

# SCIENTIFIC DATA

## OPEN Data Descriptor: Global distribution of clay-size minerals on land surface for biogeochemical and climatological studies

Received: 9 February 2017

Accepted: 3 July 2017

Published: 22 August 2017

Akihiko Ito<sup>1,2</sup> & Rota Wagai<sup>3</sup>

Clay-size minerals play important roles in terrestrial biogeochemistry and atmospheric physics, but their data have been only partially compiled at global scale. We present a global dataset of clay-size minerals in the topsoil and subsoil at different spatial resolutions. The data of soil clay and its mineralogical composition were gathered through a literature survey and aggregated by soil orders of the Soil Taxonomy for each of the ten groups: gibbsite, kaolinite, illite/mica, smectite, vermiculite, chlorite, iron oxide, quartz, non-crystalline, and others. Using a global soil map, a global dataset of soil clay-size mineral distribution was developed at resolutions of 2' to 2° grid cells. The data uncertainty associated with data variability and assumption was evaluated using a Monte Carlo method, and validity of the clay-size mineral distribution obtained in this study was examined by comparing with other datasets. The global soil clay data offer spatially explicit studies on terrestrial biogeochemical cycles, dust emission to the atmosphere, and other interdisciplinary earth sciences.

<b>Design Type(s)</b>	observational design • data integration objective
<b>Measurement Type(s)</b>	composition of soil
<b>Technology Type(s)</b>	data collection method
<b>Factor Type(s)</b>	
<b>Sample Characteristic(s)</b>	Shiga Prefecture • Taiwan Island • Indian subcontinent • Fars Province • India • Fukushima Prefecture • Northern California • Argentine Patagonia • Iran • Morteratsch Glacier • Italy • Province of Ontario • Municipality of Kiruna • Ross Dependency • Brazil • Triangulo Mineiro/Alto Paranaiba Mesoregion • Municipality of Manaus

<sup>1</sup>National Institute for Environmental Studies, Tsukuba 305-8506, Japan. <sup>2</sup>Japan Agency for Marine-Earth Science and Technology, Yokohama 236-0001, Japan. <sup>3</sup>National Agriculture and Food Research Organization, Tsukuba 305-8604, Japan. Correspondence and requests for materials should be addressed to A.I. (email: itoh@nies.go.jp).

## Background & Summary

With an equivalent spherical diameter of  $< 2 \mu\text{m}$ , clay particles are produced through various geochemical processes such as weathering and metamorphism<sup>1</sup>. Clay exists in the Earth's crust, ocean sediments, and atmospheric aerosols. Clay is one of the fundamental constituents of soil; it influences soil physical and chemical attributes, soil biogeochemical functions, and the services it provides<sup>2,3</sup>. Because of their size and microstructure, clay-size fine particle fraction (hereafter clay-size minerals) account for major portions of soil specific surface area, allowing interaction with inorganic ions and organic compounds<sup>4</sup>. Clay lends the properties of viscosity and plasticity to wet soils and causes concretion as a result of baking and burning<sup>5</sup>. Interaction of clays with organic matter facilitated by soil microbial activity leads to the formation of characteristic pore-rich structure in soil called aggregates, which enhance aeration and water retention, and stabilize soil organic matter<sup>6,7</sup>. These characteristics are critical for supplying terrestrial plants with water and nutrients for growth and survival. Moreover, clay is a major source of airborne mineral dust, which is mainly contributed to the atmosphere from arid regions. Compared with silt ( $2\text{--}20 \mu\text{m}$  diameter) and sand ( $> 20 \mu\text{m}$  diameter), airborne clay dust has a longer residence time in the atmosphere and influences radiation scattering and cloud formation<sup>8</sup>.

Because of their unique properties and importance in many natural processes, clay-size minerals are frequently of interest in soil biogeochemical and climatological studies. Dust emission schemes implemented in aerosol-chemistry models for atmospheric research consider soil particle size distribution<sup>9,10</sup>. Soil sand and clay fractions are used, together with soil bulk density, in many hydrological and biogeochemical models to determine soil water-holding capacity and hydraulic conductivity<sup>11,12</sup>. Moreover, some biogeochemical models account for the influence of clay content on the turnover rate of soil organic matter<sup>13,14</sup>. For the purposes of broad-scale application, spatially explicit datasets of major soil properties including clay content have been constructed<sup>15–17</sup>. However, clay ( $< 2 \mu\text{m}$  sized particles) encompasses a variety of minerals that differ widely in physical structure and chemical composition and reactivity, and are distributed heterogeneously on earth surface, both horizontally and vertically. Therefore, utility may be found not only in the distribution of total clay content in soil but also in the mineral composition of the clay to account for patterns in soil function or behavior. Because of the technical difficulty in quantifying clay-size minerals<sup>18</sup>, soil datasets published to date usually exclude information on clay-size mineral composition, except for limited geographical regions (e.g., Australia<sup>19</sup>). Several dust emission datasets provide global maps of soil clay-size minerals<sup>20–22</sup> for dust sources such as the near-surface soils of dry regions. However, these dust source datasets are not useful for hydrological and biogeochemical studies in humid regions where subsoil properties may be required.

We therefore developed a dataset of major soil clay-size minerals covering the global land surface for both topsoil (near-surface) and subsoil at different spatial resolutions. This dataset is intended for application in a variety of earth science fields that require interdisciplinary data. We gathered observational data on clay-size minerals through a literature survey and meta-analysis. Most observations were originally obtained by X-ray diffraction (XRD) analysis. The range of clay-size minerals that occur in soils were classified into ten groups: gibbsite, kaolinite, illite/mica, smectite, vermiculite, chlorite, iron (Fe) oxide, quartz, non-crystalline (amorphous and short-range-order minerals), and others. We then aggregated the clay-size mineral composition data on the basis of 12 soil orders<sup>23</sup>. Using a global map of soil orders and additional soil datasets, we developed global maps of clay-size mineral abundances in topsoil and subsoil at a resolution of  $2'$  grid cells (about  $3.7 \text{ km}$ ) and, by averaging, at lower spatial resolutions. We examined uncertainties in the dataset by statistical (i.e., Monte Carlo) methods and by comparison with previous datasets.

While the new dataset developed here has room for improvement, it can facilitate continental-scale studies of biogeochemistry and climatology by providing more precise soil properties related to, for example, soil cation exchange capacity and dust emission. The dataset should also find applications in interdisciplinary studies in fields such as hydrology and agronomy, both as input data for model simulations and for the interpretation of observational data.

