155(3–4), 164–174.

Zhu, J., 1988. Genesis and research significance of the plinthitic horizon. Geographical Research 7(4), 12–20. (in Chinese with English abstract)

Figure captions

Fig. 1. A schematic diagram of the boundaries between URC, RRC, weathered bedrock and bedrock: current landform is subparallel to the topography of the underlying strata (a), current landform is more rugged than the topography of the underlying strata (b) and current landform is smoother than the topography of the underlying strata (c).

Fig. 2. Study area of the Red Soil CZO and core samples: regional landform (a), elevation and drilling sites in the study area (b), upland landscape of the red soil (c), a typical RRC sample (d) and a weathered layer (e). Note that the mud was scraped from the borehole samples (d-e).

Fig. 3. Spatial distribution of the ground penetrating radar profiles surveyed by the US Radar with a 250 MHz shielded antenna in November 2016 (a) and by the Geoscanner/AKULA-9000C with 60, 120

and 200 MHz unshielded antennae in January 2018 (b). Note that old paddy fields indicate cultivation durations of greater than 100 years.

Fig. 4. Workflow of the paleotopographic reconstruction. GPR: ground-penetrating radar; GPS: global positioning system; GWR: geographically weighted regression; SCZ_S: subsurface critical zone structure.

Fig. 5. A schematic diagram of the boundaries between URC, RRC, weathered bedrock and bedrock and manually selected points. Note that these boundaries (dotted lines) are interpreted from a GPR common offset profile.

Fig. 6. Scatterplot of the chemical index of alteration (CIA) values for each borehole.

Fig. 7. Propagation velocity of electromagnetic waves interpreted by the common midpoint mode survey: GPR profile near borehole BH5 with a frequency of 60 MHz (a) and the interpreted 1D velocity model (b).

Fig. 8. GPR radargrams of Line 1 near borehole BH5 with frequencies of 200 MHz (a), 120 MHz (b) and 60 MHz (c). Note that the ground direct waves were removed, and the mean velocity was 0.07 m/ns. The land use and regolith of BH5 are upland and 7.10 m, respectively.

Fig. 9. GPR radargram of Line 2 near borehole BH3 where the land use was paddy field and regolith depth was 7.30 m.

Fig. 10. The variable importance measured in GWR models in terms of regression coefficients: URC (a), RRC (b), weathered bedrock (c) and regolith (d). Land use 1, land use 2, land use 3 and land use 4 are dummy variables. DisToChan: distance to channel; ReSlpPosi: relative slope position; TWI: topographic wetness index.

Fig. 11. Predicted thickness and elevation maps: URC layer (a), RRC layer (b), weathered bedrock layer (c), regolith thickness (d), the difference between the regolith thickness and the thickness sum of URC, RRC and weathered bedrock ($\hat{y}_{\text{Regolith}} - \hat{y}_{\text{URC}} - \hat{y}_{\text{RRC}} - \hat{y}_{\text{Weathered bedrock}}$) (e) and the elevation of the underlying paleotopography ($y_{\text{Elevation}} - \hat{y}_{\text{URC}} - \hat{y}_{\text{RRC}} - \hat{y}_{\text{Weathered bedrock}}$) (f).

Fig. 12. Analysis of geomorphological inheritance: scatterplots of the paleotopographic elevation versus the current elevation (a) and scatterplots of the paleotopographic elevation versus the elevation of the RRC layer (b).

Fig. 13. The elevation of the current land surface, RRC layer and paleotopography at transect 1 (a), transect 2 (b), transect 3 (c) and transect 4 (d). The locations of these four transects are exhibited in Fig. 14a.

Fig. 14. The slope maps of current land surface (a), RRC layer (b) and paleotopography (c).

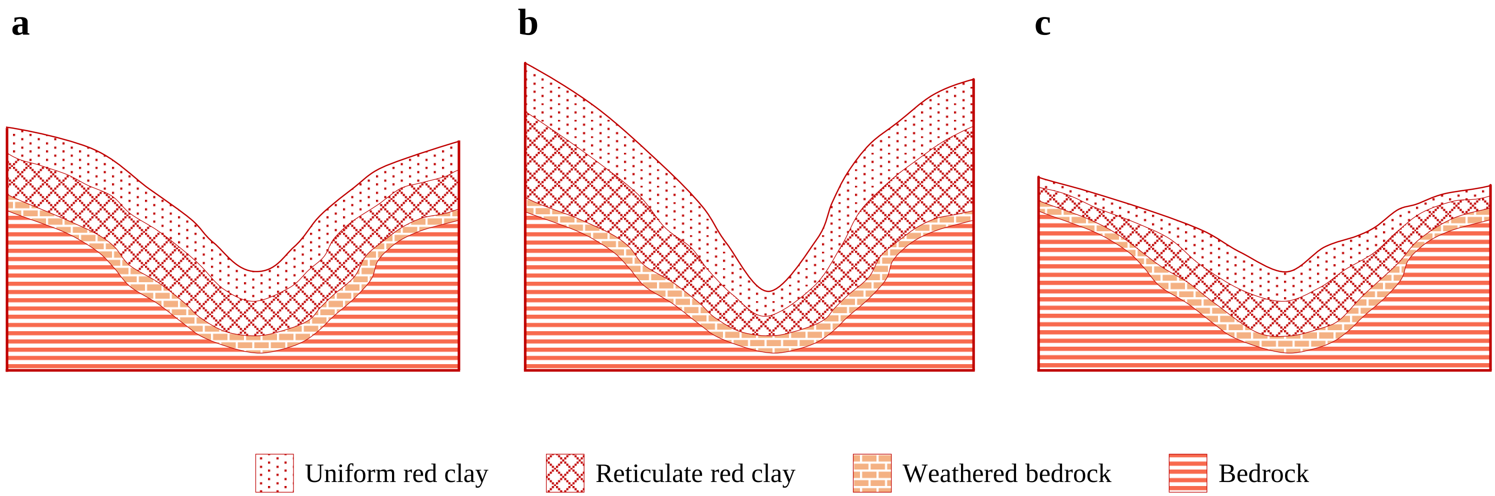
Table captions

Table 1. Summarized information of eight boreholes.

