

Supplementary Information

Regional patterns and environmental controls of topsoil pH in a karst mountainous landscape of Southwest China

Shihan Wang^{1,2}, Hongbin Liu^{1,2,*}

¹ College of Resources and Environment, Southwest University, Chongqing 400716, China

² Chongqing Key Laboratory of Digital Agriculture, Chongqing 400716, China

* Correspondence: lhbin@swu.edu.cn

Supplementary Table S1. Detailed description of the 59 primary environmental covariates used in the modelling framework.

Domain	Covariate Description	Abbreviation	Data Type	Resolution	Source / Derivation
Terrain	Elevation (m)	Ele	Continuous	30 m	SRTM DEM (USGS)
Terrain	Analytical Hillshading	AH	Continuous	30 m	Derived via SAGA GIS
Terrain	Aspect (°)	Asp	Continuous	30 m	Derived via SAGA GIS
Terrain	Convergence Index	CI	Continuous	30 m	Derived via SAGA GIS
Terrain	Channel Network Base Level (m)	CNBL	Continuous	30 m	Derived via SAGA GIS
Terrain	Channel Network Distance (m)	CND	Continuous	30 m	Derived via SAGA GIS
Terrain	LS-Factor	LS	Continuous	30 m	Derived via SAGA GIS
Terrain	Plan Curvature (m ⁻¹)	PLC	Continuous	30 m	Derived via SAGA GIS
Terrain	Profile Curvature (m ⁻¹)	PRC	Continuous	30 m	Derived via SAGA GIS
Terrain	Relative Slope Position	RSP	Continuous	30 m	Derived via SAGA GIS
Terrain	Slope (°)	Slp	Continuous	30 m	Derived via SAGA GIS
Terrain	Total Catchment Area (m ²)	TCA	Continuous	30 m	Derived via SAGA GIS
Terrain	Topographic Wetness Index	TWI	Continuous	30 m	Derived via SAGA GIS
Terrain	Valley Depth (m)	VD	Continuous	30 m	Derived via SAGA GIS
Terrain	Terrain Ruggedness Index (m)	TRI	Continuous	30 m	Derived via SAGA GIS
Terrain	Melton's Ruggedness Number	MRN	Continuous	30 m	Derived via SAGA GIS
Terrain	Mid-Slope Position	MSP	Continuous	30 m	Derived via SAGA GIS
Terrain	Vertical Distance to Channel Network (m)	VDCN	Continuous	30 m	Derived via SAGA GIS
Terrain	Topographic Relief (m)	TR	Continuous	30 m	Derived via SAGA GIS
Terrain	Slope Length (m)	Slp_l	Continuous	30 m	Derived via SAGA GIS
Terrain	Slope Height (m)	Slp_h	Continuous	30 m	Derived via SAGA GIS
Terrain	Multi-resolution Ridge Top Flatness	MRRTF	Continuous	30 m	Derived via SAGA GIS
Terrain	Multi-resolution Valley Bottom Flatness	MRVBF	Continuous	30 m	Derived via SAGA GIS
Terrain	100m radius Topographic Position Index (m)	TPI100m	Continuous	30 m	Derived via SAGA GIS
Terrain	500m radius Topographic Position Index (m)	TPI500m	Continuous	30 m	Derived via SAGA GIS
Terrain	1000m radius Topographic Position Index (m)	TPI1000m	Continuous	30 m	Derived via SAGA GIS
Terrain	Closed Depression Index	CD	Continuous	30 m	Derived via SAGA GIS
Climate	Spring Precipitation (mm)	Pre_spr	Continuous	1000 m	ERA5-Land (2000–2020)
Climate	Summer Precipitation (mm)	Pre_sum	Continuous	1000 m	ERA5-Land (2000–2020)
Climate	Autumn Precipitation (mm)	Pre_fal	Continuous	1000 m	ERA5-Land (2000–2020)
Climate	Winter Precipitation (mm)	Pre_win	Continuous	1000 m	ERA5-Land (2000–2020)
Climate	Spring Max Temperature (°C)	Tmax_spr	Continuous	1000 m	ERA5-Land (2000–2020)
Climate	Summer Max Temperature (°C)	Tmax_sum	Continuous	1000 m	ERA5-Land (2000–2020)
Climate	Autumn Max Temperature (°C)	Tmax_fal	Continuous	1000 m	ERA5-Land (2000–2020)
Climate	Winter Max Temperature (°C)	Tmax_win	Continuous	1000 m	ERA5-Land (2000–2020)

