**Supplementary data**

**This document contains**



**SI 1 Supplementary Results**

Four fitted equations for $CEC_{silt}$ by using MLR were expressed as follows. The description of four datasets were given in Environmental variables section in the manuscript, in which the $CEC_{silt}$ was dependent variable rather than $CEC_{clay}$.

$$\begin{cases} \text{Dataset 1: } CEC_{silt} = 3.464 + 0.505 \times CEC_{soil} - 0.628 \times pH - 0.199 \times SOC + 0.053 \times Fe_d + 0.045 \times Sand - 0.126 \times Clay \\ \text{Dataset 2: } CEC_{silt} = 11.328 - 0.224 \times CEC_{clay} + 0.698 \times CEC_{soil} - 0.887 \times pH - 0.213 \times SOC + 0.034 \times Fe_d - 0.168 \times Clay \\ \text{Dataset 3: } CEC_{silt} = 1.258 + 0.531 \times CEC_{soil} - 0.664 \times pH - 0.228 \times SOC + 0.053 \times Fe_d + 0.037 \times Sand - 0.128 \times Clay + 0.001 \times MAT \\ \text{Dataset 4: } CEC_{silt} = 9.582 - 0.208 \times CEC_{clay} + 0.698 \times CEC_{soil} - 0.886 \times pH - 0.227 \times SOC + 0.035 \times Fe_d - 0.165 \times Clay + 0.0007 \times MAT \end{cases}$$

Similar with $CEC_{silt}$, four fitted equations for $CEC_{soil}$ by using MLR were expressed as follows.

$$\begin{cases} \text{Dataset 1: } CEC_{soil} = 6.435 + 1.342 \times pH + 1.026 \times Fe_d - 0.059 \times Sand - 0.072 \times Clay \\ \text{Dataset 2: } CEC_{soil} = -28.990 + 0.138 \times CEC_{clay} + 0.267 \times CEC_{silt} + 0.648 \times CEC_{Min} + 0.491 \times pH + 0.101 \times SOC + 0.241 \times Sand \\ \qquad + 0.237 \times Silt + 0.261 \times Clay \\ \text{Dataset 3: } CEC_{soil} = 16.124 + 1.095 \times pH + 1.231 \times Fe_d - 0.048 \times Clay - 0.0002 \times MAP - 0.004 \times MAT \\ \text{Dataset 4: } CEC_{soil} = -28.990 + 0.138 \times CEC_{clay} + 0.267 \times CEC_{silt} + 0.648 \times CEC_{Min} + 0.491 \times pH + 0.101 \times SOC + 0.241 \times Sand \\ \qquad + 0.237 \times Silt + 0.261 \times Clay \end{cases}$$

Four fitted equations for $CEC_{Min}$ by using MLR were expressed as follows.

$$\begin{cases} \text{Dataset 1: } CEC_{Min} = 40.764 + 0.974 \times CEC_{soil} - 0.403 \times Sand - 0.399 \times Silt - 0.359 \times Clay \\ \text{Dataset 2: } CEC_{Min} = 39.459 + 0.072 \times CEC_{clay} + 0.923 \times CEC_{soil} - 0.398 \times Sand - 0.404 \times Silt - 0.349 \times Clay \\ \text{Dataset 3: } CEC_{Min} = 40.764 + 0.974 \times CEC_{soil} - 0.403 \times Sand - 0.399 \times Silt - 0.359 \times Clay \\ \text{Dataset 4: } CEC_{Min} = 39.459 + 0.072 \times CEC_{clay} + 0.923 \times CEC_{soil} - 0.398 \times Sand - 0.404 \times Silt - 0.349 \times Clay \end{cases}$$

## SI 2 Supplementary Tables

**Table S1.** Main soil types and diagnostic horizons with the Soil Taxonomy [3], in which the values of the $CEC_{clay}$ (by $NH_4OAc$ pH 7) and ECEC are required.

