# Supplementary Materials: Abiotic Controls on Macroscale Variations of Humid Tropical Forest Height. *Remote Sensing* 2016, *8*, 494

**Yan Yang, Sassan Saatchi, Liang Xu, Yifan Yu, Michael Lefsky, Lee White, Yuri Knyazikhin and Ranga Myneni**

## 1. Ecoregion-Based Stratification

The pan-tropical study region (57°S–30°N) was divided into a variety of ecoregions based on the HWSD soil map, SRTM elevation, Globcover LC map, and Vegetation Continuous Field (VCF) product of MODIS. First, we combined the HWSD soil-mapping units (more than 16,000 types) into 41 classes using the soil attributes of FAO74 and FAO90 (Figure S8 and Table S1). We further separated the pan tropical region to three different continents–America, Africa and Asia/Australia. Treating the same soil type (within those 41 classes) from each continent separately, we got a total of 123 soil types. After removing classes with less than 2000 pixels under 1-km spatial resolution, the final classification map contains 87 soil types in total. The second step is to use STRM elevation data and separate each soil type into 4 sub-categories–low elevation (0–200 m), medium elevation (200–900 m), high elevation (900–1800 m) and very high elevation (>1800 m). That results in 348 ecoregions. In addition, finally, we defined the tropical dense forest (code 40 and 160) from the Globcover map, and only pixels with vegetation fractional covers larger than 10% were considered valid according to MODIS VCF. Thus, we obtained a pan-tropical dense forest map classified into 348 ecoregions based on soil properties and ground elevation. The classification procedure is described in Figure S9. Due to the lack of soil type diversity in the African continent, we further stratified the largest soil class (class 9-Ferralsols) to 7 small soil types according to FAO74 and FAO90 (see Table S2), which was specifically used in the analyses of the African continent.

For each ecoregion, we calculated the $TCH_m$ from the GLAS shots located within the region. We used two methods for the calculation of $TCH_m$. First, we simply averaged the TCH values retrieved from all the GLAS shots within each region. Second, we calculated the area-weighted

$TCH_m$ by using random sampling. The strategy was the following: (a) The number of GLAS shots was denoted as $n_i$ within ecoregion $i$; (b) The area of ecoregion $i$ was denoted as $A$; (c) We found the minimum GLAS point density ($\rho_{\&}$ = min( $n/A$)) out of all ecoregions (in practice, it was set to the 10 percentile in GLAS point density to remove outliers); (d) for each ecoregion $j$ with GLAS point density larger than $\rho_{\&}$, we randomly selected $\rho_{\&} A_j$ ($<n_j$) GLAS points, so that the selection of GLAS points is area-weighted. Such random sampling procedure guarantees the analysis is biased toward ecogions with larger number of GLAS shots. The experiment using both methods (Figure S5) shows that they have a very good agreement for most ecoregions, indicating that sufficient GLAS shots are available to represent the regional mean regardless of region size. We adopted the first method and created a final map of pan-tropical $TCH_m$ based on soil and elevation information (Figure S7).

## 2. Comparative Spatial Regression Results

In order to have a comparative analysis to show consistencies of spatial regression results, we performed another spatial regression method: the generalized least squares (GLS or Kriging/Geostatistical regression) approach–modeling the spatial autoregression using semi-variagrams and transferring the spatial information into error terms [43,45]. Like SEVM, GLS is also statistically rigorous and aim to retrieve the best linear unbiased estimators of regression coefficients. Following the same procedure of SEVM, we summarized our GLS results in Figure S11, and Tables S3 and S4.