Domain	Covariate Description	Abbreviation	Data Type	Resolution	Source / Derivation
Climate	Spring Min Temperature (°C)	Tmin_spr	Continuous	1000 m	ERA5-Land (2000–2020)
Climate	Summer Min Temperature (°C)	Tmin_sum	Continuous	1000 m	ERA5-Land (2000–2020)
Climate	Autumn Min Temperature (°C)	Tmin_fal	Continuous	1000 m	ERA5-Land (2000–2020)
Climate	Winter Min Temperature (°C)	Tmin_win	Continuous	1000 m	ERA5-Land (2000–2020)
Climate	Mean Land Surface Temperature (°C)	LST_m	Continuous	1000 m	MOD11A2 / GEE (2000–2020)
Climate	Mean Sensible Heat Flux (W/m²)	H_m	Continuous	1000 m	ERA5-Land (2000–2020)
Climate	Mean Evaporation (mm)	E_m	Continuous	1000 m	ERA5-Land (2000–2020)
Vegetation	Normalized Difference Vegetation Index	NDVI	Continuous	250 m	MOD13Q1 / GEE
Vegetation	Land Surface Water Index	LSWI	Continuous	250 m	MOD09GA / GEE
Vegetation	Enhanced Vegetation Index	EVI	Continuous	250 m	MOD13Q1 / GEE
Vegetation	Modified Normalized Difference Water Index	MNDWI	Continuous	250 m	Sentinel-2 / GEE
Vegetation	Soil Adjusted Vegetation Index	SAVI	Continuous	250 m	MOD13Q1 / GEE
Vegetation	Bare Soil Index	BSI	Continuous	250 m	Sentinel-2 / GEE
Vegetation	Green Normalized Difference Vegetation Index	GNDVI	Continuous	250 m	Sentinel-2 / GEE
Vegetation	Modified Soil Adjusted Vegetation Index	MSAVI	Continuous	250 m	Sentinel-2 / GEE
Vegetation	Normalized Difference Red Edge Index	NDRE	Continuous	250 m	Sentinel-2 / GEE
Vegetation	Net Primary Productivity (g C/m²/yr)	NPP	Continuous	250 m	MOD17A3H / GEE
Geology	Geological strata	SCID	Categorical	1:200,000	Regional geological maps
Geology	Lithology	LIID	Categorical	1:50,000	Regional geological maps
Geology	Parent material	PSID	Categorical	30 m	Regional geological maps
Soil	Sand content (%)	Sand	Continuous	30 m	Soil surveys
Soil	Silt content (%)	Silt	Continuous	30 m	Soil surveys
Soil	Clay content (%)	Clay	Continuous	30 m	Soil surveys
Human Act.	Land use type	LUID	Categorical	30 m	High-res land-use products

Note: The 59 primary covariates comprised 27 terrain, 15 climate, 10 vegetation, 3 geology and parent-material, 3 soil-texture, and 1 land-use variable. The categorical variables SCID, LIID, PSID, and LUID were transformed into 143 numerical features by one-hot encoding, yielding 198 candidate predictors in total.

Supplementary Table S2. Classification mapping of the parent-material (PSID) categorical variable after one-hot encoding.