| Diagnostic horizons | Diagnostic criteria | Soil type [a)] | |
|---|---|---|---|
| | | **Soil orders** | **Great groups** |
| Kandic Horizon | Has an apparent CEC of 16 cmol(+) or less per kg clay (by 1N $NH_4OAc$ pH 7) and an apparent ECEC of 12 cmol(+) or less per kg clay. | Alfisols | Kandiaqualfs, Palecryalfs, Ferrudalfs, Kandiudalfs, Kanhapludalfs, Kandiustalfs, Kanhaplustalfs, Rhodoxeralfs, Palexeralfs |
| | | Oxisols | Acraquox, Acroperox, Kandiperox, Acroperox, Acrotorrox, Acrudox, Kandiudox, Acrustox, Kandiustox |
| | | Ultisols | Albaquults, Kandiaquults, Kanhaplaquults, Kandihumults, Kanhaplohumults, Kandiudults, Kanhapludults, Kandiustults, Kanhaplustults |
| Oxic Horizon | Has an apparent CEC of 16 cmol(+) or less per kg clay (by 1N $NH_4OAc$ pH 7) and an apparent ECEC of 12 cmol(+) or less per kg clay | Oxisols | Acraquox, Acroperox, Acrotorrox, Acrudox, Acrustox |

[a)] The $CEC_{clay}$ is also used as a criterion in the Kandic and Kanhaplic subgroups of Alfisols and Ultisols, the Udoxic and Ustoxic subgroups of Quartzipsamments, and the Oxic subgroups of Inceptisols and Mollisols.

**Table S2.** Reference soil groups and qualifiers in the World Reference Base for Soil Resources (WRB) [4], in which the values of the $CEC_{clay}$ (by $NH_4OAc$ pH 7) are required.

| **Soil type or Qualifiers** | **Definitions of the criteria** |
|---|---|
| *Reference Soil Groups* | |
| Ferralsols | Having a CEC (by 1 *M* $NH_4OAc$, pH 7) of < 16 $cmol_c$ $kg^{-1}$ clay and a sum of exchangeable bases (by 1 *M* $NH_4OAc$, pH 7) plus exchangeable Al (by 1 *M* KCl, unbuffered) of < 12 $cmol_c$ $kg^{-1}$ clay. |
| Acrisols | Having a CEC (by 1 *M* $NH_4OAc$, pH 7) of < 24 $cmol_c$ $kg^{-1}$ clay in some part of the argic horizon. |
| Lixisols | Having a CEC (by 1 *M* $NH_4OAc$, pH 7) of < 24 $cmol_c$ $kg^{-1}$ clay in some part of the argic horizon. |
| *Qualifiers* | |
| Acric | Having a CEC (by 1 *M* $NH_4OAc$, pH 7) of < 24 cmolc $kg^{-1}$ clay |
| Alic | Having a CEC (by 1 *M* $NH_4OAc$, pH 7) of pH 7) of < 24 cmolc $kg^{-1}$ clay |
| Lixic | Having a CEC (by 1 *M* $NH_4OAc$, pH 7) of < 24 $cmol_c$ $kg^{-1}$ clay |
| Luvic | Having a CEC (by 1 *M* $NH_4OAc$, pH 7) of ≥ 24 $cmol_c$ $kg^{-1}$ clay |
| Hypersideralic | Having a CEC (by 1 *M* $NH_4OAc$, pH 7) of < 16 $cmol_c$ $kg^{-1}$ clay |
| Sideralic | Having a CEC (by 1 *M* $NH_4OAc$, pH 7) of < 24 $cmol_c$ $kg^{-1}$ clay |
| Ferralic | Having a CEC (by 1 *M* $NH_4OAc$, pH 7) of < 16 $cmol_c$ $kg^{-1}$ clay and a sum of exchangeable bases (by 1 *M* $NH_4OAc$, pH 7) plus exchangeable Al (by 1 *M* KCl, unbuffered) of < 12 $cmol_c$ $kg^{-1}$ clay. |
| Xanthic | Having a CEC (by 1 *M* $NH_4OAc$, pH 7) of < 16 $cmol_c$ $kg^{-1}$ clay and a sum of exchangeable bases (by 1 *M* $NH_4OAc$, pH 7) plus exchangeable Al (by 1 *M* KCl, unbuffered) of < 12 $cmol_c$ $kg^{-1}$ clay. |