## Methods

### Meta-analysis

Records of the clay-size mineral composition of terrestrial soils were mainly collected by literature survey using the search engines Web of Science (Thomson Reuters, New York, NY) and Google Scholar (Alphabet Inc., Mountain View, CA). We attempted to gather as much data as possible for all the orders of Soil Taxonomy<sup>23</sup>, a soil classification system that is both practical and genetic as it considers all soil formation factors (i.e., climate, parent material, topography, biota, and time). It also captures the variation in the degree of weathering from immature Entisols and young Inceptisols to highly weathered Ultisols and Oxisols. This soil classification system has been commonly used in soil science and is compatible with the World Reference Base<sup>24</sup>, an international standardized soil classification system. Histosol order was excluded from the present study because it is distributed mainly in peat lands and is primarily composed of organic materials with little mineral matter. For the same reason, data from organic layers (O or  $A_0$  horizon) above mineral soil layers were not considered. We selected records from papers that contained all the following information for a soil sample: the sampling location, the soil order at the site (or a name from another soil classification system that could be translated to a Soil Taxonomy order), the depth or soil layer sampled, and quantitative data on clay-size mineral composition

(e.g., percent of the clay fraction of each constituent, referred as % clay hereafter). Papers that only qualitatively described which clay-size mineral was predominant in the soil sample were not used. In the literature we collected, the majority of measurements of clay-size mineral composition used the XRD analysis method<sup>25,26</sup>. Several studies also adopted other analytical methods to identify and quantify clay-size minerals, such as scanning or transmission electron microscopy, infrared spectroscopy, differential thermal analysis, and thermogravimetric analysis<sup>18</sup> as well as selective dissolution<sup>27,28</sup> for the non-crystalline category. When available, other information on the soil samples was extracted for future analyses. These comprised bulk density; sand, silt, and clay composition; cation exchange capacity; specific surface area; and organic carbon content. At present, we gathered 168 records from 27 papers (Data Citation 1), which satisfied the conditions mentioned above, representing all the soil orders and different continental regions.

In the collected literature, clay-size fraction (not only clay minerals but also other substances) was classified using various methods and nomenclatures. To maximize the applicability of the data to biogeochemical studies, we aggregated clay-size minerals into ten groups (Table 1): gibbsite, kaolinite, illite/mica, smectite, vermiculite, chlorite, iron oxide, quartz, non-crystalline, and others. Illite and mica were grouped together, because illite may form from the weathering of potassium-rich mica (muscovite), the most common mineral in mica family. Among 2:1 type layer silicate clay minerals, vermiculite and smectite were separated from illite due to their expandability of their interlayers that lead to much higher chemical reactivity. Smectite-group minerals (e.g., montmorillonite) is particularly expandable and thus separated from vermiculite. Hydroxyl-interlayered-vermiculite (HIV) and hydroxyl-interlayered smectite (SIV), often present in acidic soils, are included under vermiculite category. Iron oxide, hydroxide, and oxyhydroxide are collectively called Fe oxides here. Fe oxide, mainly present as hematite and goethite, is separately considered from other clay-size minerals because this is an important clay-sized mineral group especially in weathered tropical soils and those derived from mafic materials and plays a unique role in plant nutrition and soil microbial metabolism<sup>29</sup>. The non-crystalline category, estimated from selective dissolution data, contained several different amorphous and short-range order minerals including volcanic glass in Andosols and acid oxalate soluble phases of Al and Fe oxides. For each record, total clay-size mineral fraction of 10 groups in Table 1 was normalized to 100%. The others category was adopted as a catch-all for clay-size constituents with a low frequency of occurrence, including feldspar, and mixed-layer clay-size minerals. The classification is not intended to remain fixed and could be revised in response to improved understanding or to meet the requirements of different applications.

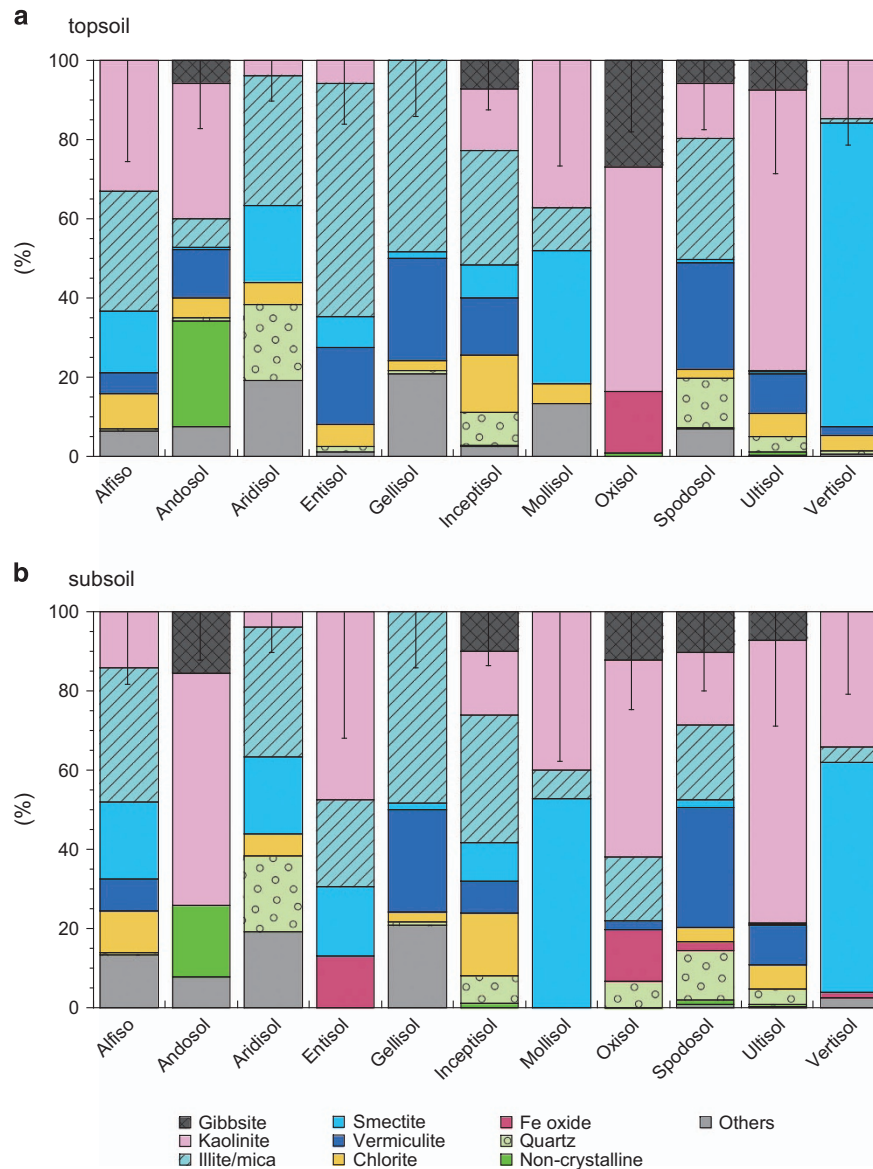
In this study, we adopted a global definition of topsoil as the soil layer from the ground surface to 0.3 m depth when the sampling depth was given, or as soil in the A horizon when the sampling depth was described in terms of soil profile layers. Similarly, subsoil was defined as soil below 0.3 m depth or the B horizon. Soils obtained from below 2 m depth or the C horizon were not recorded, because our focus was primarily on the biogeochemically active soil components. To avoid overrepresentation of a few intensive studies, multiple samples from a single site were averaged by site and soil layer. We then calculated statistical metrics (e.g., mean, median, and standard deviation) of the clay-size mineral fractions for the topsoil and subsoil of each Soil Taxonomy order. The mean clay-size mineral composition for the 11 soil orders obtained from the meta-analysis clearly indicated that clay-size mineral composition differs widely among soil orders (Fig. 1).