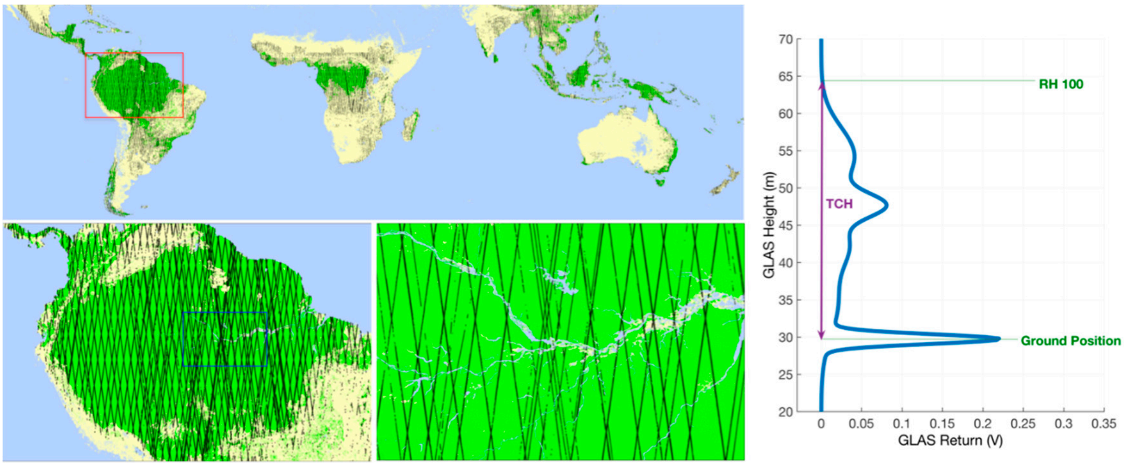
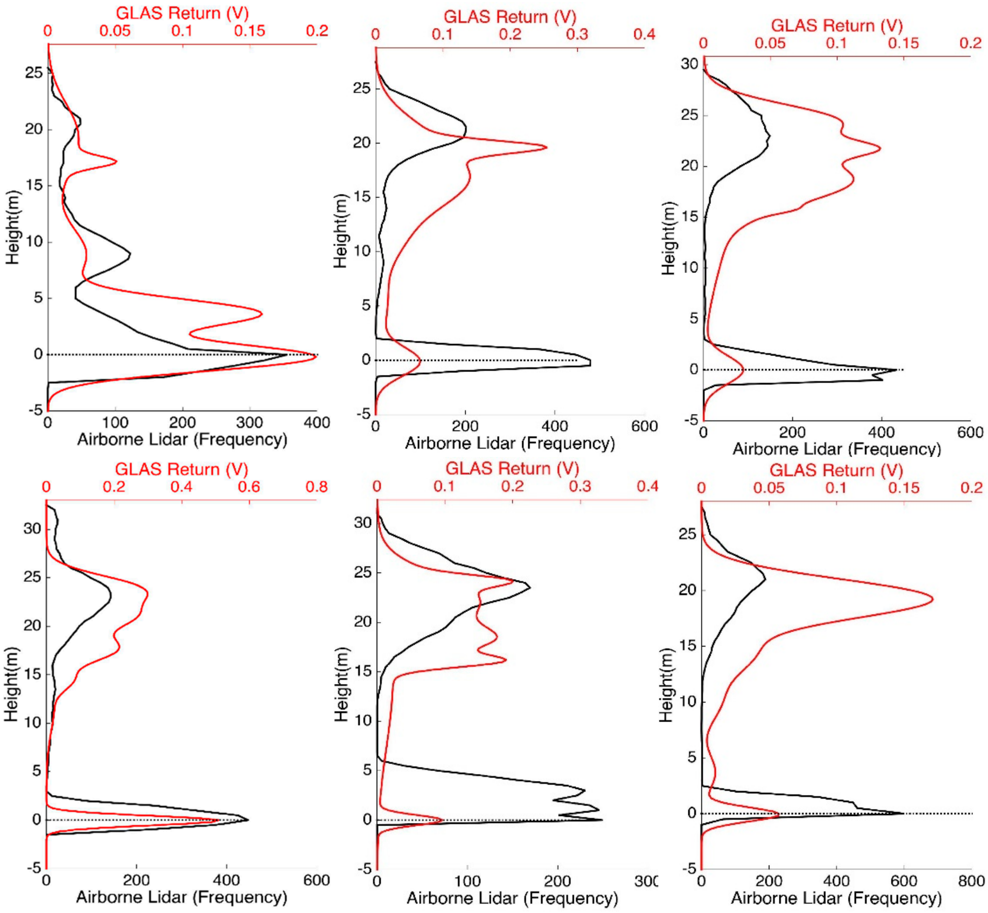
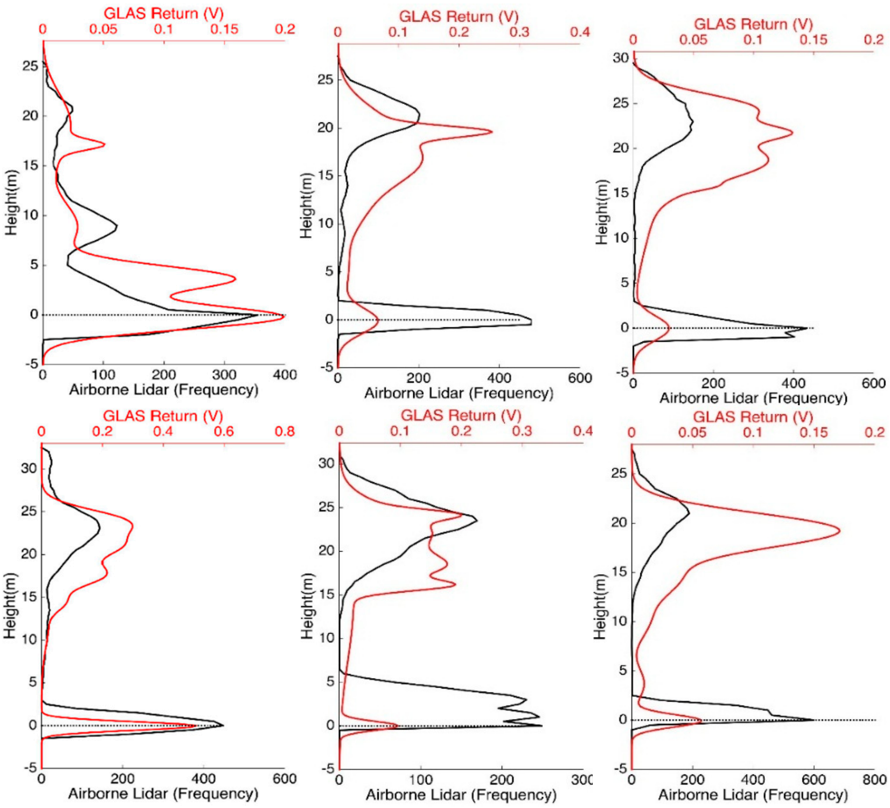



**Figure S1.** Systematic sampling of GLAS lidar shots over tropics. The upper panel is pan-tropical, the lower left panel is the region of South America, and the lower right panel is the enlarged blue rectangle showing the details of GLAS tracks in the lower left panel.

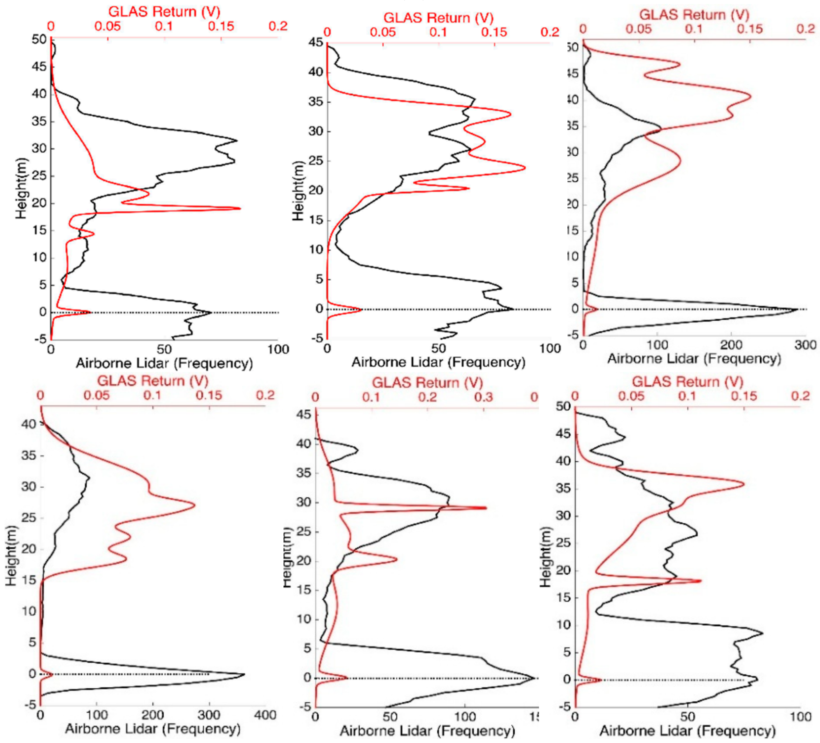



**(A)**

**Figure S2.** *Cont.*




(**B**)




(**C**)

**Figure S2.** Compassion of vertical profiles between GLAS points and associated Airborne LIDAR points in three continents: (**A**) Amazon; (**B**) Africa; (**C**) Asia. The red lines represent the profiles of GLAS waveforms, and the black lines are the profiles derived from Airborne LIDAR data. Airborne Lidar data were collected from different ecological campaigns and the vertical profiles were calculated from the aggregation of DTM (digital terrain model) products under the footprints of GLAS shots.