**Table S3.** Diagnostic subsurface horizons and diagnostic characteristics in the Chinese Soil Taxonomy (CST) [5], in which the values of the $CEC_{clay}$ (by $NH_4OAc$ pH 7) and ECEC are required.

| Type | Soil orders | Soil groups or subgroups | Diagnostic criteria |
|---|---|---|---|
| *Diagnostic subsurface horizons* | | | |
| Ferralic horizon | Ferralosols | Rhodi-Udic Ferralosols, Xanthi-Udic Ferralosols, Hapli-Udic Ferralosols | Having a CEC (by 1 *M* $NH_4OAc$, pH 7) of < 16 $cmol_c\ kg^{-1}$ clay and an ECEC of < 12 $cmol_c\ kg^{-1}$ clay. |
| LAC-ferric horizon | Ferrosols | Argi-Ustic Ferrosols, Hapli-Ustic Ferrosols, Carbonati-Perudic Ferrosols, Alliti-Perudic Ferrosols, Hapli-Perudic Ferrosols, Carbonati-Udic Ferrosols, Hi-weatheri-Udic Ferrosols, Alliti-Udic Ferrosols, Argi-Udic Ferrosols, Hapli-Udic Ferrosols | Having a CEC (by 1 *M* $NH_4OAc$, pH 7) of < 24 $cmol_c\ kg^{-1}$ clay |
| *Diagnostic characteristics* | | | |
| Alic property | | Ali-Perudic Argosols, Ali-Udic Argosols, Ali-Perudic Cambosols, Ali-Udic Cambosols, Alic Humi-Udic Andosols, Alic Hapli-Udic Andosols, Alic Acidi-Udic Argosols, Alic Bori-Perudic Cambosols, Alic Acidi-Perudic Cambosols | Having a CEC (by 1 *M* $NH_4OAc$, pH 7) of ≥ 24 $cmol_c\ kg^{-1}$ clay |

**Table S4.** Correlation coefficients between soil properties and environmental variables.

| | $CEC_{silt}$ | $CEC_{soil}$ | $CEC_{OM}$ | pH | SOC | $Fe_d$ | Silt | Clay | Elevation | Slope | MRRTF | TWI | SPI | MAP | MAT |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $CEC_{clay}$ | 0.272** | 0.580** | 0.525** | 0.121 | -0.067 | -0.283** | 0.424** | -0.263** | 0.192* | 0.116 | -0.196* | -0.090 | 0.240** | -0.299** | -0.376** |
| $CEC_{silt}$ | 1 | 0.600** | 0.459** | 0.068 | 0.073 | 0.140 | 0.108 | -0.337** | 0.188* | 0.118 | -0.216* | -0.116 | 0.121 | -0.069 | -0.024 |
| $CEC_{soil}$ | | 1 | 0.920** | 0.279** | 0.048 | 0.258** | 0.160 | 0.065 | 0.356** | 0.086 | -0.194* | -0.129 | 0.147 | -0.365** | -0.310** |
| $CEC_{OM}$ | | | 1 | 0.238** | 0.004 | 0.298** | 0.085 | 0.237** | 0.335** | 0.046 | -0.137 | -0.089 | 0.104 | -0.378** | -0.342** |
| pH | | | | 1 | -0.001 | 0.048 | 0.225* | -0.067 | 0.231* | 0.356** | -0.078 | -0.272** | 0.427** | -0.397** | -0.048 |
| SOC | | | | | 1 | 0.188* | 0.032 | -0.171 | 0.020 | -0.018 | -0.014 | 0.039 | -0.018 | 0.022 | 0.228* |
| $Fe_d$ | | | | | | 1 | -0.130 | 0.541** | 0.383** | -0.079 | 0.012 | -0.025 | -0.120 | -0.245** | -0.077 |
| Silt | | | | | | | 1 | -0.352** | 0.120 | 0.145 | -0.026 | 0.038 | 0.101 | -0.239** | -0.213* |
| Clay | | | | | | | | 1 | 0.076 | -0.225* | 0.171 | 0.082 | -0.228* | -0.163 | -0.166 |

* Significant at the 0.05 level.