Feature	Parent-material class	Feature	Parent-material class
PSID_1	Quaternary (第四系)	PSID_36	Lower Silurian Hanjiadian Formation (志留系下统韩家店组)
PSID_2	Upper Cambrian Gengjiadian Formation (寒武系上统耿家店组)	PSID_37	Lower Silurian Xintan Formation (志留系下统新滩组)
PSID_3	Middle Triassic Badong Formation (三叠系中统巴东组)	PSID_38	Lower Silurian Hanjiadian Formation (志留系下统韩家店组)
PSID_4	Middle Triassic Badong Formation (三叠系中统巴东组)	PSID_39	Middle Triassic Badong Formation (三叠系中统巴东组)
PSID_5	Lower Triassic Daye Formation (三叠系下统大冶组)	PSID_40	Lower Silurian Hanjiadian Formation (志留系下统韩家店组)
PSID_6	Middle Ordovician Guniutan Formation (奥陶系中统牯牛潭组)	PSID_41	Middle Triassic Badong Formation (三叠系中统巴东组)
PSID_7	Lower Triassic Jialingjiang Formation (三叠系下统嘉陵江组)	PSID_42	Lower Silurian Hanjiadian Formation (志留系下统韩家店组)
PSID_8	Upper Permian Changxing Formation (二叠系上统长兴组)	PSID_43	Middle Triassic Badong Formation (三叠系中统巴东组)
PSID_9	Lower Triassic Jialingjiang Formation (三叠系下统嘉陵江组)	PSID_44	Middle Cretaceous Zhengyang Formation (白垩系中统正阳组)
PSID_10	Upper Permian Wujiaping Formation (二叠系上统吴家坪组)	PSID_45	Lower Silurian Shiniulan Formation (志留系下统石牛栏组)
PSID_11	Lower Triassic Jialingjiang Formation (三叠系下统嘉陵江组)	PSID_46	Lower Silurian Rongxi Formation (志留系下统溶溪组)
PSID_13	Lower Silurian Huixingshao Formation (志留系下统回星哨组)	PSID_47	Lower Silurian Xiushan Formation (志留系下统秀山组)
PSID_16	Upper Cambrian Maotian Formation (寒武系上统毛田组)	PSID_48	Middle Cambrian Gaotai Formation (寒武系中统高台组)
PSID_18	Lower Silurian Xiaoheba Formation (志留系下统小河坝组)	PSID_49	Lower Cambrian Palang Formation (寒武系下统杷榔组)
PSID_19	Lower Silurian Longmaxi Formation (志留系下统龙马溪组)	PSID_50	Middle Permian Liangshan Formation (二叠系中统梁山组)
PSID_20	Middle Cambrian Shilengshui Formation (寒武系中统石冷水组)	PSID_51	Lower Cambrian Bianmachong Formation (寒武系下统变马冲组)
PSID_21	Lower Triassic Jialingjiang Formation (三叠系下统嘉陵江组)	PSID_52	Lower Cambrian Qingxudong Formation (寒武系下统清虚洞组)
PSID_22	Middle-Lower Ordovician Dawan Formation (奥陶系中下统大湾组)	PSID_54	Middle Cambrian Pingjing Formation (寒武系中统平井组)
PSID_23	Lower Ordovician Fenxiang Formation (奥陶系下统分乡组)	PSID_55	Middle Jurassic Xintiangou Formation (侏罗系中统新田沟组)
PSID_24	Upper Devonian Shuicheping Formation (泥盆系上统水车坪组)	PSID_56	Lower Silurian Longmaxi Formation (志留系下统龙马溪组)
PSID_25	Lower Jurassic Ziliujing Formation (侏罗系下统自流井组)	PSID_57	Upper Ordovician Wufeng Formation (奥陶系上统五峰组)
PSID_26	Lower Jurassic Ziliujing Formation (侏罗系下统自流井组)	PSID_58	Lower Silurian Longmaxi Formation (志留系下统龙马溪组)

Feature	Parent-material class	Feature	Parent-material class
PSID_27	Lower Jurassic Zhenzhuchong Formation (侏罗系下统珍珠冲组)	PSID_59	Middle-Upper Ordovician Guniutan Formation (奥陶系中上统牯牛潭组)
PSID_28	Lower Jurassic Ziliujing Formation (侏罗系下统自流井组)	PSID_60	Upper Permian Wujiaping Formation (二叠系上统吴家坪组)
PSID_29	Upper Triassic Xujiache Formation (三叠系上统须家河组)	PSID_61	Lower Triassic Jialingjiang Formation (三叠系下统嘉陵江组)
PSID_30	Middle Jurassic Shaximiao Formation (侏罗系中统沙溪庙组)	PSID_63	Upper Ordovician (奥陶系上统)
PSID_31	Middle Permian Qixia Formation (二叠系中统栖霞组)	PSID_64	Upper Ordovician Baota Formation (奥陶系上统宝塔组)
PSID_32	Upper Permian Wujiaping Formation (二叠系上统吴家坪组)	PSID_65	Lower Ordovician Honghuayuan Formation (奥陶系下统红花园组)
PSID_33	Middle Permian Maokou Formation (二叠系中统茅口组)	PSID_67	Lower Ordovician Nanjinguan Formation (奥陶系下统南津关组)
PSID_34	Middle Permian Liangshan Formation (二叠系中统梁山组)	PSID_68	Lower Ordovician Tongzi Formation (奥陶系下统桐梓组)
PSID_35	Upper Permian Wujiaping Formation (二叠系上统吴家坪组)	PSID_69	Middle-Upper Ordovician (奥陶系中上统)