## Mapping

Several methods are available to convert georeferenced point data into a continuous domain, such as linear interpolation and geostatistical techniques like kriging. In light of the discontinuous nature of clay composition and the sparsity of the available data, we adopted a simple and robust assignment method on the basis of the soil-type map. We used a global map of Soil Taxonomy orders and suborders

Groups	Major clay-size minerals	Structure*
Gibbsite	gibbsite	1 (octahedral)
Kaolinite	kaolinite, halloysite	1:1
Illite/mica	illite, mica	2:1
Smectite	montmorillonite	2:1
Vermiculite	vermiculite	2:1
Chlorite	chlorite	2:1:1
Iron (Fe) oxide	goethite, hematite	mixed
Quartz	quartz	crystal
Non-crystalline	allophane, imogolite, ferrihydrite	amorphous
Others	feldspar, etc.	mixed

**Table 1. Summary of clay-size mineral groups.** \*Silicate layer or other structural characteristics of clay-size minerals. Numbers indicate the numbers of tetrahedral: octahedral (: inter-layer) sheets.



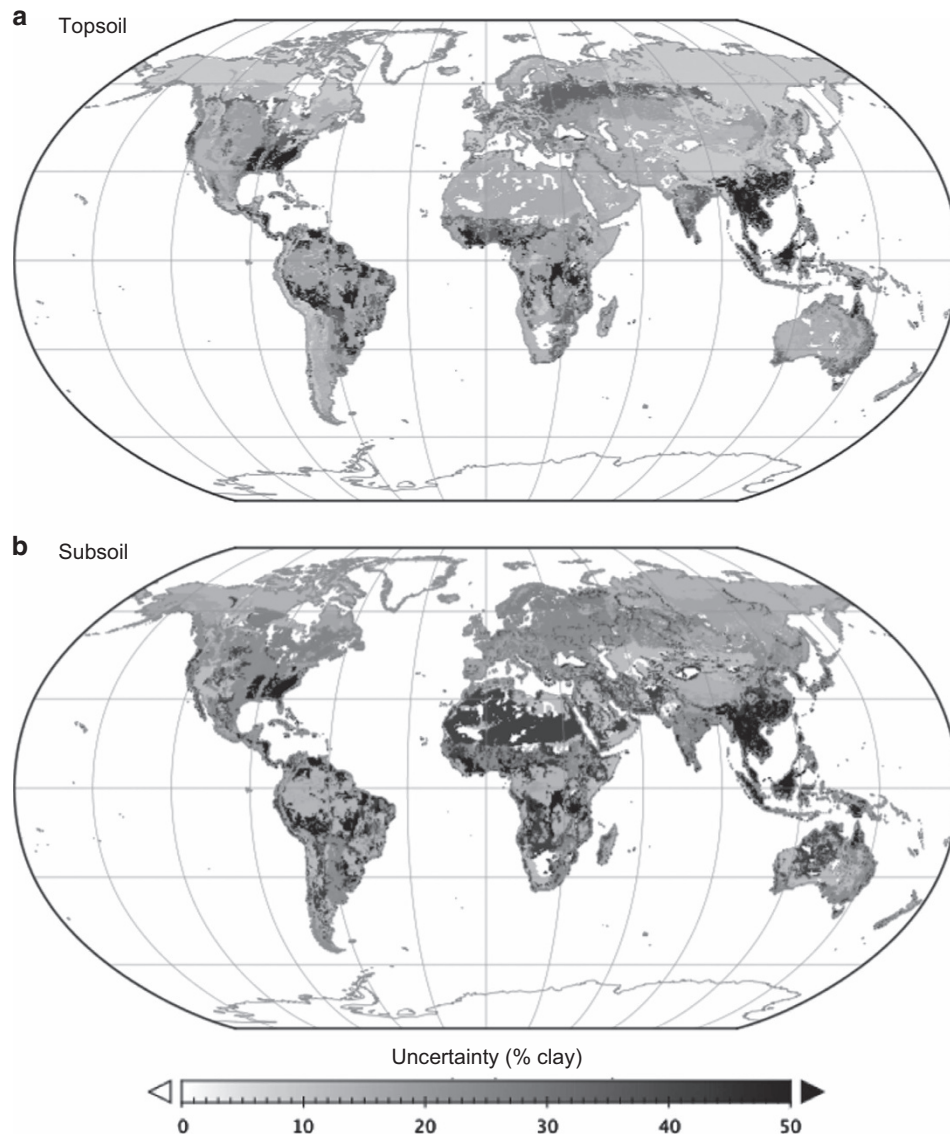
**Figure 1.** Mean clay-size mineral composition for each of the soil orders of Soil Taxonomy. The data were derived from the literature, and shown for the (a) topsoil and (b) subsoil, separately. The error bar at the top of each column shows the range of the fraction-weighted standard deviation obtained from data in the literature.

produced by the Natural Resources Conservation Service of the United States Department of Agriculture (data obtained from: [https://www.nrcs.usda.gov/Internet/FSE\\_DOCUMENTS/nrcs142p2\\_052837.zip](https://www.nrcs.usda.gov/Internet/FSE_DOCUMENTS/nrcs142p2_052837.zip), last accessed: July 7, 2017) produced by reclassifying the base maps of the FAO/UNESCO Soil Map of the World. The spatial resolution of the soil map is a grid defined by cells of 2' latitude and longitude (about 3.7 km on the equator). The average clay composition (Fig. 1) was assigned to each grid cell on the basis of the dominant soil order for the cell (areas of glaciers and ice sheets, Histosol in wetlands, and shifting desert sands were excluded). To enable the data to be used as inputs for global modeling at spatial resolutions coarser than 2', we also prepared versions of the maps with larger grids (Data Citation 1). The multiple 2' values contained in a larger grid cell would be averaged to produce a value for the larger cells. We produced versions with cell sizes of 5', 15', 30', 1°, and 2°.