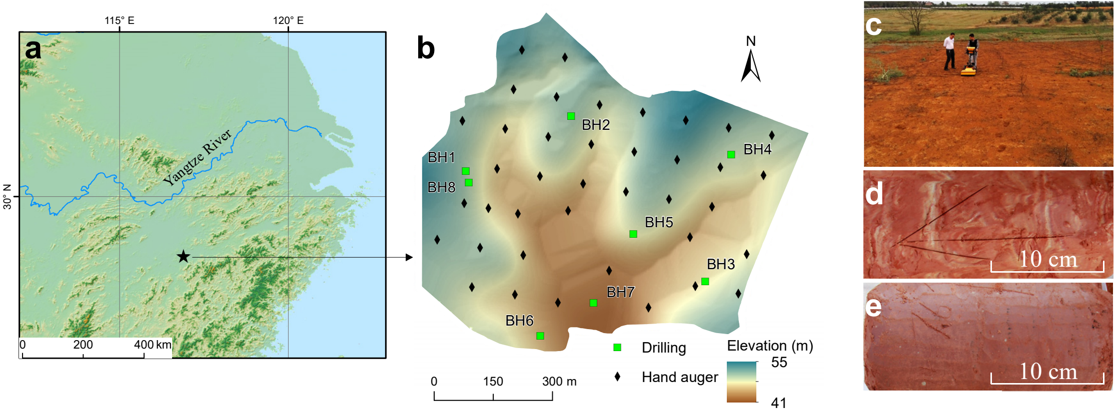
Table 2. The VIFs of the geographically weighted regression (GWR) models for the spatial prediction. Note that the mean VIF was used for each predictor of the GWR model (*N*=294).

Table 3. The validation results of different predictive models based on thickness data interpreted by GPR and observations.

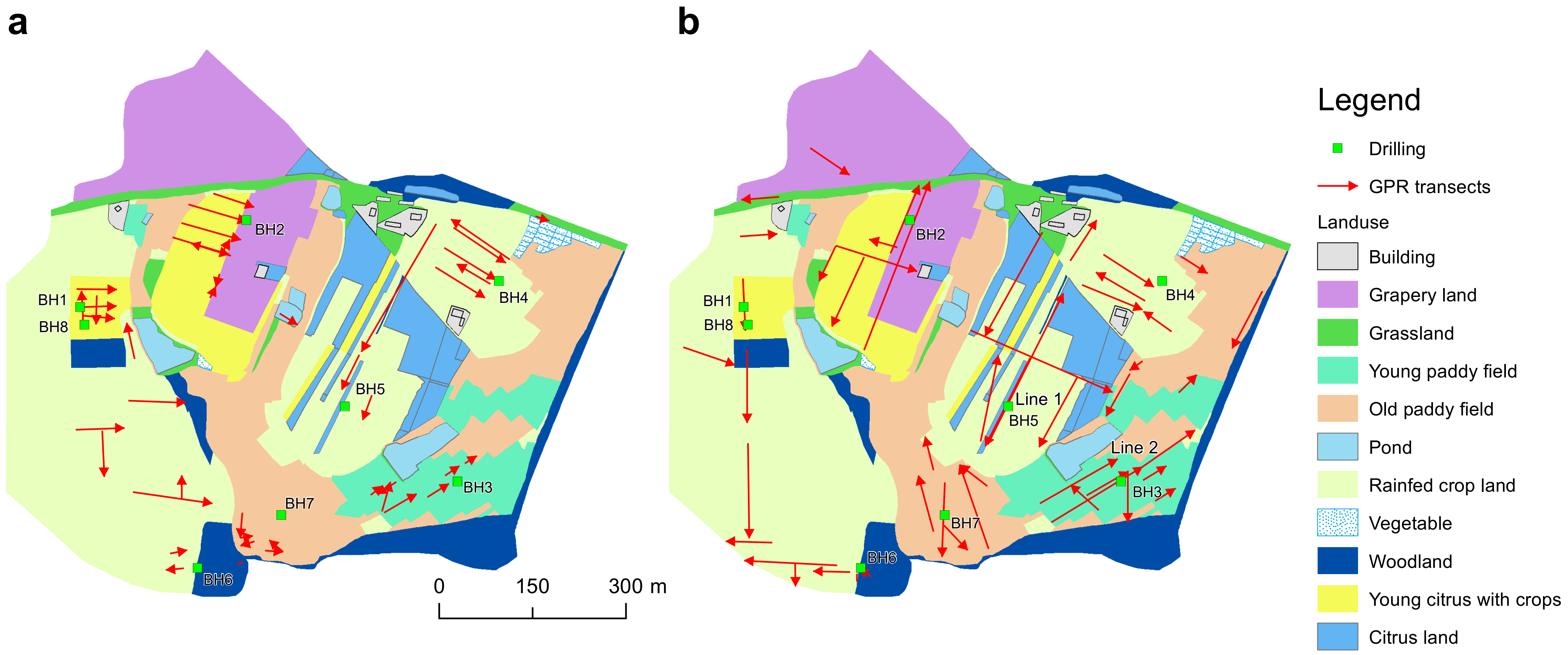
Table 4. Statistics of the predicted thickness maps (m) (*N*=20,367).