Note: The original Chinese stratigraphic nomenclature was translated into standard English geological formations. Multiple encoded IDs assigned to the same formation represent spatially distinct sub-members or separate polygons in the original geological vector data.

Supplementary Table S3. Classification mapping of the lithology (LIID) categorical variable after one-hot encoding.

Feature	Lithology Class	Feature	Lithology Class
LIID_1	Alluvium (冲积物)	LIID_5	Shale (页岩)
LIID_2	Dolomite (白云岩)	LIID_6	Sandstone (砂岩)
LIID_3	Limestone (灰岩)	LIID_7	Conglomerate (砾岩)
LIID_4	Mudstone (泥岩)	LIID_8	Limestone and Shale (灰岩、页岩)

Note: Lithology classes retain the codes used in the modelling dataset.

Supplementary Table S4. Classification mapping of the land-use (LUID) categorical variable after one-hot encoding.

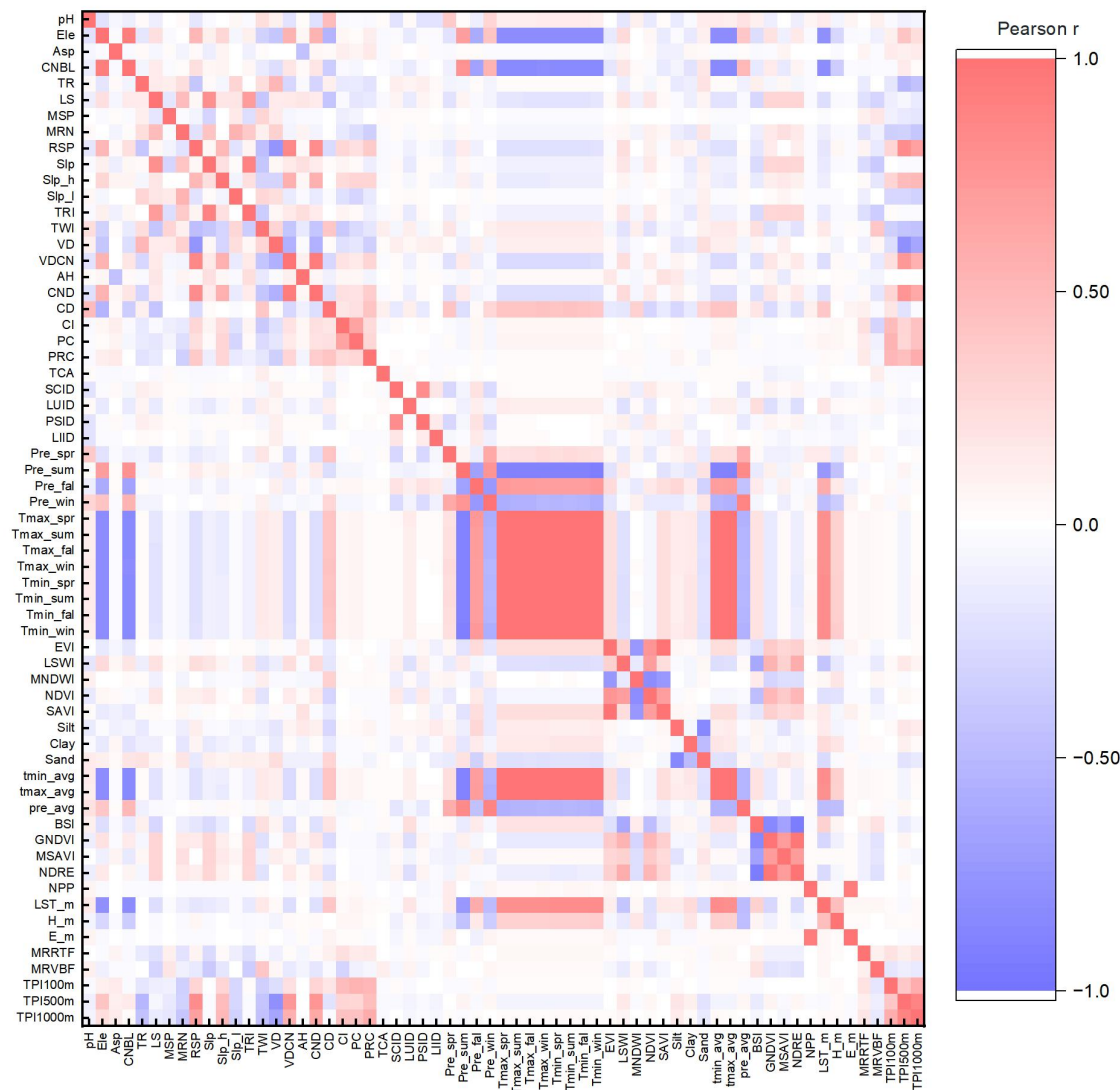
Feature	Land Use Class	Feature	Land Use Class
LUID_2	Arbor forest (乔木林地)	LUID_7	Other grassland (其他草地)
LUID_3	Bamboo forest (竹林地)	LUID_8	Other orchard (其他园地)
LUID_4	Shrub forest (灌木林地)	LUID_17	Other forest land (其他林地)
LUID_5	Dry land (旱地)	LUID_19	Orchard (果园)
LUID_6	Paddy field (水田)	LUID_22	Tea garden (茶园)