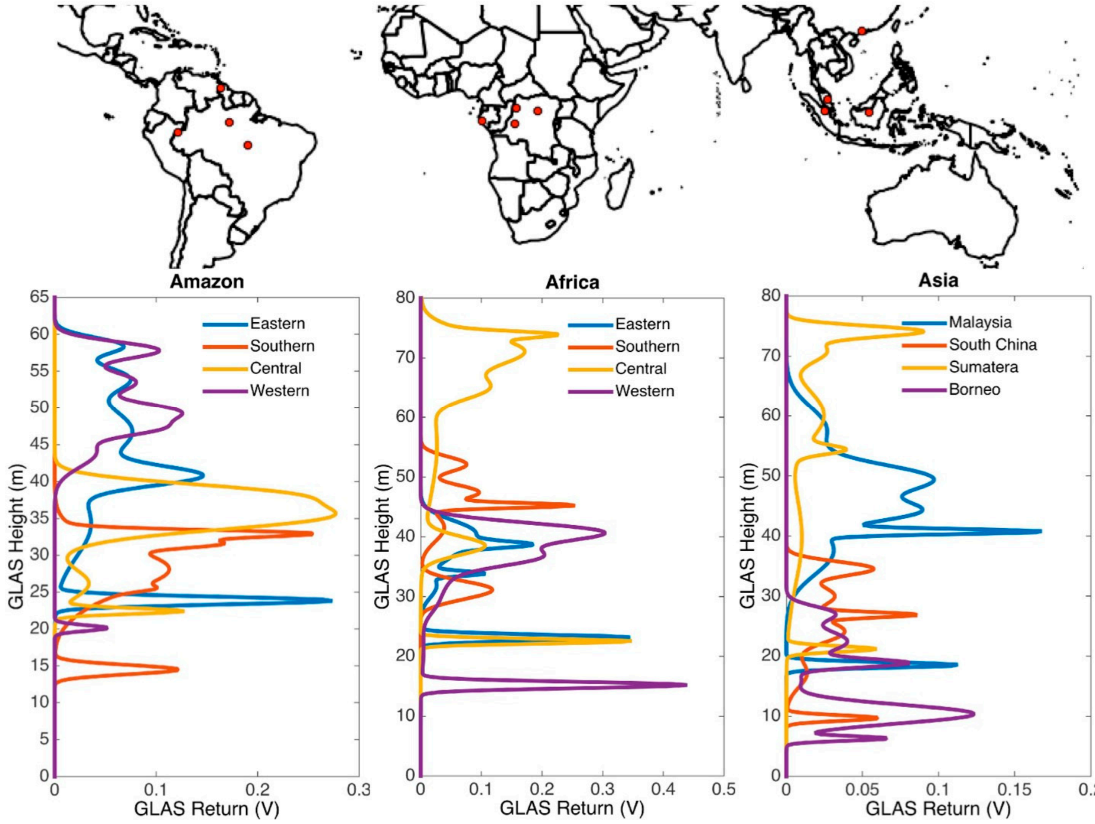



**Figure S3.** Vertical profile of the GLAS footprints in three continents. The upper panel shows the locations of the selected GLAS shots in three continents (the small red circles). The lower panel is the vertical profiles in Amazon (left), in Africa (central), and in Asia (right).

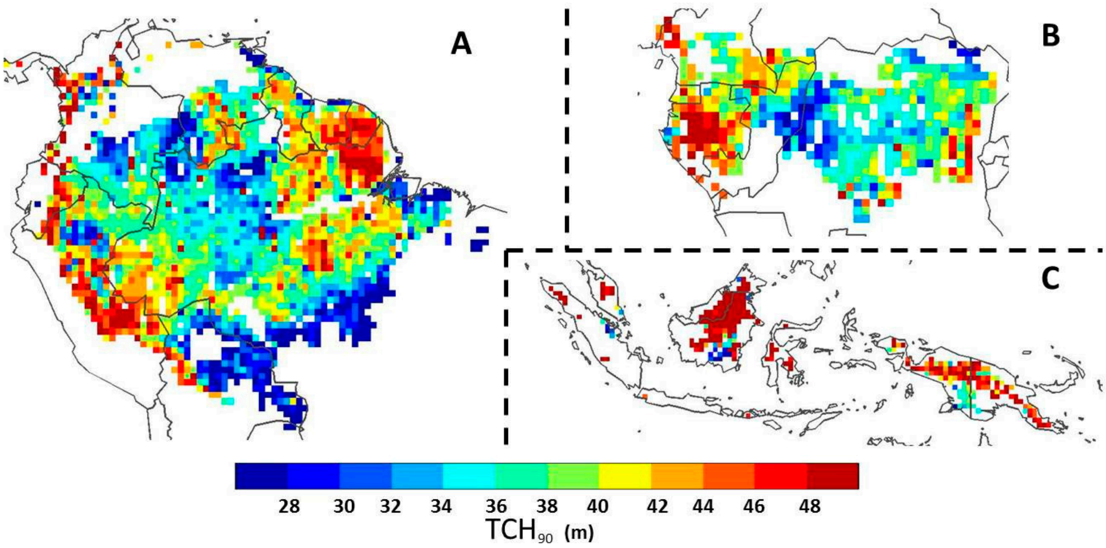



**Figure S4.** $TCH_{90}$ calculated from GLAS dataset in 0.5-deg resolution. (**A**) $TCH_{90}$ of South America; (**B**) $TCH_{90}$ of Central Africa; and (**C**) $TCH_{90}$ of Southeast Asia. Pixels were colored white and marked invalid if there are less than 50 GLAS points available in each pixel.