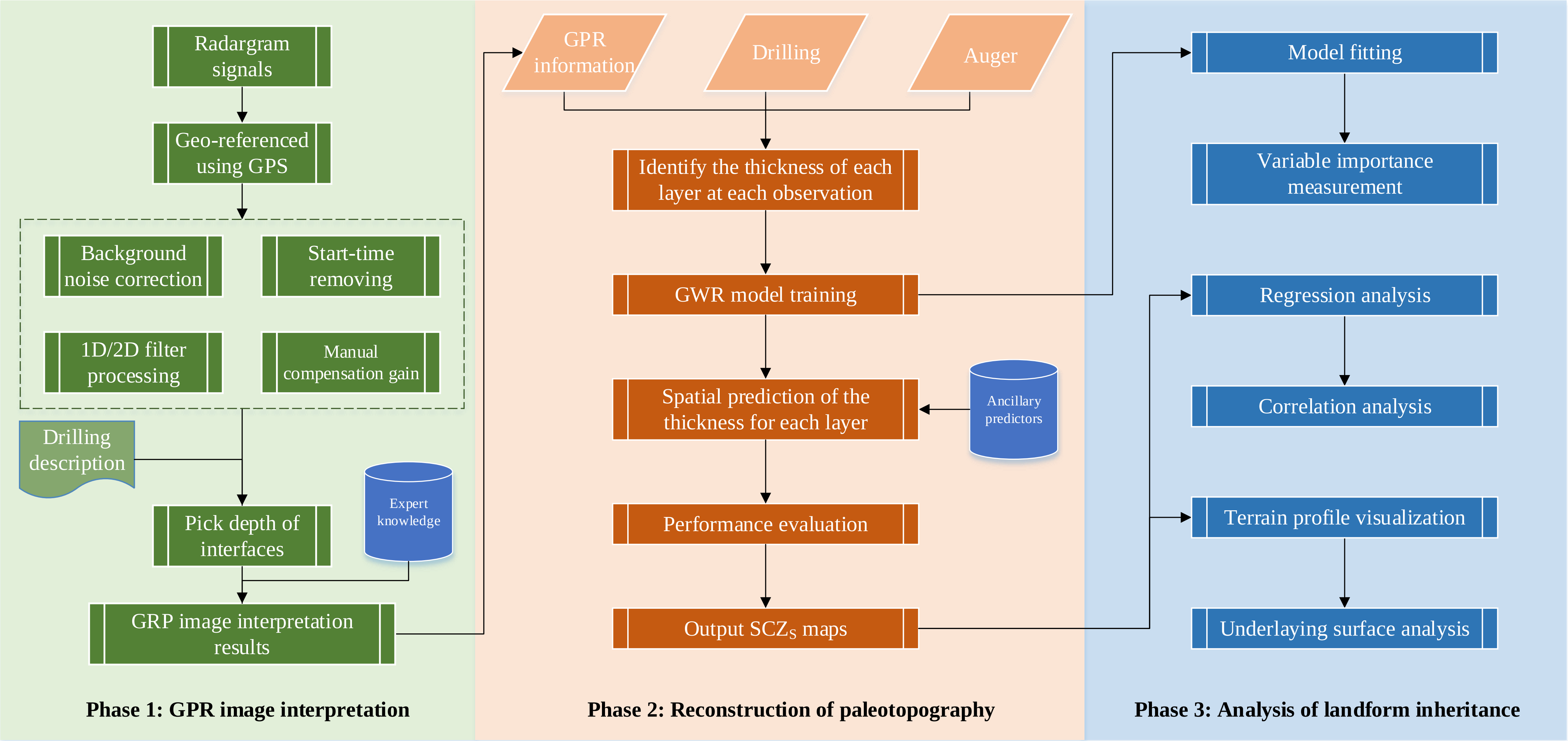
a
b
c
Uniform red clay
Reticulate red clay
Weathered bedrock
Bedrock

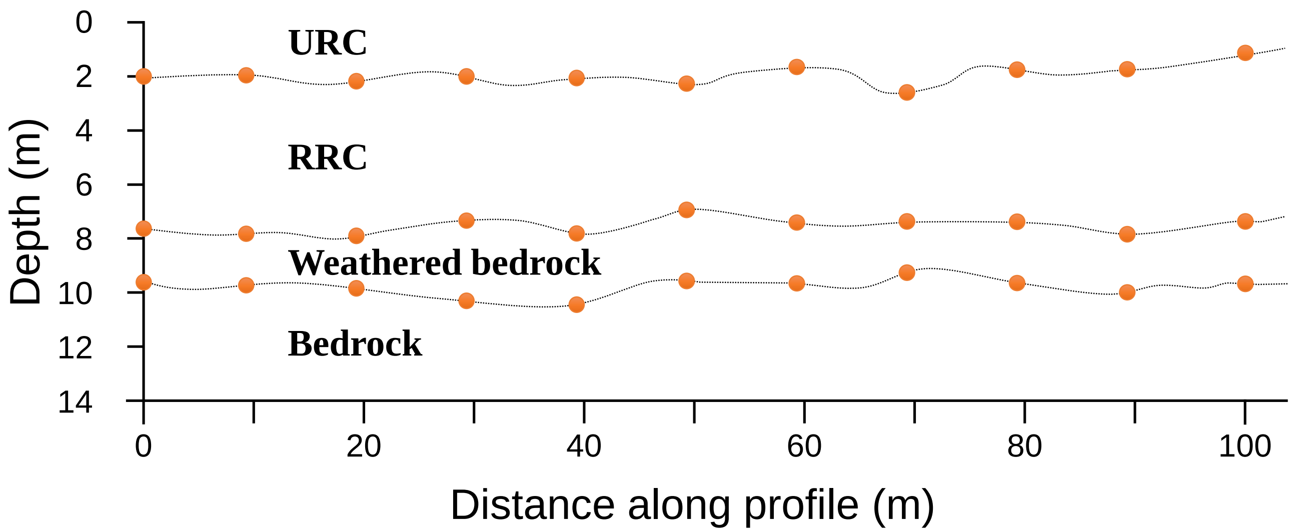


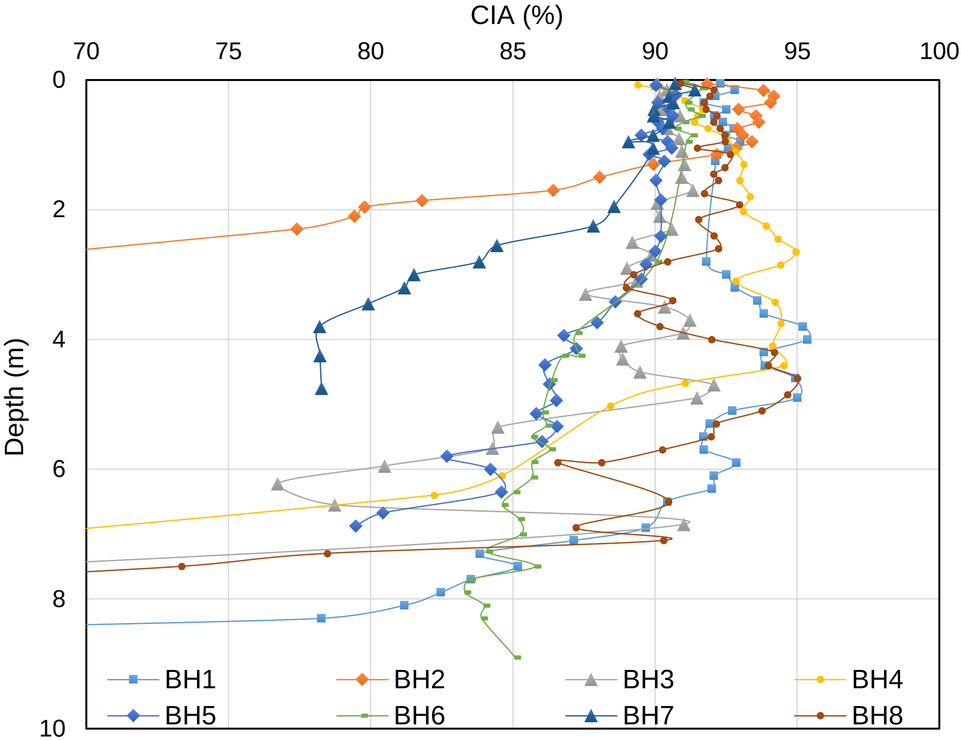
a
115° E
120° E
30° N
Yangtze River
0
200
400 km
b
N
BH1
BH2
BH3
BH4
BH5
BH6
BH7
BH8
Drilling
Hand auger
Elevation (m)
55
41
0
150
300 m
c
d
10 cm
e
10 cm