The clay composition data mentioned above indicate  $F_{CM}$  (% clay), which is the proportion of the soil's clay occupied by a given clay-size mineral group, for each of the ten groups. To obtain the absolute amount of a clay-size mineral in the soil column (i.e., weight of mineral per unit area,  $W_{CM}$ ,  $\text{kg m}^{-2}$ ), we used the following equation:

$$W_{CM} = ST \times BD \times (1 - F_{GV}/100) \times F_{CL}/100 \times F_{CM}/100$$

where  $ST$  is soil depth or layer thickness (m),  $BD$  is soil bulk density ( $\text{kg m}^{-3}$ ),  $F_{GV}$  is the gravel fraction (%), and  $F_{CL}$  is the soil clay content (%). These additional soil data ( $BD$ ,  $F_{GV}$ , and  $F_{CL}$ ) are available from



**Figure 2.** Distribution of combined weighted uncertainty in the estimates of clay-size mineral composition. Data are shown for (a) topsoil and (b) subsoil.

global datasets such as the Harmonized World Soil Database (HWSD), version 1.21, developed by the FAO and International Institute of Applied System Analysis<sup>16</sup>. The HWSD was compiled from different source soil maps and pedon data to maximize the usability of existing soil data.

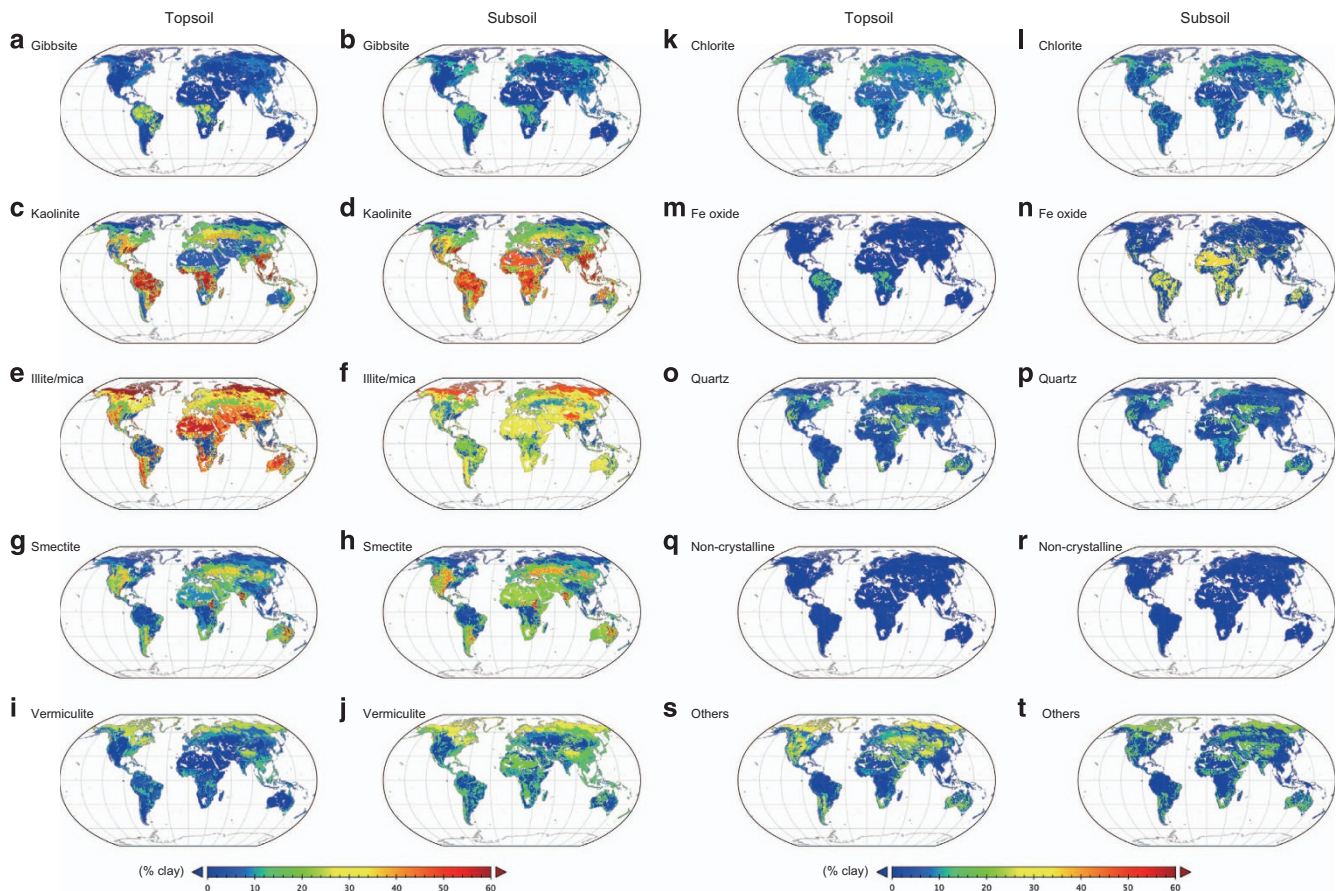
#### Code availability

The global map datasets were produced using a purpose-built program written in C with the standard mathematical library. The code has been compiled for a variety of platforms (e.g., Windows, UNIX, and MacOS), requires 1.4 to 11 GB memory to run dependent on platform and spatial resolution, and is available from the corresponding author upon request.

#### Data Records

The meta-analysis dataset (clay\_meta\_data\_v1.csv; Data Citation 1) includes the following fields for each record:

- Record ID
- Soil order name
- Topsoil (T) or subsoil (S)
- $F_{CM}$  for gibbsite group (% clay)
- $F_{CM}$  for kaolinite group (% clay)
- $F_{CM}$  for illite/mica group (% clay)
- $F_{CM}$  for smectite group (% clay)



**Figure 3.** Global terrestrial distribution of clay-size mineral groups as a percent of the soil clay fraction (% clay), excluding areas of ice sheet, glaciers, shifting desert sands and dunes, and wetlands (histosols). (a, b) Gibbsite, (c, d) kaolinite, (e, f) illite/mica, (g, h) smectite, (i, j) vermiculite, (k, l) chlorite, (m, n) Fe oxide, (o, p) quartz, (q, r) non-crystalline, and (s, t) others. Data for topsoil (a, c, e, g, i, k, m, o, q, s) and subsoil (b, d, f, h, j, l, n, p, r, t) are shown.