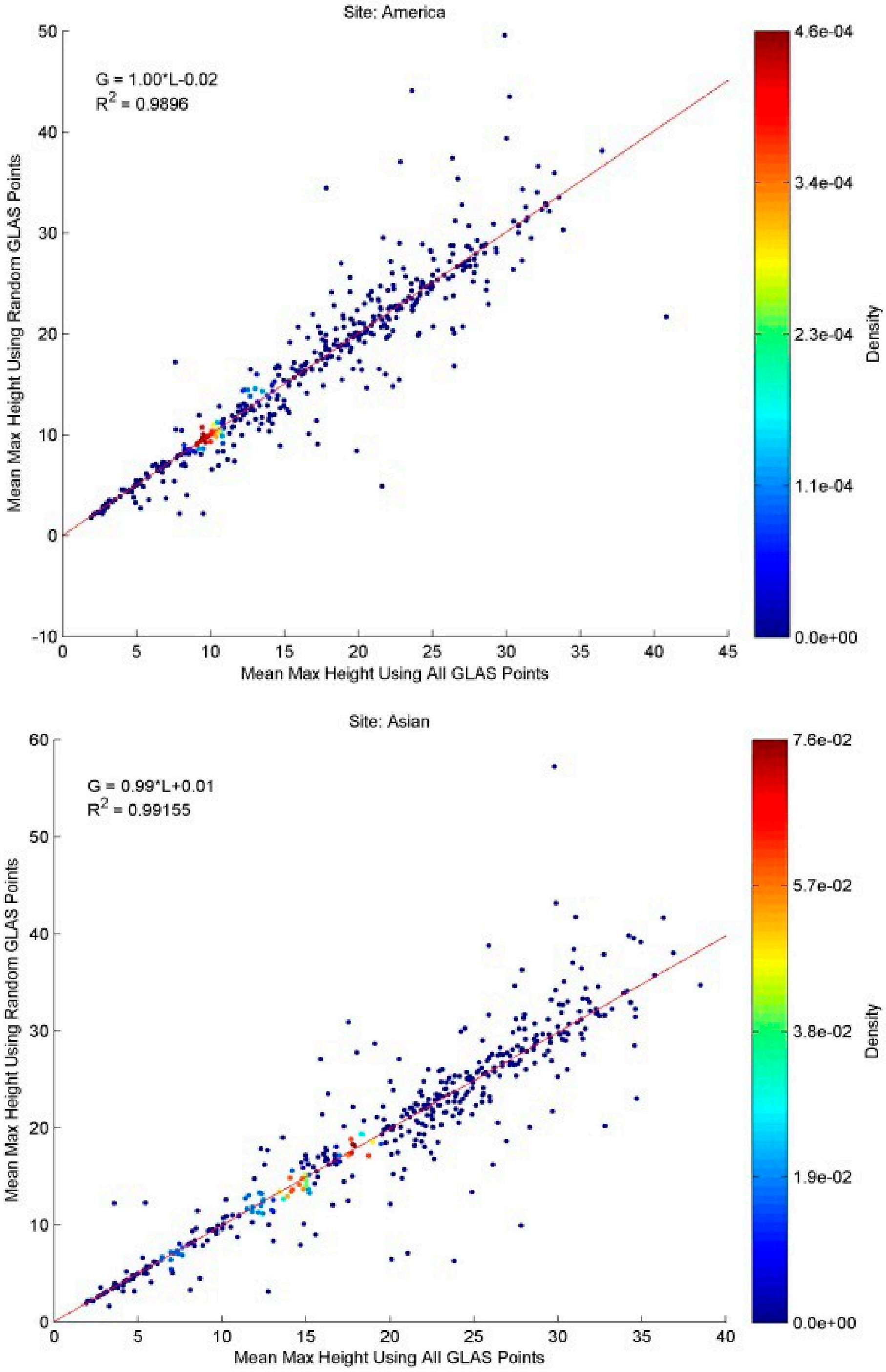



**Figure S5.** Relationship between $TCH_m$ using all GLAS points and $TCH_m$ from random sampling. The Top panel is the scatterplot from America, and the bottom panel is the scatterplot from Asia.