Note: Land-use classes follow the Chinese Third National Land Survey classification (GB/T 21010-2017).

Supplementary Table S5. Classification mapping of the geological-strata (SCID) categorical variable after one-hot encoding.

Feature	Geological Code (DCDM)	Feature	Geological Code (DCDM)
SCID_1	Qh	SCID_37	S1x
SCID_2	E3g	SCID_38	S1h3
SCID_3	T2b1	SCID_39	T2b
SCID_4	T2b2	SCID_40	S1h2
SCID_5	T1d	SCID_41	T2b3
SCID_6	O2g	SCID_42	S1h1
SCID_7	T1j2	SCID_43	T2b4
SCID_8	P3c	SCID_44	K2z
SCID_9	T1j1	SCID_45	S1s
SCID_10	P3w	SCID_46	S1r
SCID_11	T1j4	SCID_47	S1xs
SCID_13	S1hx	SCID_48	E2g
SCID_16	E3m	SCID_49	E1p
SCID_18	S1xh	SCID_50	P2l-q
SCID_19	S1l	SCID_51	E1b
SCID_20	E2s	SCID_52	E1q
SCID_21	T1j3	SCID_54	E2p
SCID_22	O1-2d	SCID_55	J2x
SCID_23	O1f	SCID_56	S1l1
SCID_24	D3s	SCID_57	O3w-g
SCID_25	J1zlda	SCID_58	S1l2
SCID_26	J1zlm	SCID_59	O2-3g-d
SCID_27	J1z	SCID_60	P3w-c
SCID_28	J1zld	SCID_61	T1j3-4
SCID_29	T3xj	SCID_63	O3
SCID_30	J2s	SCID_64	O3b-l
SCID_31	P2q	SCID_65	O1h
SCID_32	P3w1	SCID_66	O1n
SCID_33	P2m	SCID_67	O1n-f
SCID_34	P2l	SCID_68	O1t
SCID_35	P3w2	SCID_69	O2-3
SCID_36	S1h		

Note: Geological codes follow standard stratigraphic nomenclature; prefixes denote geological periods, including Q (Quaternary), T (Triassic), P (Permian), S (Silurian), O (Ordovician), and E (Cambrian).



Supplementary Figure S1. Pearson correlation matrix of topsoil pH and environmental covariates. Colours represent Pearson's correlation coefficient (r), ranging from -1 (blue) to +1 (red), with white indicating zero. Correlations were calculated using all 4,578 valid observations.