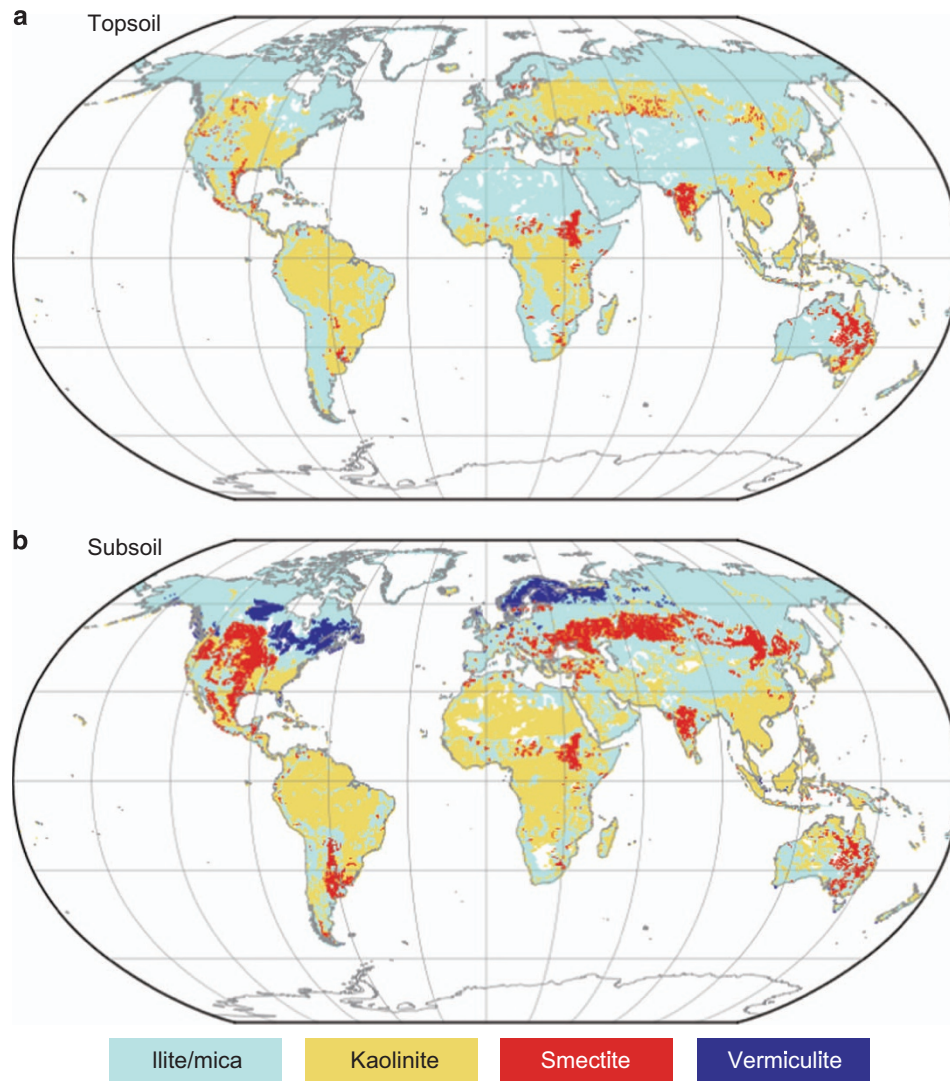
$F_{CM}$  for vermiculite group (% clay)  
 $F_{CM}$  for chlorite group (% clay)  
 $F_{CM}$  for iron oxide group (% clay)  
 $F_{CM}$  for quartz group (% clay)  
 $F_{CM}$  for non-crystalline group (% clay)  
 $F_{CM}$  for others group (% clay).

#### References

The global map data file includes the following variables for each grid cell of the topsoil and subsoil files, respectively:

$F_{CM}$  and  $W_{CM}$  for gibbsite group (% clay or  $\text{kg m}^{-2}$ )  
 $F_{CM}$  and  $W_{CM}$  for kaolinite group (% clay or  $\text{kg m}^{-2}$ )  
 $F_{CM}$  and  $W_{CM}$  for illite/mica group (% clay or  $\text{kg m}^{-2}$ )  
 $F_{CM}$  and  $W_{CM}$  for smectite group (% clay or  $\text{kg m}^{-2}$ )  
 $F_{CM}$  and  $W_{CM}$  for vermiculite group (% clay or  $\text{kg m}^{-2}$ )  
 $F_{CM}$  and  $W_{CM}$  for chlorite group (% clay or  $\text{kg m}^{-2}$ )  
 $F_{CM}$  and  $W_{CM}$  for quartz group (% clay or  $\text{kg m}^{-2}$ )  
 $F_{CM}$  and  $W_{CM}$  for iron oxide group (% clay or  $\text{kg m}^{-2}$ )  
 $F_{CM}$  and  $W_{CM}$  for non-crystalline group fraction (% clay or  $\text{kg m}^{-2}$ )  
 $F_{CM}$  and  $W_{CM}$  for others group (% clay or  $\text{kg m}^{-2}$ )  
 Dominant (i.e., most abundant) clay-size mineral group  
 Uncertainty due to sample variability (%)

The files (Data Citation 1) are in NetCDF-4 format, library version 4.3.1.1 (<http://www.unidata.ucar.edu/software/netcdf/>), and available as clay\_XXXX\_YY\_v1r1.nc4, where XXXX indicates fraction or weight data and YY denotes spatial resolution. Files with different mesh sizes are provided: 2' (YY = 2 m),



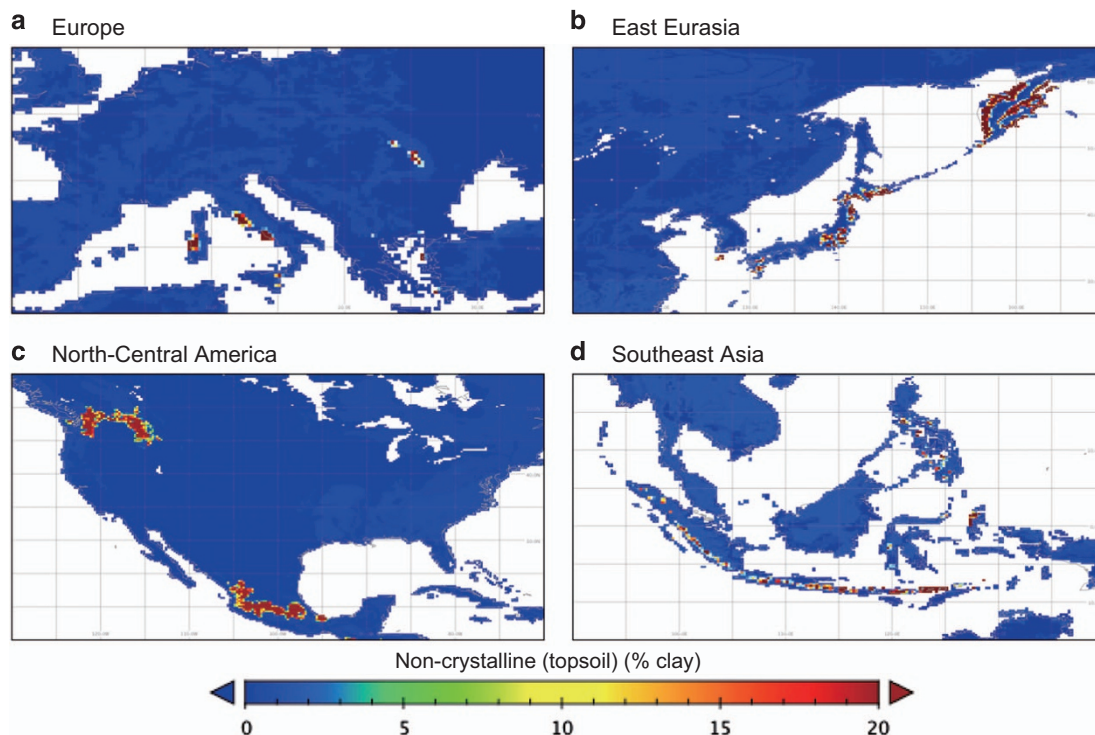
**Figure 4.** Distribution of the most abundant clay-size mineral group in each grid cell. Data is shown for (a) topsoil and (b) subsoil. The gibbsite, chlorite, quartz, and non-crystalline groups are not represented because they were never the most abundant.