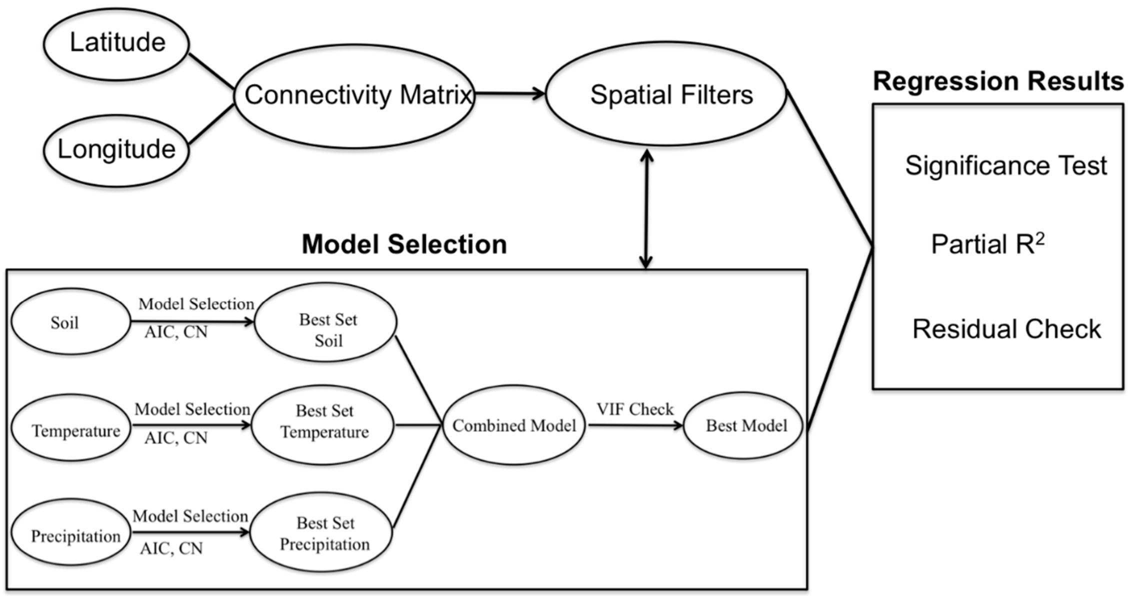



**Figure S6.** Diagram of the processing steps of spatial regression analysis (see Section 2.4).

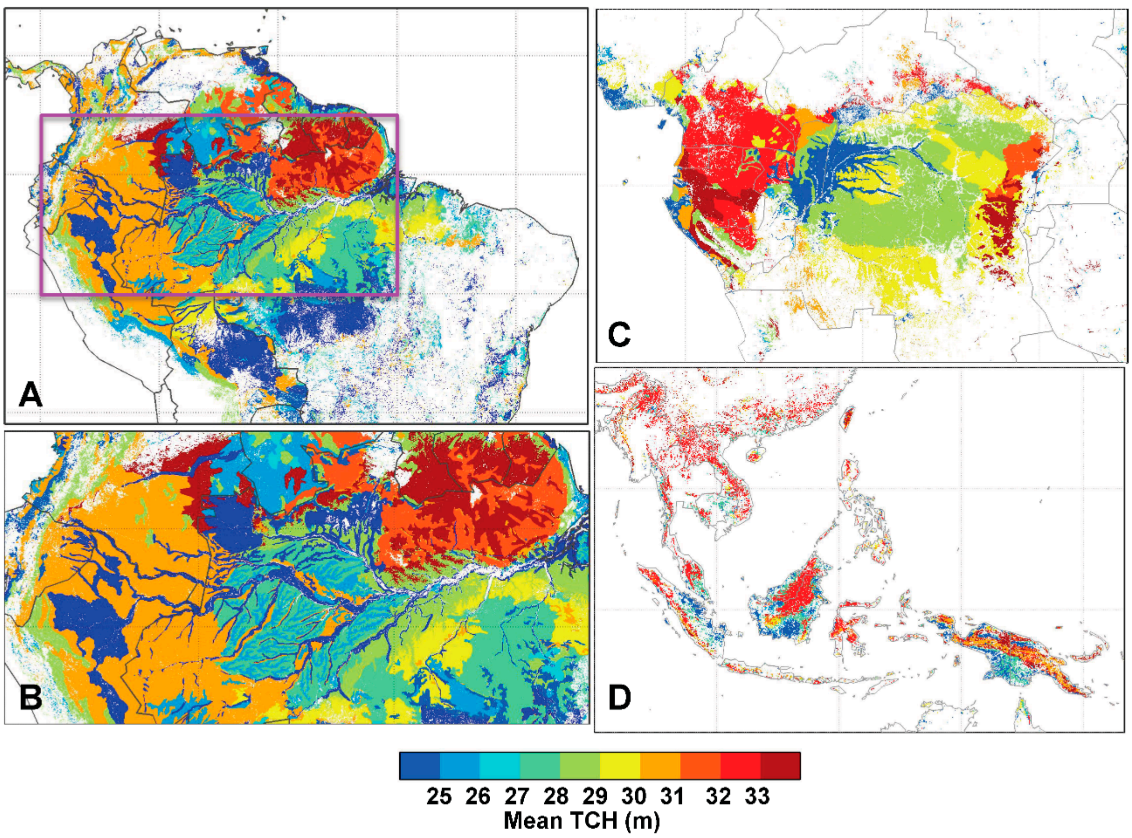



**Figure S7**. Mean TCH calculated from GLAS dataset Based on Soil types. (**A**) $TCH_m$ of America; (**B**) $TCH_m$ in the purple rectangle of panel A; (**C**) $TCH_m$ of Africa; and (**D**) $TCH_m$ in the Asia. Pixels were colored white and marked invalid if there are less than 100 GLAS points available in each soil type.

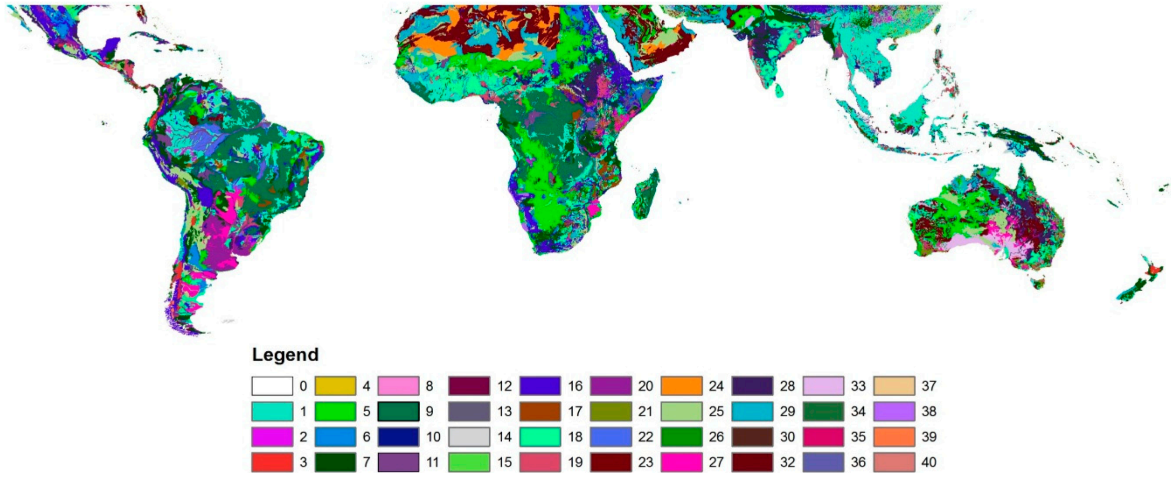



**Figure S8.** The tropical soil classification map that combines FAO 74 and FAO 94 attributes from HWSD database (see Table S1).

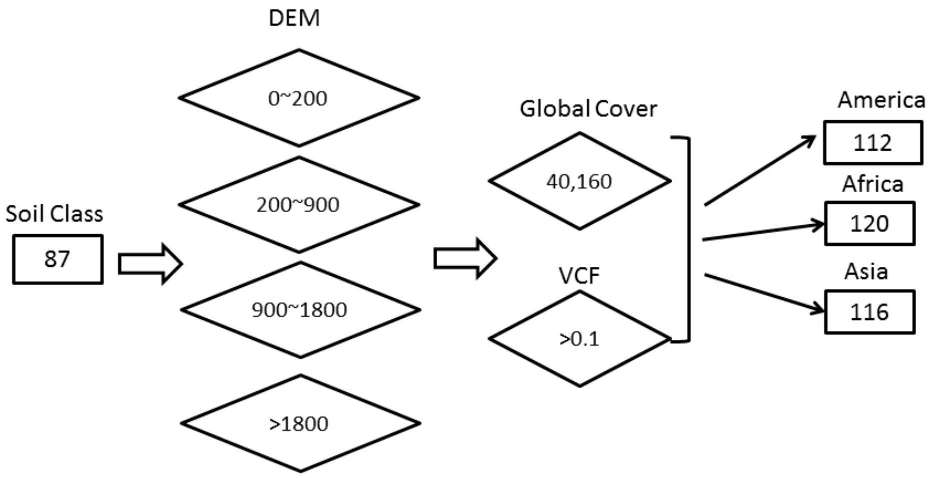



**Figure S9.** Diagram of the ecoregion classification. The number in the rectangle is the number of soil types. The range in the rhombus separates each soil type to further refined soil types based on ground elevation and land cover.