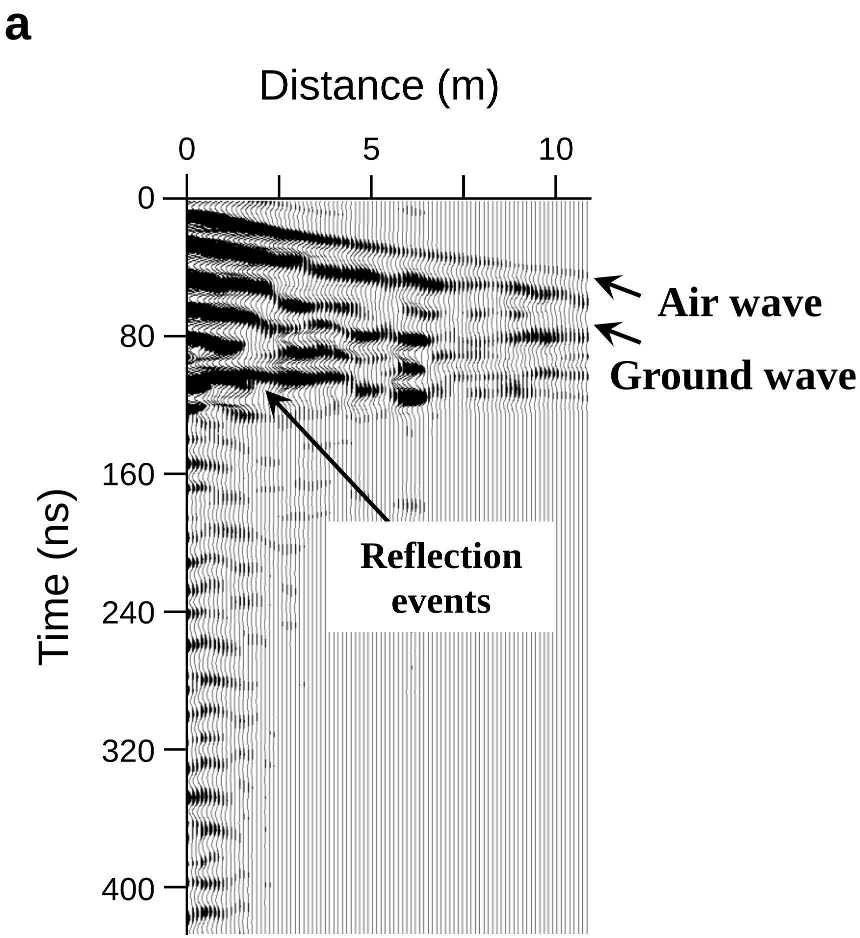
a
b
BH1
BH8
BH2
BH4
BH5
BH7
BH3
BH6
0
150
300 m
Line 1
Line 2
Legend
Drilling
GPR transects
Landuse
Building
Grapery land
Grassland
Young paddy field
Old paddy field
Pond
Rainfed crop land
Vegetable
Woodland
Young citrus with crops
Citrus land


Radargram signals
Geo-referenced using GPS
Background noise correction
Start-time removing
1D/2D filter processing
Manual compensation gain
Drilling description
Pick depth of interfaces
Expert knowledge
GRP image interpretation results
GPR information
Drilling
Auger
Identify the thickness of each layer at each observation
GWR model training
Spatial prediction of the thickness for each layer
Ancillary predictors
Performance evaluation
Output SCZs maps
Model fitting
Variable importance measurement
Regression analysis
Correlation analysis
Terrain profile visualization
Underlaying surface analysis
Phase 1: GPR image interpretation
Phase 2: Reconstruction of paleotopography
Phase 3: Analysis of landform inheritance