5' (YY = 5 m), 15' (YY = qd), 30' (YY = hd), 1° (YY = 1d), and 2° (YY = 2d); note that 2' data were divided into three files due to their large size.

### Technical Validation

#### Estimation uncertainty

We evaluated the uncertainty of the calculated clay-size mineral composition values using the Monte Carlo method. The standard deviation values of clay-size mineral composition obtained by the meta-analysis were used to define ranges for random error generation, assuming a Gaussian distribution, to add to the average composition values (Fig. 1). Uncertainty in the clay-size mineral abundance determinations in the source data (e.g., several data were given as abundance level from + to +++) was incorporated by assigning different values to each category. Namely, we assigned 1–10% to the low abundance level of '+' and 20–60% to the high abundance level of '++++'; intermediate values were assigned to '++' and '+++'. After assigning the data uncertainty and random error, a total 30,000 sets of global maps were generated and normalized to total 100% as mentioned previously. Note that multiple initial values were used when generating random numbers. Based on the 30,000 combinations of clay-size mineral fractions at each grid, standard deviations were calculated for each clay-size mineral and weighted by the mean clay composition of that mineral before mapping the distribution of uncertainty (Fig. 2).



**Figure 5.** Regional distribution of non-crystalline (principally allophane) clay-size fraction group. Data are shown at 15' resolution for (a) Europe, (b) East Eurasia, (c) North-Central America, and (d) Southeast Asia.

#### Comparison with other datasets

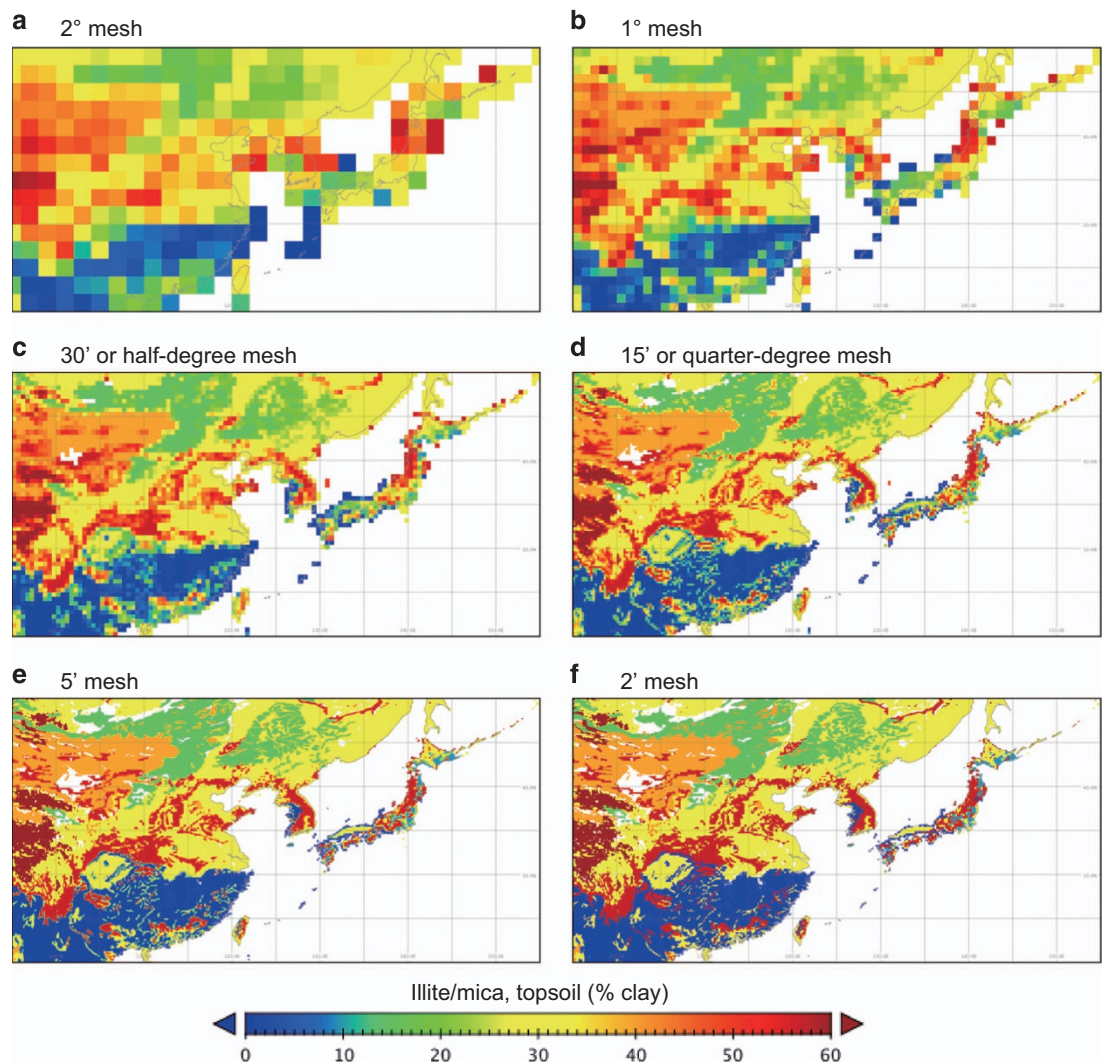
Several clay-size mineral maps have been developed for estimating dust emission<sup>20–22</sup>. The datasets of Clanquin *et al.*<sup>20</sup> and GMINER30 by Nickovic *et al.*<sup>21</sup> cover only arid regions such as deserts and dunes, because little dust emission occurs from humid vegetated areas. The present global dataset is consistent with these datasets with respect to major features of clay-size mineral distribution, such as the abundance of illite/mica in arid regions in West Asia and North Africa. The dataset of Clanquin *et al.*<sup>20</sup> and Nickovic *et al.*<sup>21</sup> show high quartz content in arid regions in Africa, Asia, and Australia because these datasets also include data for silt. The dataset of Journet *et al.*<sup>22</sup> covers both arid and humid land areas, except glaciers and wetlands, but only for topsoil subject to aeolian deflation. For dust emission estimation, the dataset separates clay-size fraction into groups for calcite, goethite, feldspar, and hematite, which were aggregated in our dataset. Again, our dataset is consistent with that of Journet *et al.*<sup>22</sup>, in that illite/mica content is high (40–50%) in arid regions and low (<10%) in the humid tropics. The three antecedent datasets focused on dust emission to the atmosphere, ignoring clay functions in terrestrial ecosystems. For example, these datasets do not consider non-crystalline clay, which exerts strong control on soil carbon storage and nutrient cycling<sup>30</sup>. The dataset we have produced provides clay-size mineral composition for both topsoil and subsoil, and for both arid and humid land areas, so unlike the pre-existing datasets it can be used for integrated studies of biogeochemistry and climatology.