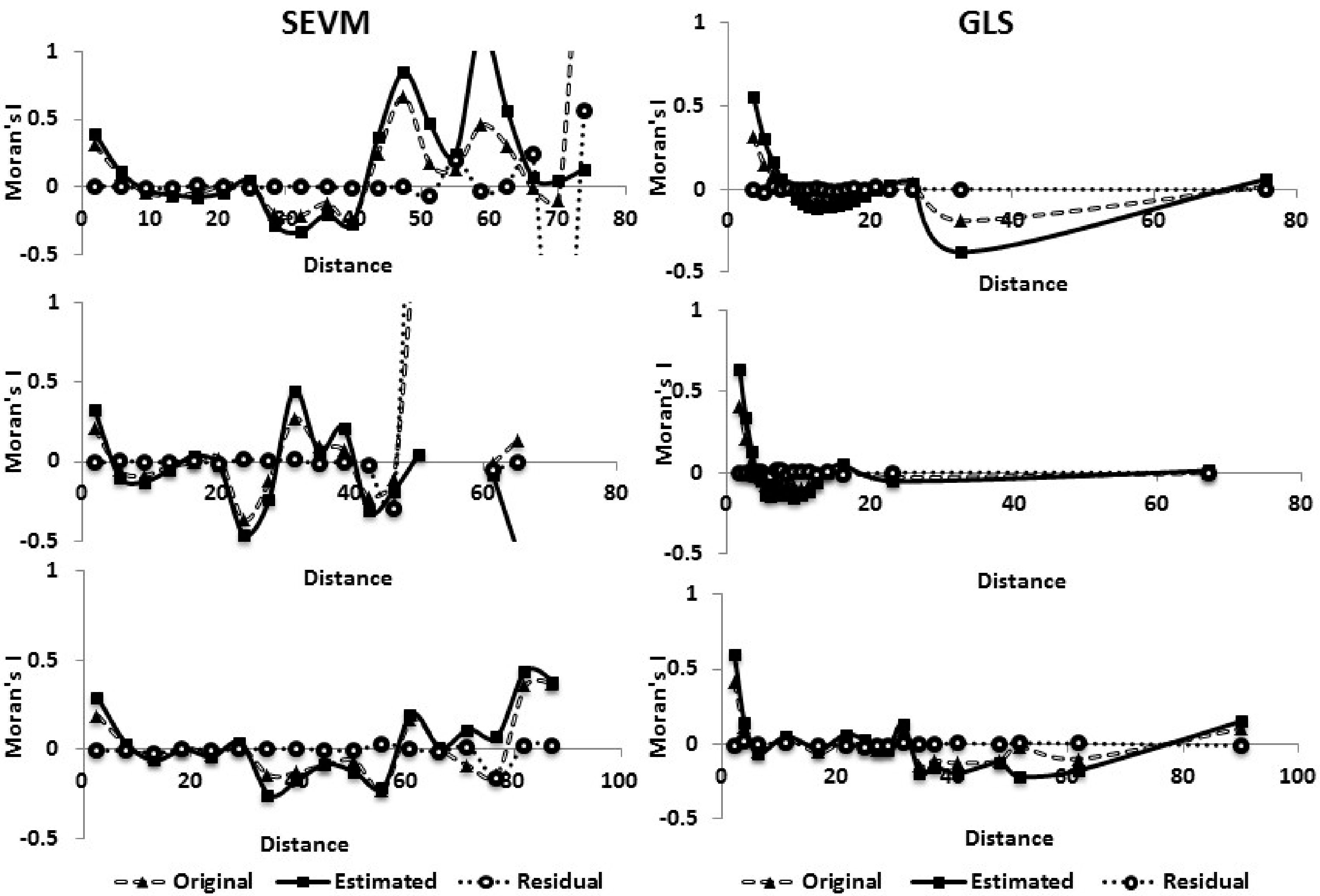



**Figure S10.** Spatial autocorrelations in terms of Moran's I. "Original" curves show spatial autocorrelations existing in the original $TCH_m$ data in America (upper panels), Africa (central panels) and Asia (lower panels). "Estimated" curves show the predicted $TCH_m$ from spatial regression results of SEVM and GLS, while the "Residual" curves show results of spatial regressions that successfully remove the spatial effects.