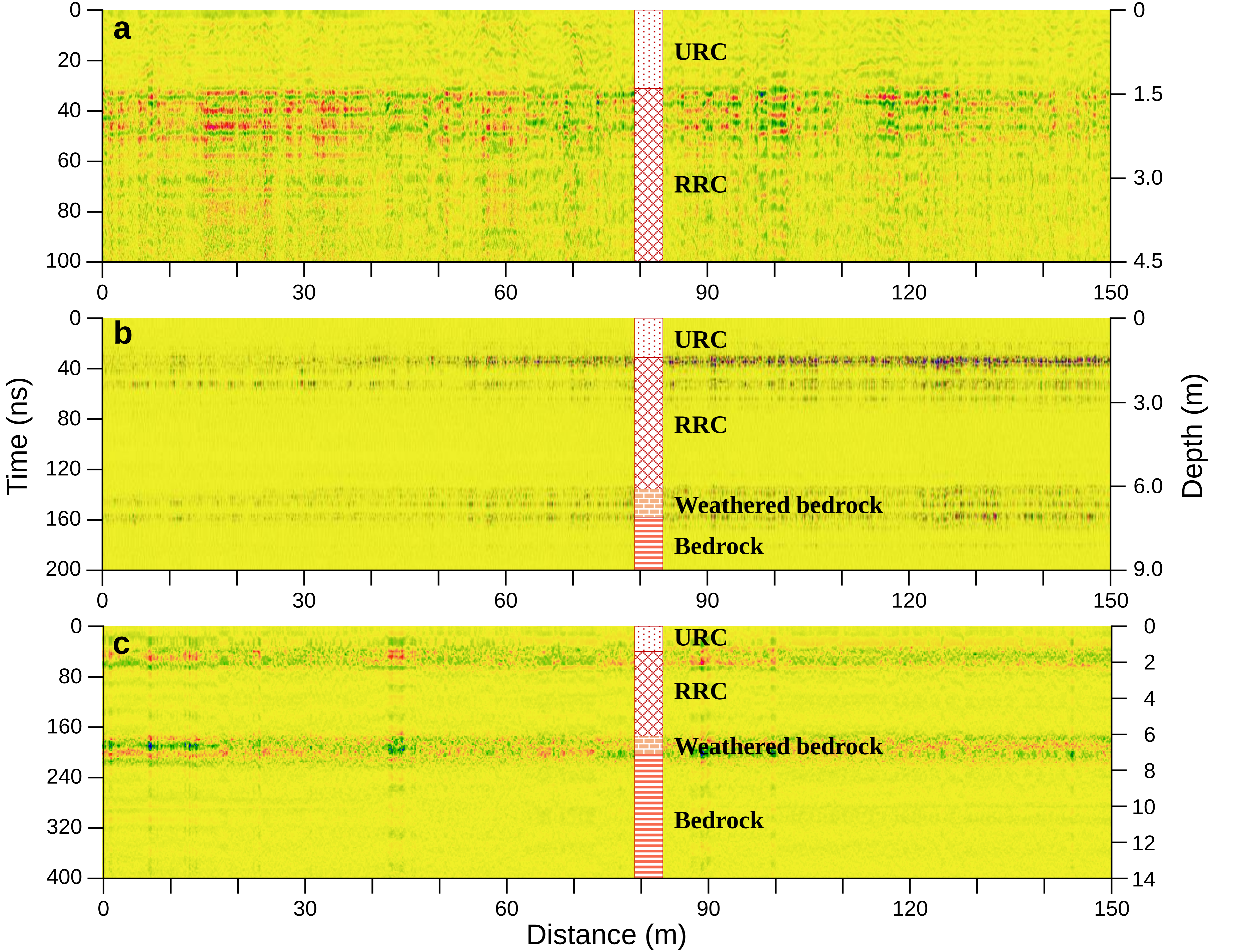

URC
RRC
Weathered bedrock
Bedrock
Depth (m)
0
2
4
6
8
10
12
14
0
20
40
60
80
100
Distance along profile (m)



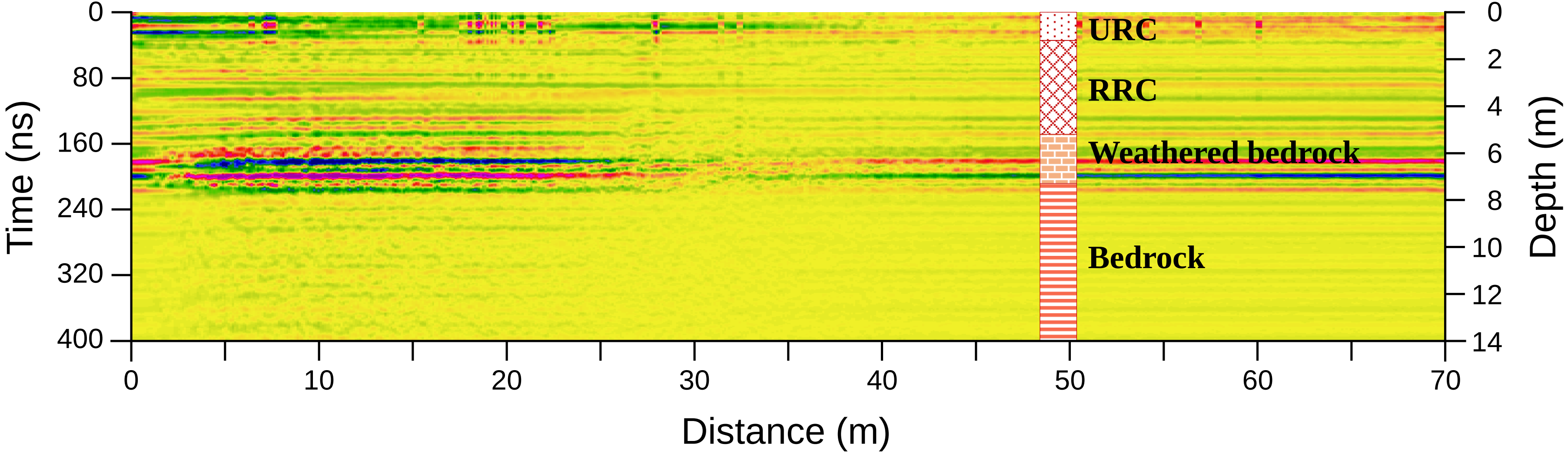
CIA (%)
70
75
80
85
90
95
100
Depth (m)
0
2
4
6
8
10
BH1
BH2
BH3
BH4
BH5
BH6
BH7
BH8