** Significant at the 0.01 level.

**Table S5.** Performance assessment ($R^2$ and RMSE) of multiple linear regression for the PTFs of CEC of silt ($CEC_{silt}$), CEC of the fine earth fraction (<2 mm) ($CEC_{soil}$) and the CEC of mineral fractions ($CEC_{Min}$). For each prediction case, the mean values and standard deviations of $R^2$ and RMSE based on 100 runs are shown.

| **Dataset** | **$CEC_{silt}$** | **$CEC_{soil}$** | **$CEC_{Min}$** |
|---|---|---|---|
| *$R^2$* | | | |
| Dataset 1 | 0.51±0.21 | 0.21±0.18 | 0.85±0.09 |
| Dataset 2 | 0.53±0.22 | 0.22±0.18 | 0.86±0.07 |
| Dataset 3 | 0.54±0.23 | 0.21±0.21 | 0.85±0.08 |
| Dataset 4 | 0.55±0.22 | 0.22±0.18 | 0.86±0.08 |
| *RMSE* | | | |
| Dataset 1 | 2.62±0.40 | 4.33±0.54 | 1.85±0.37 |
| Dataset 2 | 2.60±0.41 | 4.28±0.84 | 1.84±0.44 |
| Dataset 3 | 2.58±0.41 | 4.27±0.99 | 1.83±0.44 |
| Dataset 4 | 2.56±0.39 | 4.28±0.84 | 1.84±0.44 |