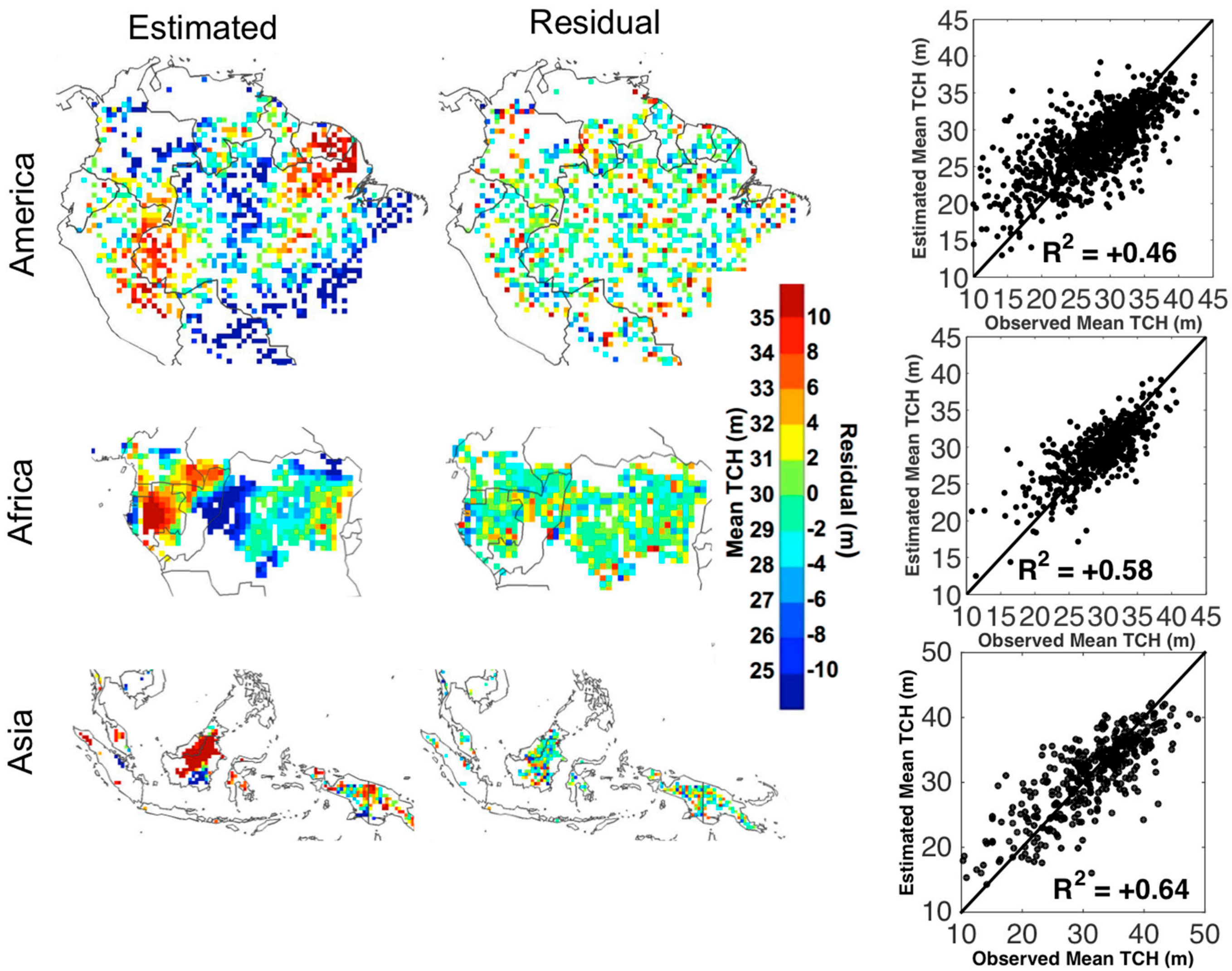



**Figure S11.** Spatial regression results using GLS for tropical forests in America, Africa and Asia between $TCH_m$ and all the selected environmental variables.

**Table S1.** Soil classification used in the ecoregion stratification in the tropics.

| Soil Value | Soil Class Type | Description of Soil Class Type |
|---|---|---|
| **1** | **ACRISOLS (AC)** | Soils with subsurface accumulation of low activity clays and low base saturation |
| **2** | **ALISOLS (AL)** | Soils with sub-surface accumulation of high activity clays, rich in exchangeable aluminum |
| **3** | **ANDOSOLS (AN)** | Young soils formed from volcanic deposits |
| **4** | **ANTHROSOLS (AT)** | Soils in which human activities have resulted in profound modification of their properties |
| **5** | **ARENOSOLS (AR)** | Sandy soils featuring very weak or no soil development |
| **6** | **CALCISOLS (CL)** | Soils with accumulation of secondary calcium carbonates |
| **7** | **CAMBISOLS (CM)** | Weakly to moderately developed soils |
| **8** | **CHERNOZEMS CH)** | Soils with a thick, dark topsoil, rich in organic matter with a calcareous subsoil |
| **9** | **FERRALSOLS (FR)** | Deep, strongly weathered soils with a chemically poor, but physically stable subsoil |
| **10** | **FLUVISOLS (FL)** | Young soils in alluvial deposits |
| **11** | **GLEYSOLS (GL)** | Soils with permanent or temporary wetness near the surface |
| **12** | **GREYZEMS (GR)** | Acid soils with a thick, dark topsoil rich in organic matter |
| **13** | **GYPSISOLS (GY)** | Soils with accumulation of secondary gypsum |
| **14** | **HISTOSOLS (HS)** | Soils which are composed of organic materials |
| **15** | **KASTANOZEMS (KS)** | Soils with a thick, dark brown topsoil, rich in organic matter and a calcareous or gypsum-rich subsoil |
| **16** | **LEPTOSOLS (LP)** | Very shallow soils over hard rock or in unconsolidated very gravelly material |
| **17** | **LIXISOLS (LX)** | Soils with subsurface accumulation of low activity clays and high base saturation |
| **18** | **LUVISOLS (LV)** | Soils with subsurface accumulation of high activity clays and high base saturation |
| **19** | **NITISOLS (NT)** | Deep, dark red, brown or yellow clayey soils having a pronounced shiny, nut-shaped structure |
| **20** | **PHAEOZEMS (PH)** | Soils with a thick, dark topsoil rich in organic matter and evidence of removal of carbonates |
| **21** | **PLANOSOLS (PL)** | Soils with a bleached, temporarily water-saturated topsoil on a slowly permeable subsoil |

**Table S1.** *Cont.*

| | | |
|---|---|---|
| **22** | **PLINTHOSOLS (PT)** | Wet soils with an irreversibly hardening mixture of iron, clay and quartz in the subsoil |
| **23** | **PODZOLS (PZ)** | Acid soils with a subsurface accumulation of iron-aluminum-organic compounds |
| **24** | **PODZOLUVISOLS (PD)** | Acid soils with a bleached horizon penetrating into a clay-rich subsurface horizon |
| **25** | **REGOSOLS (RG)** | Soils with very limited soil development |
| **26** | **SOLONCHAKS (SC)** | Strongly saline soils |
| **27** | **SOLONETZ (SN)** | Soils with subsurface clay accumulation, rich in sodium |
| **28** | **VERTISOLS (VR)** | Dark-colored cracking and swelling clays |
| **29** | **LITHOSOLS** | US a type of azonal soil consisting chiefly of unweathered or partly weathered rock fragments, usually found on steep slopes |
| **30** | **RENDZINAS** | a dark, grayish-brown, humus-rich, intrazonal soil |
| **31** | **RANKERS** | soils developed over non-calcareous material, usually rock |
| **32** | **YERMOSOLS** | semi-desert gray soil arid region |
| **33** | **XEROSOLS** | Soils containing low organic matter; the top layer is of a light color, and underlying layers may contain clayish and/or salt minerals such as carbonates and sulfates. |
| **34** | **Rock Outcrops(RK)** | |
| **35** | **Sand Dunes(DS)** | |
| **36** | **Water Bodies (WR)** | |
| **37** | **Urban, mining, *etc.* (UR)** | |
| **38** | **Glaciers(GG)** | |
| **39** | **No data (NI)** | |
| **40** | **IS** | |
| **41** | **HD** | |