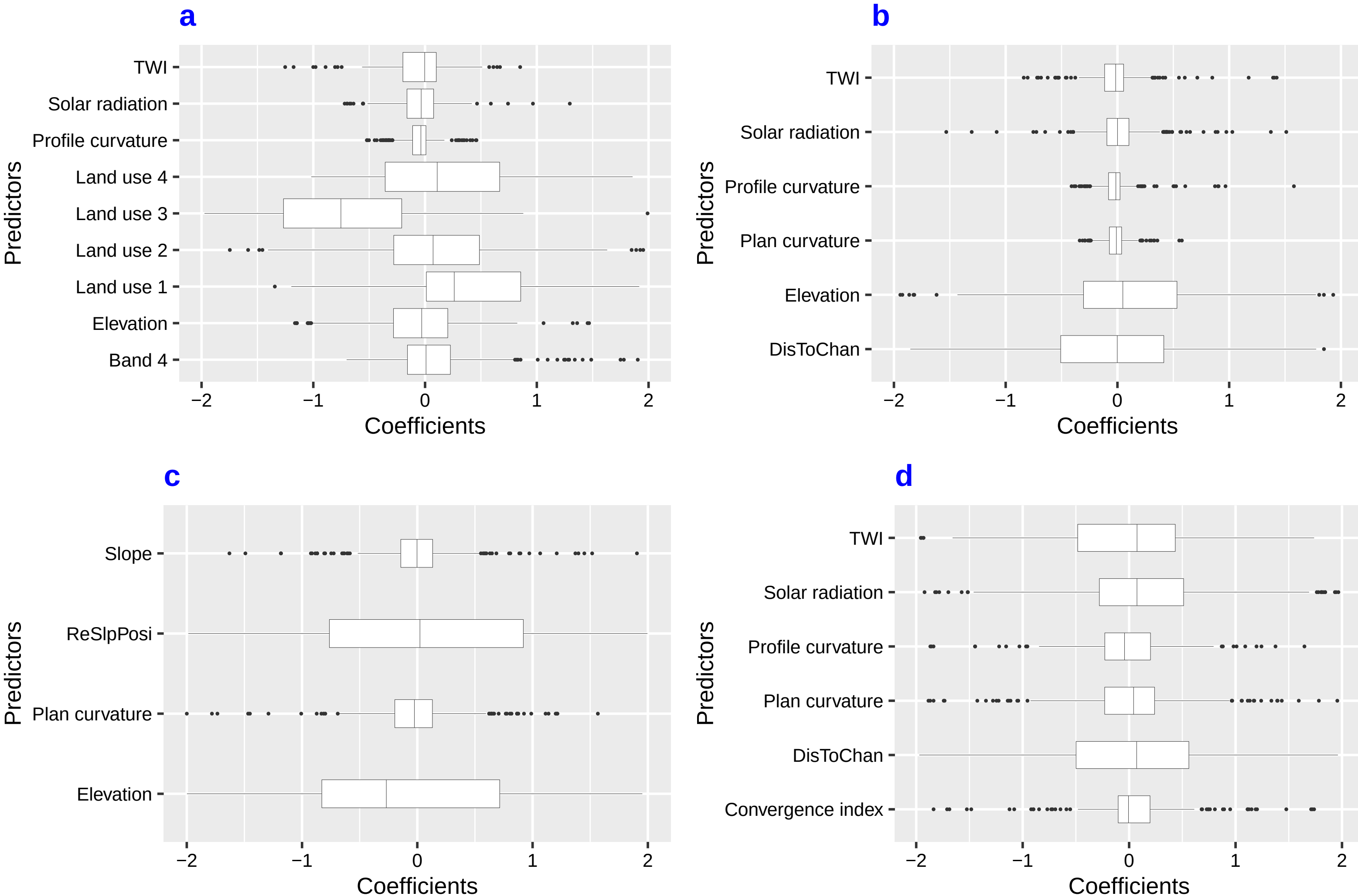


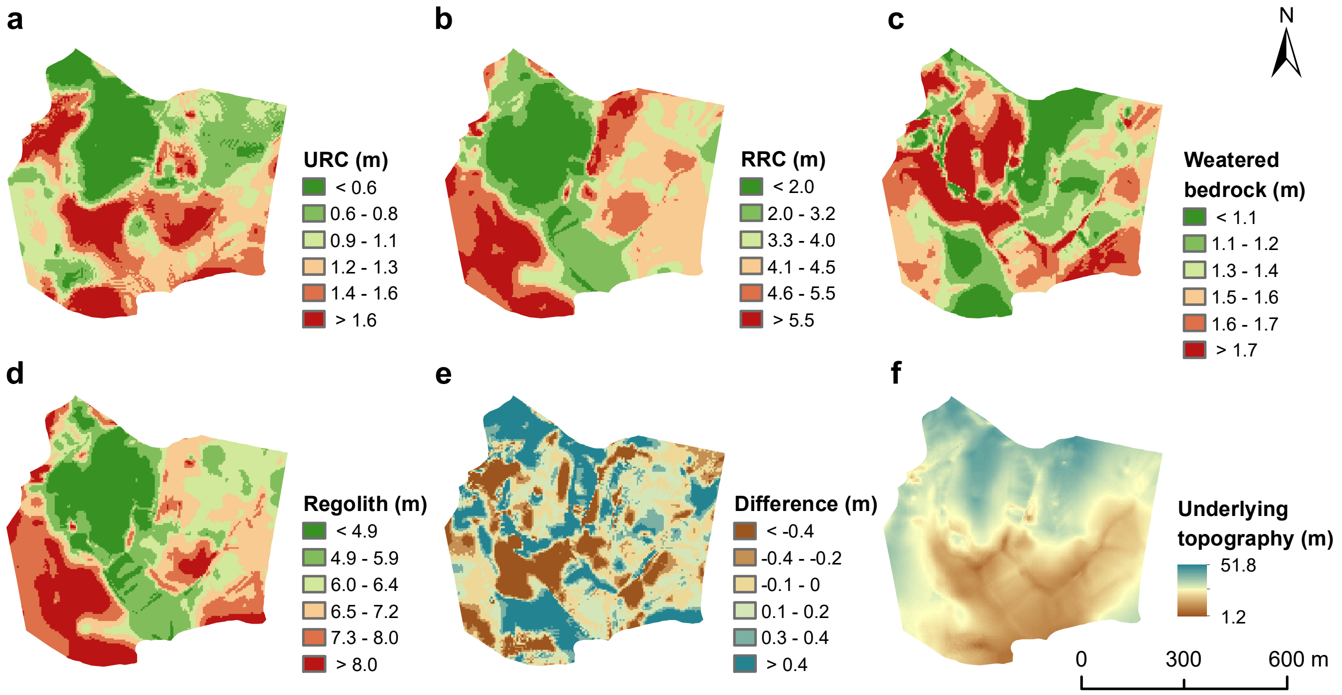
a
Distance (m)
0
5
10
0
80
160
240
320
400
Time (ns)
Air wave
Ground wave
Reflection events
b
Velocity (m/ns)
0.050
0.075
0.100
0
3
6
9
12
Depth (m)