## SI 3 Supplementary Figures

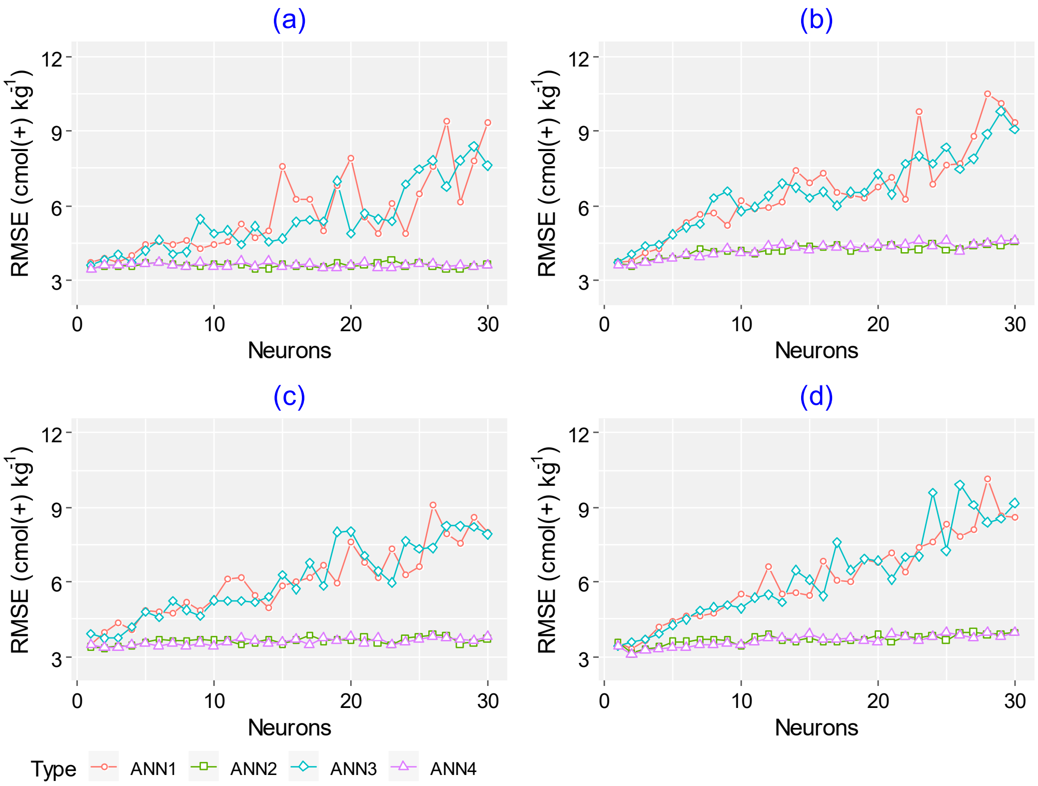


**Figure S1.** RMSE values of ANN models with 1 hidden layer for datasets 1 (a), 2 (a), 3 (c) and 4 (d), in which the number of neurons in the hidden layer ranged from 1 to 30. ANN1 and ANN2 refer to the resilient backpropagation algorithm with weight backtracking using hyperbolic tangent and logistic sigmoid activation functions, respectively. ANN3 and ANN4 refer to the resilient backpropagation algorithm without weight backtracking using hyperbolic tangent and logistic sigmoid activation functions, respectively.