**Table S2.** Separation of Ferralsols soil type into 7 classes in the African forests.

| Soil Value | Soil Type |
|---|---|
| 1 | Haplic Ferralsols (FRh) |
| 2 | Xanthic Ferralsols (FRx) |
| 3 | Rhodic Ferralsols (FRr) |
| 4 | Humic Ferralsols(FRu) |
| 5 | Geric Ferralsols(FRg) |
| 6 | Plinthic Ferralsols(FRp) |
| 7 | Orthic Ferralsols(Fo) |

**Table S3.** Spatial regression results using GLS method for $TCH_m$.

| $TCH_m$ (GLS) | | | | | |
|---|---|---|---|---|---|
| America | | Africa | | Asia | |
| Variable | Coeff. | Variable | Coeff. | Variable | Coeff. |
| CEC_T | −0.12 * | CEC_T | −0.262 * | CEC_S | −0.236 *** |
| SILT_S | −0.08 * | SILT_T | 0.08 | CLAY_T | 0.185 ** |
| OC_T | −0.09 | OC_T | 0.253 ** | PH_T | 0.131 ** |
| OC_S | −0.11 ** | CLAY_S | 0.055 | SAND_S | −0.092 * |
| CALY_S | 0.164 *** | PH_T | 0.142 * | E | −0.11 |
| PH_T | −0.11 ** | E | −0.295 ** | M Diural Range | 0.076 |
| SAND_T | 0.196 *** | T Seasonality | 0.208 | Max T warmest m | −0.08 |
| SAND_S | −0.05 | Max T warmest m | −0.073 | P seasonality | −0.364 *** |
| E | −0.09 | T Annual Range | 0.009 | P wettest Q | 0.153 |
| Isothermality | 0.016 | P seasonality | −0.05 | P warmest Q | 0.216 * |
| T Annual Range | −0.04 | P warmest Q | −0.05 | STRM | −0.037 |
| M T warmest Q | −0.05 | P coldest Q | −0.206 * | STRM SD | 0.051 |
| Annual P | 0.072 | STRM | 0.035 | | |
| P seasonality | −0.14 * | STRM SD | 0.154 * | | |
| P warmest Q | <0.001 | | | | |
| P coldest Q | −0.04 | | | | |
| STRM | −0.08 | | | | |
| STRM SD | 0.1 | | | | |

* $p$-Value < 0.05; ** $p$-Value < 0.01; *** $p$-Value < 0.001.

**Table S4.** Spatial regression results using GLS method for $TCH_{90}$.

| $TCH_{90}$ (GLS) | | | | | |
|---|---|---|---|---|---|
| America | | Africa | | Asia | |
| Variable | Coeff. | Variable | Coeff. | Variable | Coeff. |
| CEC_S | −0.074 | CEC_T | 0.02 | CEC_S | −0.171 * |
| OC_T | −0.143 *** | SILT_S | <0.001 | OC_T | −0.071 |
| OC_S | −0.126 *** | OC_S | −0.014 | CLAY_T | 0.139 * |
| CLAY_S | 0.149 *** | CLAY_S | 0.033 | PH_T | 0.105 * |
| PH_T | −0.109 ** | PH_T | 0.047 | SAND_T | −0.09 * |
| SAND_T | 0.145 *** | SAND_S | <0.001 | E | −0.461 *** |
| E | −0.165 * | E | −0.251 *** | M Diural Range | −0.026 |
| Max T Warmest m | −0.138 | Max T Warmest m | 0.002 | Min T Coldest m | −0.542 *** |
| M T driest Q | 0.018 | T Annual Range | 0.079 | P seasonality | −0.335 *** |
| P driest M | 0.167 ** | P westtest M | −0.189 * | P warmest Q | 0.226 ** |
| P wettest Q | 0.029 | P driest Q | −0.214 * | STRM | −0.049 |
| STRM | −0.044 | P warmest Q | 0.133 | STRM SD | 0.083 |
| STRM SD | 0.19 *** | STRM | −0.063 | | |
| | | STRM SD | 0.361 *** | | |

* $p$-Value < 0.05; ** $p$-Value < 0.01; *** $p$-Value < 0.001.

**Table S5.** Example regression coefficients table of TCHm using SEVM method in Africa.

| Variable | Coeff. | Std Coeff. | VIF | Std Error | *t* | *p* Value |
|---|---|---|---|---|---|---|
| Constant | 29.323 | 0 | 0 | 0.124 | 236.413 | 0 |
| CEC_T | −1.471 | −0.312 | 8.849 | 0.369 | −3.983 | <0.001 |
| SILT_T | −0.033 | −0.007 | 3.335 | 0.227 | −0.144 | 0.886 |
| OC_T | 1.088 | 0.231 | 5.499 | 0.291 | 3.737 | <0.001 |
| CLAY_S | 0.374 | 0.079 | 2.17 | 0.183 | 2.046 | 0.041 |
| PH_T | 0.731 | 0.155 | 1.885 | 0.17 | 4.288 | <0.001 |
| E | −1.461 | −0.31 | 2.822 | 0.209 | −7.005 | <0.001 |
| T Seasonality | 1.116 | 0.237 | 5.535 | 0.292 | 3.82 | <0.001 |
| Max T warmest m | 0.196 | 0.042 | 5.088 | 0.28 | 0.701 | 0.484 |
| T Annual Range | −0.506 | −0.108 | 3.379 | 0.228 | −2.218 | 0.027 |
| P seasonality | −1.155 | −0.245 | 3.632 | 0.237 | −4.882 | <0.001 |
| P warmest Q | −0.061 | −0.013 | 3.672 | 0.238 | −0.258 | 0.797 |
| P coldest Q | −0.926 | −0.197 | 3.016 | 0.216 | −4.295 | <0.001 |
| STRM | −0.698 | −0.148 | 6.537 | 0.317 | −2.198 | 0.028 |
| STRM_SD | 1.72 | 0.365 | 2.764 | 0.206 | 8.334 | <0.001 |
| LCF | 1 | 0.718 | 1.648 | 0.047 | 21.199 | 0 |