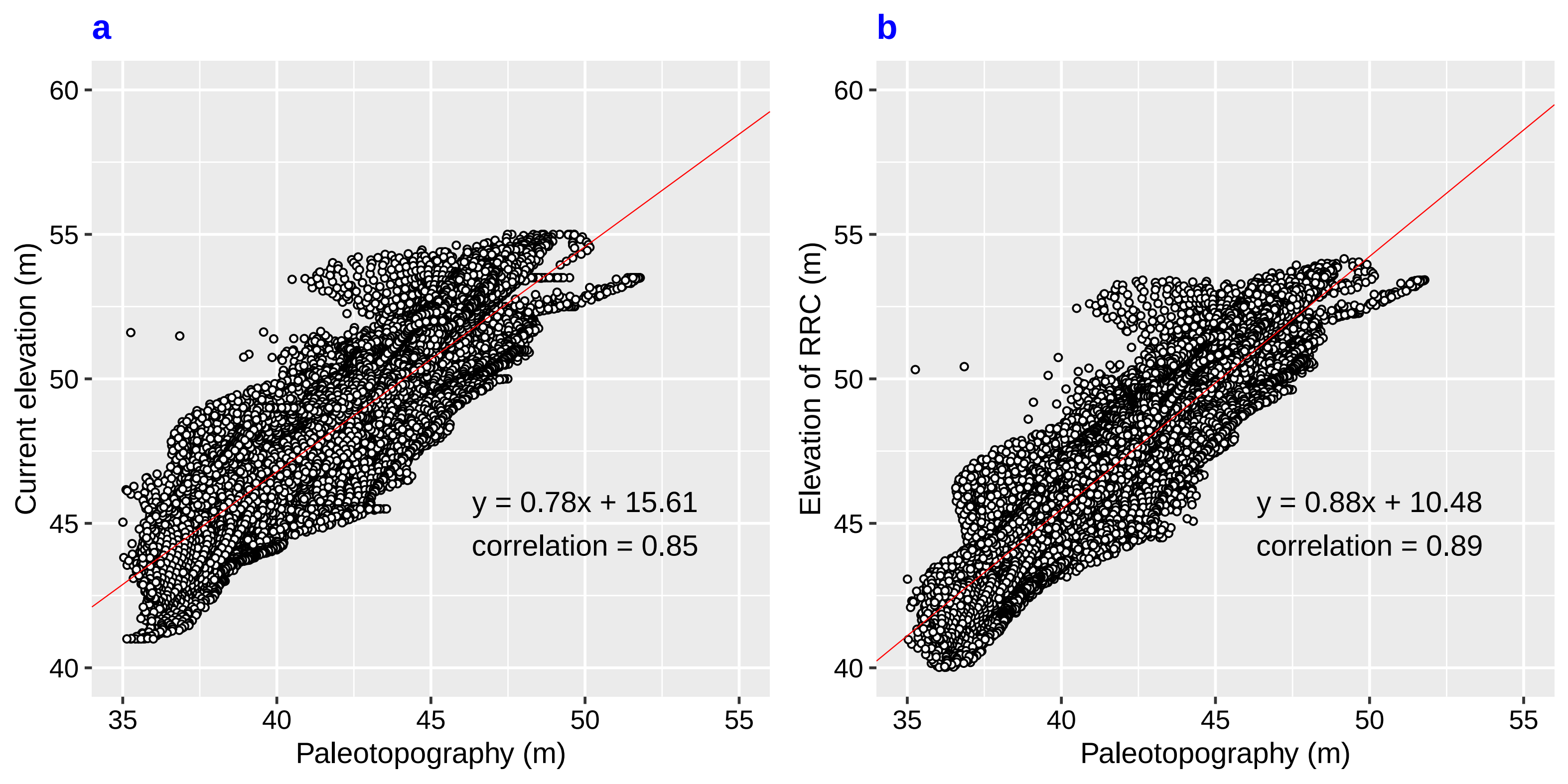
a
URC
RRC
b
URC
RRC
Weathered bedrock
Bedrock
c
URC
RRC
Weathered bedrock
Bedrock
Time (ns)
Depth (m)
Distance (m)
0
20
40
60
80
100
1.5
3.0
4.5
120
160
200
6.0
9.0
240
320
400
2
4
6
8
10
12
14
30
90
120
150