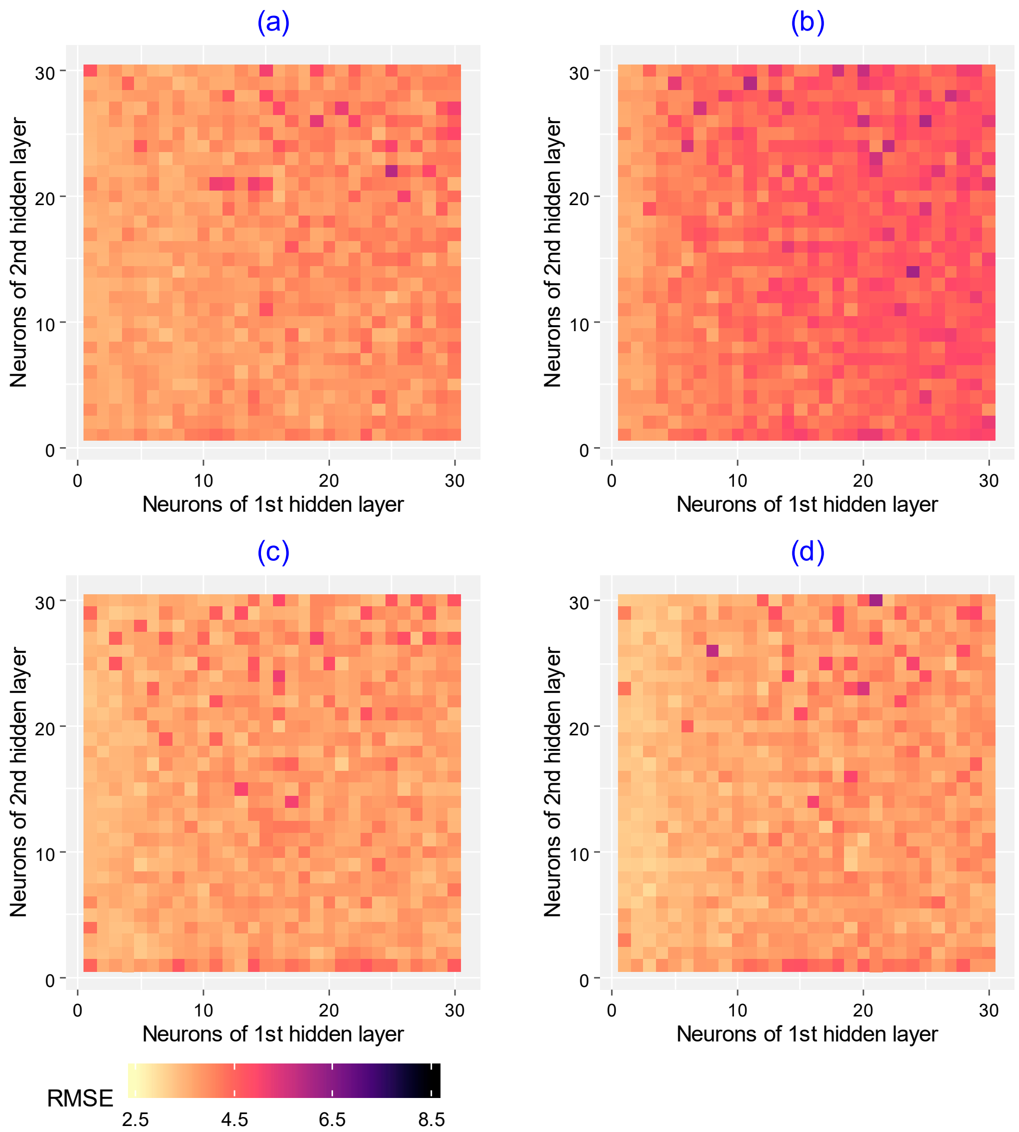


**Figure S2.** RMSE values of ANN models with 2 hidden layers for datasets 1 (a), 2 (a), 3 (c) and 4 (d), in which the number of neurons in the first and second hidden layer ranged from 1 to 30.