#### Assessment of the global patterns of clay-size minerals

The global distributions of individual clay-size mineral groups (Fig. 3) and dominant (most abundant) clay-size mineral groups in the topsoil and subsoil (Fig. 4) can be compared with known patterns of distribution. The kaolinite group with 1:1 type layer structure is predominant in the humid tropics, and the illite/mica group is abundant in arid and high-latitude regions. Such distributions are consistent with pedological knowledge about the relationship between weathering intensity and clay-size mineral formation<sup>2,31</sup>. Independent evidence supports the validity of the present dataset, at least for continental-scale patterns. The global pattern of dominant clay-size minerals in ocean sediments<sup>32</sup> reflects riverine transportation from the land and is consistent with the terrestrial distributions revealed in our dataset (Fig. 4).

Several clay-size minerals are known to have localized distributions, and the dataset appropriately captures their characteristics. For example, Vertisols contain a high proportion of smectite in the topsoil, and both this soil order and the distribution of topsoil smectite in our dataset occur mainly in eastern Australia, central India, and east-central Africa. Allophane and imogolite, major constituents in the non-crystalline group, are formed mainly by the weathering of volcanic parent materials (e.g., tephra), so the non-crystalline clay-size minerals group shows a highly localized distribution in areas associated with





**Figure 6.** Spatial pattern of clay-size mineral distribution shown at different spatial resolutions. Data of illite/mica clay-size mineral group, as percent of the clay fraction (% clay) of topsoil in East Asia, are shown at spatial resolutions of (a) 2°, (b) 1°, (c) 30', (d) 15', (e) 5', and (f) 2'.

geologically recent volcanic activity in East Asia, Southeast Asia, northwestern North America, Central to South America, and southern Europe (Fig. 5). The distributions of these soil clay-size minerals provided by the dataset would be effective in characterizing associated soil functions at each grid cell.

The map data can be displayed at different spatial resolutions (Fig. 6). Apparently, the base map with 2' grid captures spatial heterogeneity in the greatest detail. The low-resolution maps lose fine-scale heterogeneity, but capture broad variations such as the change in clay-size mineral composition with climate gradients and parent materials. The lower resolution data were produced as a secondary exercise to support global modeling studies.

#### Ways to improve dataset

While the quick-look comparison and assessment above showed reasonable results, there are multiple ways to improve the dataset. Our approach was based on (i) the global soil distribution map at soil 'order' level in Soil Taxonomy system, and (ii) the assumption of relatively small mineralogical variation within each soil order.

One could further take advantage of the wealth of information present in the soil classification systems used internationally such as World Reference Base for Soil Resources (WRB) and Soil Taxonomy. The WRB system classifies global soils into 32 'reference soil groups' instead of 12 soil 'orders' in Soil Taxonomy. While the availability is limited to some geographic regions, 'soil family' level information in Soil Taxonomy is directly tied to dominant soil mineral types (e.g., kaolinitic, smectitic, gibbsitic) and particle-size distribution (e.g., clayey, loamy, sandy).

Fe oxide and non-crystalline groups are important mineral phases not adequately represented in the current dataset because the techniques to assess silicate clays (e.g., XRD) are not the most suitable method

to characterize Fe oxides and other non-crystalline phases such as allophane. However, the quantitative estimation of these mineral groups is possible by selective dissolution approaches<sup>27,28</sup>. In fact, these are standard techniques in soil science and thus a large literature is present (often without XRD data), which can be incorporated into the dataset.

The assumption of small within-soil-order variation needs to be critically examined to improve the quality of the dataset. Some soil orders are characterized by specific set of mineral types due to the strong influence of unique parent material, climate regime, and weathering intensity (e.g., Andisols, Vertisols, and Oxisols). On the other hand, less developed soil orders such as Inceptisols and Entisols are likely to show much higher variation in clay-size mineral composition depending on soil formation factors (especially parent material, topography, climate). Furthermore, human activity such as deforestation, land-use change, and tillage could significantly degrade soils and change mineralogical composition of clay-size fraction within each soil order, for instance, by erosion-induced displacement<sup>3</sup>.

The present dataset is, therefore, a first step towards a complete global distribution map for clay-size minerals and relevant soil properties. Nevertheless, the current dataset has some values to facilitate ongoing biogeochemical studies and we can make improvements as discussed above.

## Usage Notes

The meta-analysis dataset can be used to determine clay-size mineral composition for the soil orders of Soil Taxonomy and corresponding soil categories from other classifications. The map data can be used to specify the regional distribution of characteristic clay-size constituents (such as smectite and non-crystalline mineral including allophane) affecting phenomena of interest in the fields of geomorphology and plant productivity. For global studies, the dataset can be used to characterize soil properties such as water-holding capacity, cation exchange capacity, and dust emissivity for studies in the fields of biogeochemistry and climatology. The dataset would be useful for planning soil management in land systems or regions subject to human pressure<sup>33,34</sup>. For example, the current dataset could be used to improve our estimation of soils capacity to form stable aggregate against erosion<sup>35</sup> and to store C as soil organic matter<sup>36,37</sup>. For example, a soil stability index could be developed on the basis of the clay mineralogy and relevant properties data.