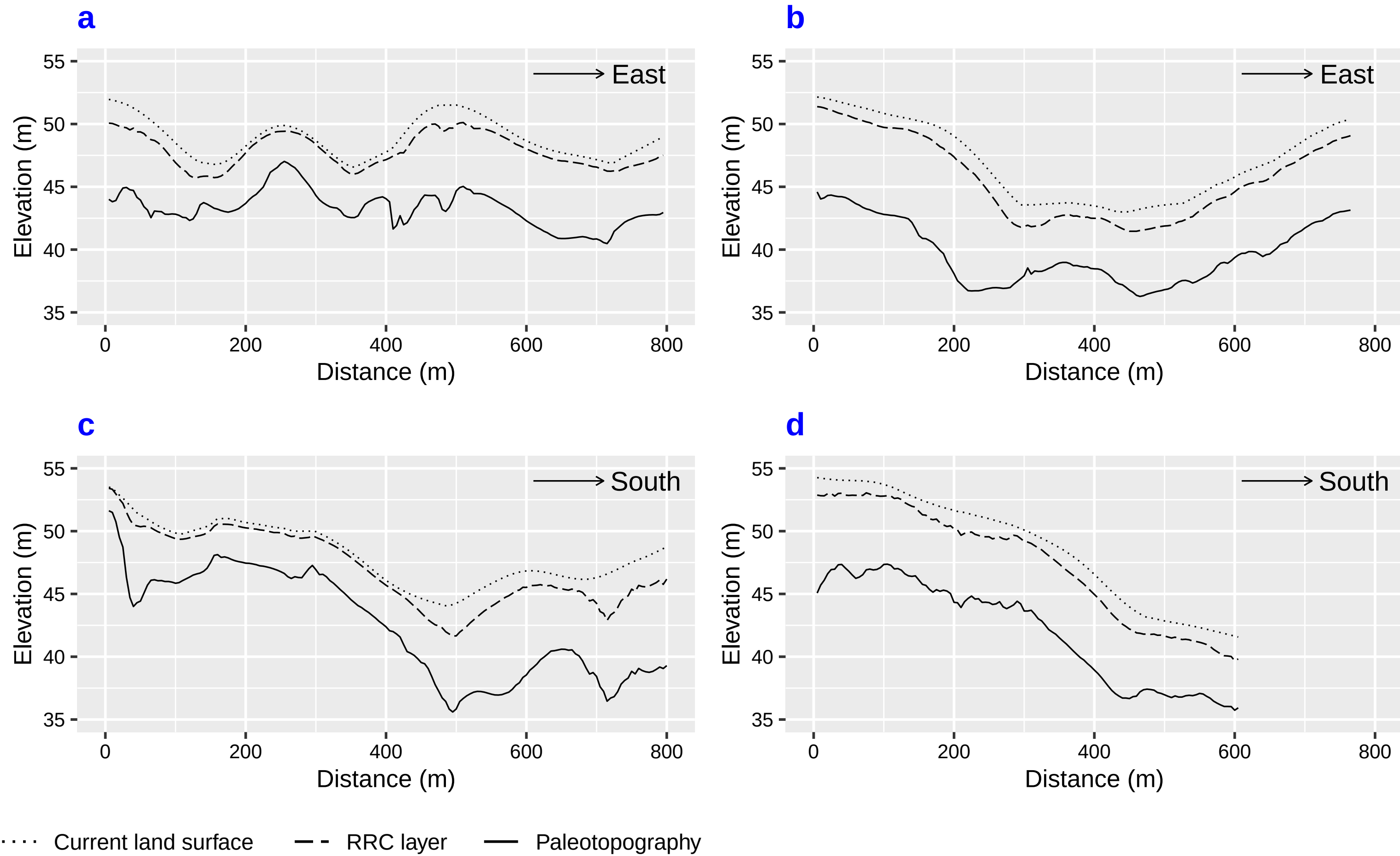

URC
RRC
Weathered bedrock
Bedrock
Time (ns)
Depth (m)
Distance (m)
0
80
160
240
320
400
0
2
4
6
8
10
12
14
0
10
20
30
40
50
60
70


a
TWI
Solar radiation
Profile curvature
Land use 4
Land use 3
Land use 2
Land use 1
Elevation
Band 4
Predictors
−2
−1
0
1
2
Coefficients
b
TWI
Solar radiation
Profile curvature
Plan curvature
Elevation
DisToChan
Predictors
−2
−1
0
1
2
Coefficients
c
Slope
ReSlpPosi
Plan curvature
Elevation
Predictors
−2
−1
0
1
2
Coefficients
d
TWI
Solar radiation
Profile curvature
Plan curvature
DisToChan
Convergence index
Predictors
−2
−1
0
1
2
Coefficients

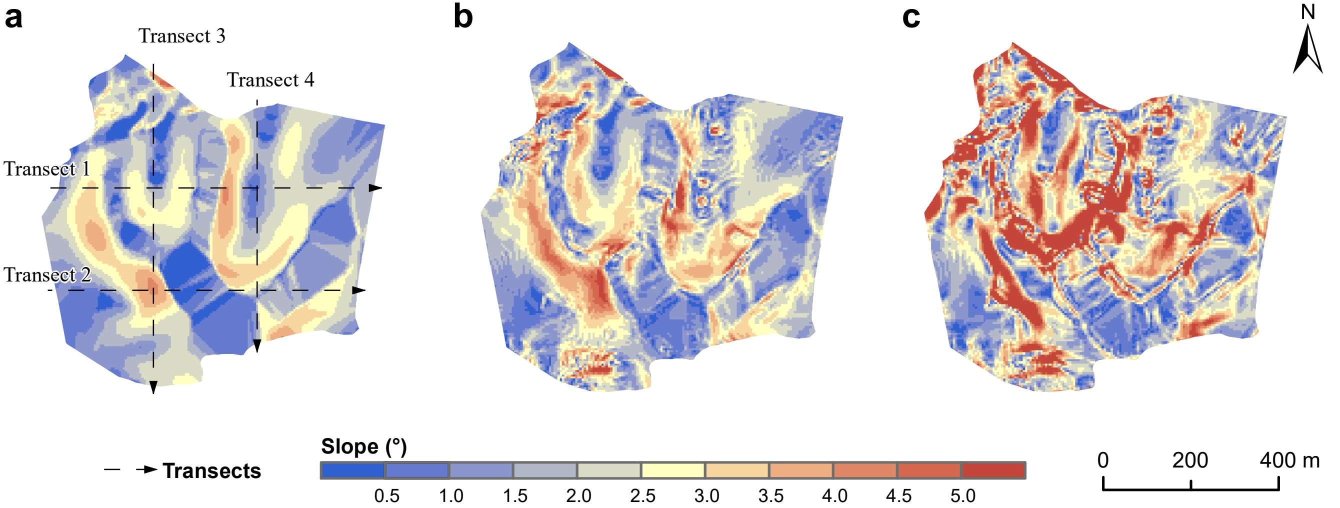


a
URC (m)
< 0.6
0.6 - 0.8
0.9 - 1.1
1.2 - 1.3
1.4 - 1.6
> 1.6
b
RRC (m)
< 2.0
2.0 - 3.2
3.3 - 4.0
4.1 - 4.5
4.6 - 5.5
> 5.5
c
N
Weatered
bedrock (m)
< 1.1
1.1 - 1.2
1.3 - 1.4
1.5 - 1.6
1.6 - 1.7
> 1.7
d
Regolith (m)
< 4.9
4.9 - 5.9
6.0 - 6.4
6.5 - 7.2
7.3 - 8.0
> 8.0
e
Difference (m)
< -0.4
-0.4 - -0.2
-0.1 - 0
0.1 - 0.2
0.3 - 0.4
> 0.4
f
Underlying
topography (m)
51.8
1.2
0
300
600 m


a
b
Current elevation (m)
Elevation of RRC (m)
Paleotopography (m)
Paleotopography (m)
y = 0.78x + 15.61
correlation = 0.85
y = 0.88x + 10.48
correlation = 0.89
35
40
45
50
55
60



a
b
c
d
East
East
South
South
Elevation (m)
Distance (m)
55
50
45
40
35
0
200
400
600
800
Current land surface
RRC layer
Paleotopography


a
b
c
Transect 3
Transect 4
Transect 1
Transect 2
N
Transects
Slope (°)
0.5
1.0
1.5
2.0
2.5
3.0
3.5
4.0
4.5
5.0
0
200
400 m

Declaration of interests

☒ The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

☐The authors declare the following financial interests/personal relationships which may be considered as potential competing interests:

None

Author statement

Song Xiao-Dong: Data curation, Investigation, Methodology, Software, Validation, Visualization, Roles/Writing – original draft, Writing – review & editing.

Wu Hua-Yong: Formal analysis, Investigation, Resources, Software, Validation, Roles/Writing – original draft.

Hallett Paul D.: Conceptualization, Funding acquisition, Project administration, Supervision, Roles/Writing – original draft, Writing – review & editing,

Pan Xi-Cai: Methodology, Software, Validation, Visualization.

Hu Xue-Feng: Methodology, Validation, Roles/Writing – original draft.

Cao Qi: Investigation, Methodology, Roles/Writing – original draft,

Zhao Xiao-Rui: Investigation, Resources, Validation.

Zhang Gan-Lin: Conceptualization, Funding acquisition, Project administration, Resources, Supervision, Writing – review & editing.