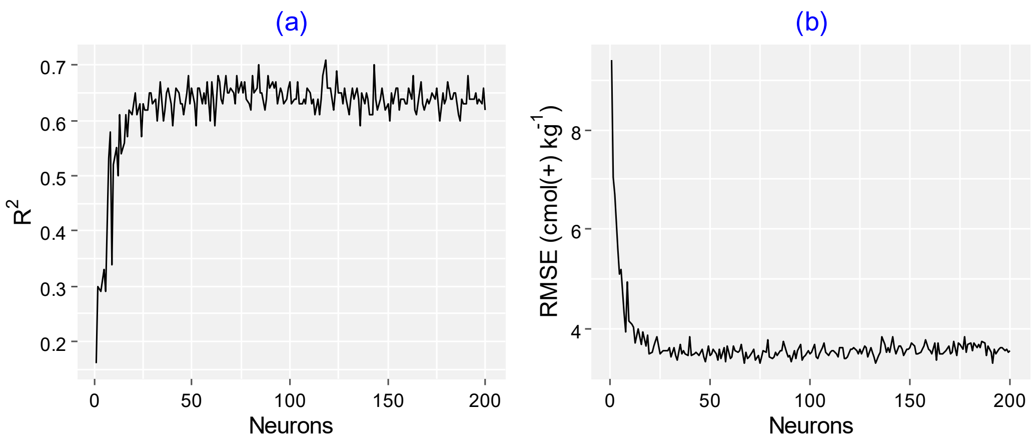


**Figure S3.** $R^2$ (a) and RMSE (b) values of DBN models with 1 hidden layer for dataset 1.

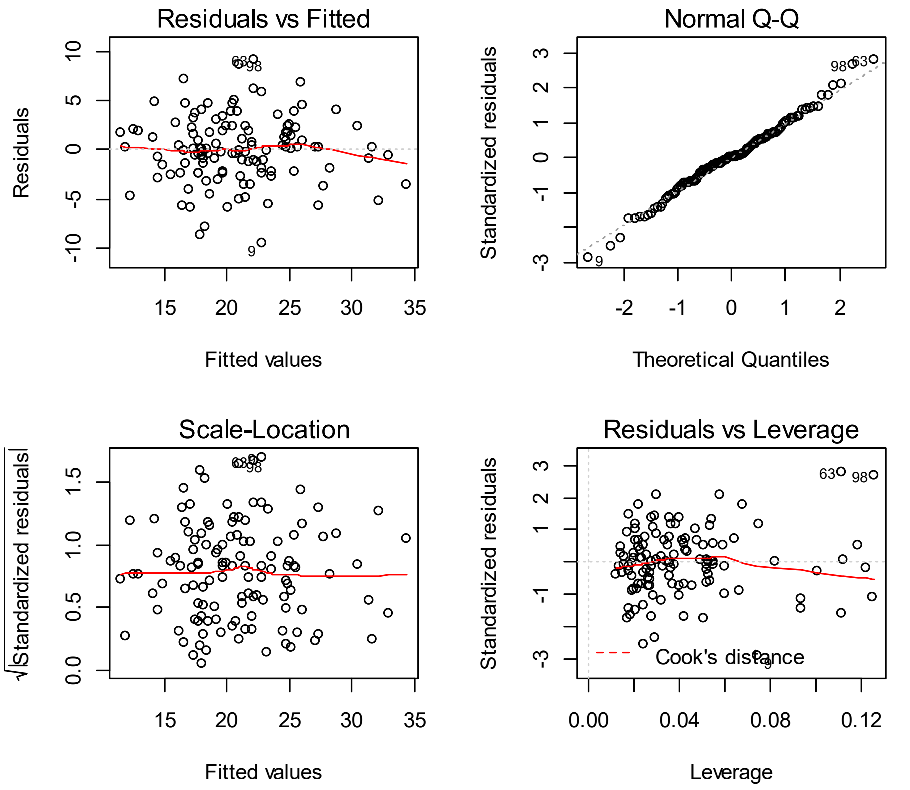


**Figure S4.** Residuals distribution check of dataset 1.

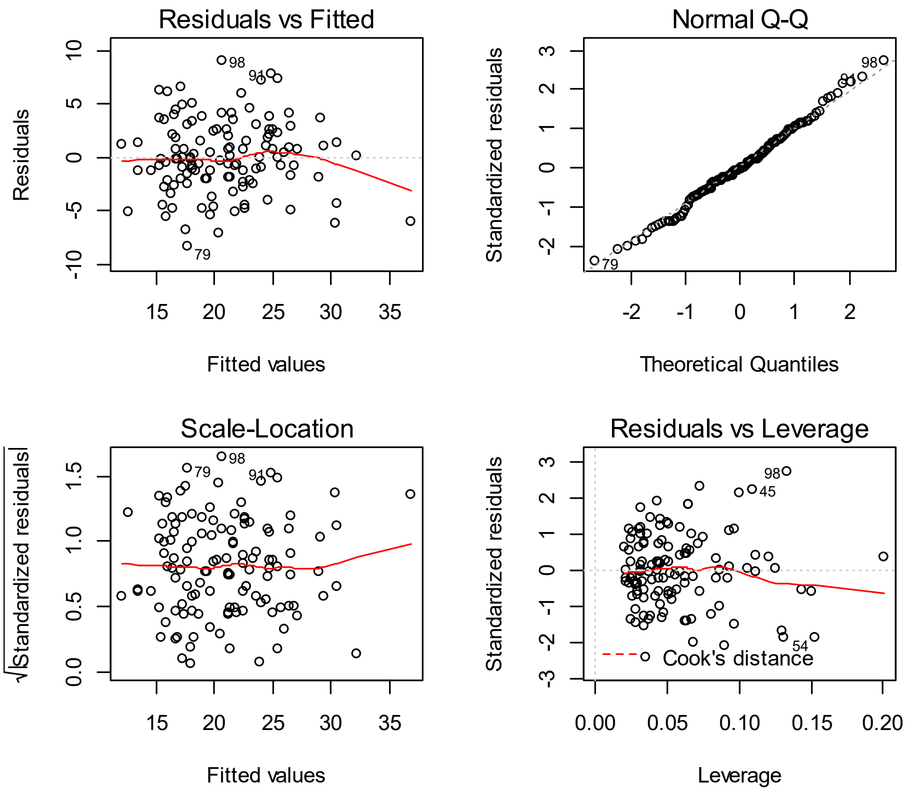


**Figure S5.** Residuals distribution check of dataset 2.

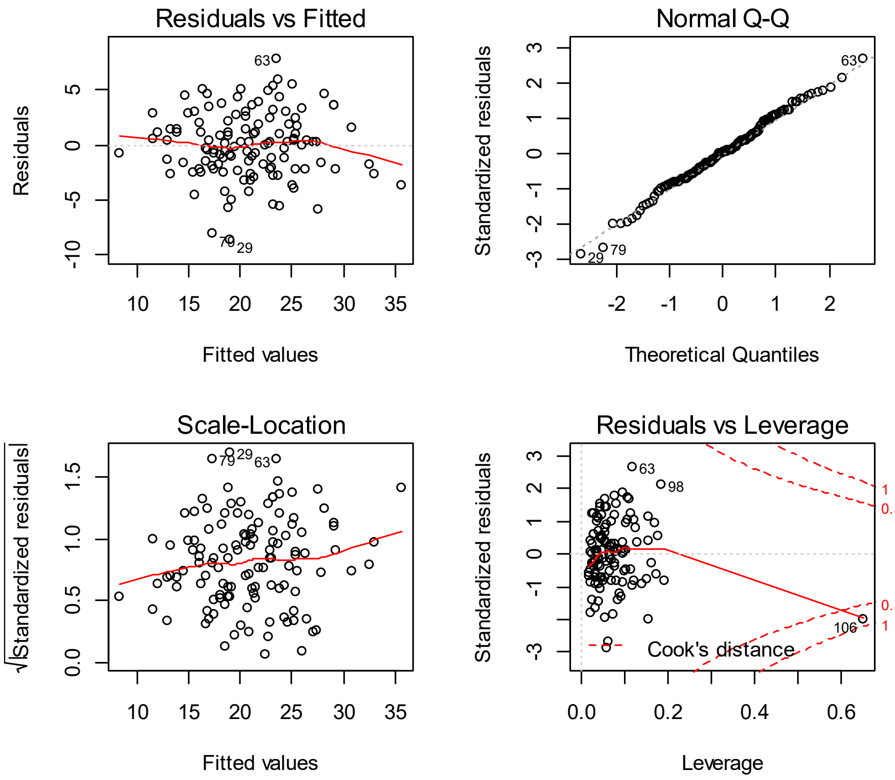