## References

- Bergaya, F. & Lagaly, G. General introduction: Clays, clay minerals, and clay science. *Dev. Clay Sci.* **5A**, 1–19 (2013).
- Wilson, M. J. The origin and formation of clay minerals in soils: past, present and future perspectives. *Clay Minerals* **34**, 7–25 (1999).
- Amundson, R. *et al.* Soil and human security in the 21st century. *Science* **348**, 647 (2015).
- Brigatti, M. F., Galán, E. & Theng, B. K. G. Structure and mineralogy of clay minerals. *Dev. Clay Sci.* **5A**, 21–81 (2013).
- Murray, H. H. Applied clay mineralogy today and tomorrow. *Clay Minerals* **34**, 39–49 (1999).
- Torn, M. S., Trumbore, S. E., Chadwick, O. A., Vitousek, P. M. & Hendricks, D. M. Mineral control of soil organic carbon storage and turnover. *Nature* **389**, 170–173 (1997).
- Six, J., Conant, R. T., Paul, E. A. & Paustian, K. Stabilization mechanisms of soil organic matter: Implications for C-saturation of soils. *Plant Soil* **241**, 155–176 (2002).
- Tegen, I., Laci, A. A. & Fung, I. The influence on climate forcing of mineral aerosols from disturbed soils. *Nature* **380**, 419–422 (1996).
- Laurent, B., Marticorena, B., Bergametti, G., Léon, J. F. & Mahowald, N. M. Modeling mineral dust emissions from the Sahara desert using new surface properties and soil database. *J. Geophys. Res.* **113**, D14218 (2008).
- Kok, J. F. Does the size distribution of mineral dust aerosols depend on the wind speed at emission? *Atm. Chem. Phys.* **11**, 10149–10156 (2011).
- Clapp, R. B. & Hornberger, G. M. Empirical equations for some soil hydraulic properties. *Water Res. Res.* **14**, 601–604 (1978).
- Saxton, K. E., Rawls, W. J., Romberger, J. S. & Papendick, R. I. Estimating generalized soil-water characteristics from texture. *Soil Sci. Soc. Am. J.* **50**, 1031–1036 (1986).
- Jenkinson, D. S. The turnover of organic carbon and nitrogen in soil. *Phil. Trans. Roy. Soc. L* **B329**, 361–368 (1990).
- Schimel, D. S. *et al.* Climatic, edaphic, and biotic controls over storage and turnover of carbon in soils. *Global Biogeochem. Cycles* **8**, 279–293 (1994).
- Zobler, L. *A world soil file for global climate modeling*, Report No. 87802 (NASA Goddard Institute for Space Studies, 1986).
- FAO/IIASA/ISRIC/ISSCAS/JRC. *Harmonized World Soil Database (version 1.2)* (FAO and IIASA, 2012).
- Batjes, N. H. Harmonized soil property values for broad-scale modeling (WISE30sec) with estimates of global soil carbon stock. *Geoderma* **269**, 61–68 (2016).
- Środoń, J. Identification and quantitative analysis of clay minerals. *Dev. Clay Sci.* **5B**, 25–49 (2013).
- Rossel, R. A. V. Fine-resolution multiscale mapping of clay minerals in Australian soils measured with near infrared spectra. *J. Geophys. Res.* **116**, F04023 (2011).
- Clauquin, T. & Schulz, M. Modeling the mineralogy of atmospheric dust sources. *J. Geophys. Res.* **104**, 22243–22256 (1999).
- Nickovic, S., Vukovic, A., Vujadinovic, M., Djurdjevic, V. & Pejanovic, G. technical note: High-resolution mineralogical database of dust-productive soils for atmospheric dust modeling. *Atm. Chem. Phys.* **12**, 845–855 (2012).
- Journet, E., Balkanski, Y. & Harrison, S. P. A new data set of soil mineralogy for dust-cycle modeling. *Atm. Chem. Phys.* **14**, 3801–3816 (2014).
- Soil Survey Staff. *Keys to Soil Taxonomy*. 8th edn (Pocahontas Press, 1999).
- FAO/ISSS/ISRIC. *World Reference Base for Soil Resources*, Report No. 84, 108 (Food and Agriculture Organization, International Society of Soil Science International Soil Reference and Information Centre, 1998).
- Moore, D. M. & Reynolds, R. C. *X-ray Diffraction and the Identification and Analysis of Clay Minerals* (Oxford University Press, 1997).
- Kahle, M., Kleber, M. & Jahn, R. Review of XRD-based quantitative analyses of clay minerals in soils: the suitability of mineral intensity factors. *Geoderma* **109**, 191–205 (2002).
- Parfitt, R. L. & Childs, C. W. Estimation of forms of Fe and Al: A review, and analysis of contrasting soils by dissolution and Moessbauer methods. *Austral. J. Soil Res.* **26**, 121–144 (1988).

28. Loeppert, R. H., Inskeep, W. P. in *Methods of Soil Analysis. Part 3, Chemical Methods Vol. 5 Soil Science Society of America Book Series* (ed. Sparks, D. L.) 632–650 (ASA and SSSA, 1996).
29. Schwertmann, U. in *Iron nutrition and interactions in plants*, (eds Chen, Y. & Hader, Y.) 3–27 (Kluwer Academic Publishers, 1991).
30. Sollins, P., Robertson, G. P. & Uehara, G. Nutrient mobility in variable- and permanent-charge soils. *Biogeochemistry* **6**, 181–199 (1988).
31. Jagadamma, S., Mayes, M. A., Zinn, Y. L., Gísladóttir, G. & Russell, A. E. Sorption of organic carbon compounds to the fine fraction of surface and subsurface soils. *Geoderma* **213**, 79–86 (2014).
32. Griffin, J. J., Windom & Goldberg, E. D. The distribution of clay minerals in the World Ocean. *Deep-Sea Res.* **15**, 433–459 (1968).
33. Lal, R. Soil carbon sequestration impacts on global climate change and food security. *Science* **304**, 1623–1627 (2004).
34. Paustian, K. *et al.* Climate-smart soils. *Nature* **532**, 49–57 (2016).
35. Six, J., Elliott, E. T. & Paustian, K. Soil structure and soil organic matter: II. a normalized stability index and the effect of mineralogy. *Soil Sci. Soc. Am. J.* **64**, 1042–1049 (2000).
36. Doetterl, S. *et al.* Soil carbon storage controlled by interactions between geochemistry and climate. *Nature Geosci.* **8**, 780–783 (2015).
37. Lawrence, C. R., Harden, J. W., Xu, X., Schulz, M. S. & Trumbore, S. E. Long-term controls on soil organic carbon with depth and time: A case study from the Cowlitz River Chronosequence, WA USA. *Geoderma* **247/248**, 73–87 (2015).

## Data Citation

1. Ito, A. & Wagai, R. PANGAEA <https://doi.org/10.1594/PANGAEA.868929> (2016).

## Acknowledgements

This study used the Harmonized World Soil Database, version 1.21, developed at the International Institute of Applied System Analysis. This study was partly supported by Japan Society for the Promotion of Science (JSPS) KAKENHI Grant Numbers JP15H02810 and JP17H01867.

## Author Contributions

A.I. designed the study, conducted the meta-analysis and mapping, and drafted the manuscript. R.W. advised on pedological considerations for data collection, classification, and analyses.

## Additional Information

**Competing interests:** The authors declare no competing financial interests.

**How to cite this article:** Ito, A. & Wagai, R. Global distribution of clay-size minerals on land surface for biogeochemical and climatological studies. *Sci. Data* **4**:170103 doi: 10.1038/sdata.2017.103 (2017).

**Publisher's note:** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>

The Creative Commons Public Domain Dedication waiver <http://creativecommons.org/publicdomain/zero/1.0/> applies to the metadata files made available in this article.

© The Author(s) 2017