**Figure S6.** Residuals distribution check of dataset 3.

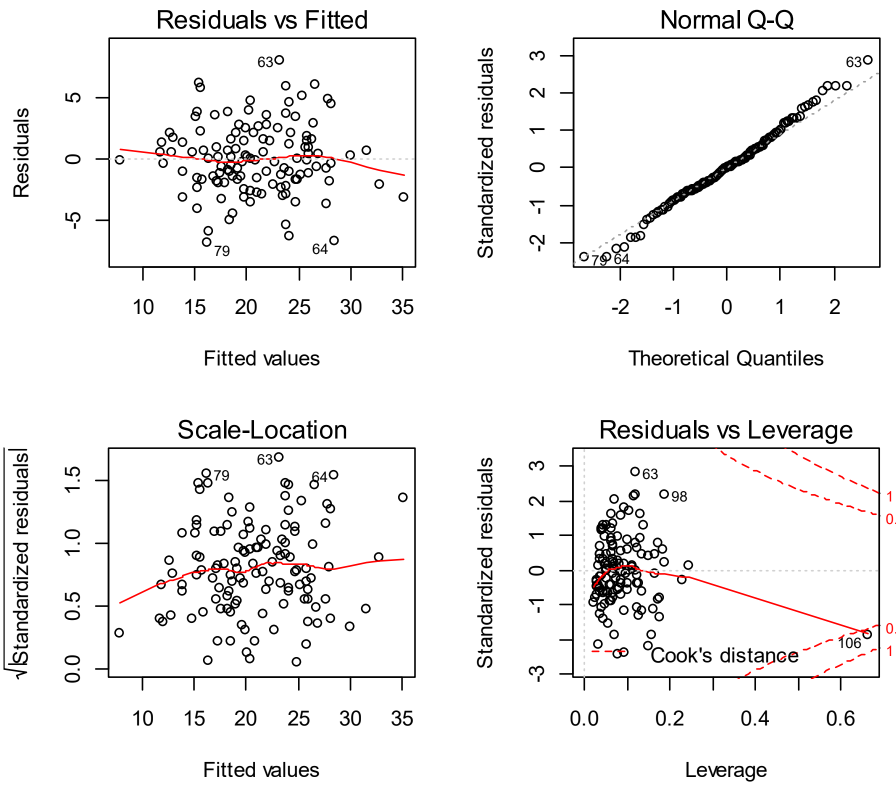


**Figure S7.** Residuals distribution check of dataset 4.

(a) (b) (c) (d)


**Figure S8.** Representative pedons of four soil series, of which the soil type should be referenced to Ferrosols: Liangtian (a), Jinji (b), Datuo (c), and Dengta (d). The land use type of Liangtian and Datuo is forest, and that of Jinji and Dengta is upland.