RESEARCH ARTICLE

Critical Evaluation of Particle Size Distribution Models Using Soil Data Obtained with a Laser Diffraction Method

Wang Weipeng^{1,2}, Liu Jianli^{1*}, Zhao Bingzi¹, Zhang Jiabao¹, Li Xiaopeng¹, Yan Yifan^{1,2}

1 State Key Laboratory of Soil and Sustainable Agriculture, Institute of Soil Science, Chinese Academy of Sciences, Nanjing, China, **2** Graduate University of Chinese Academy of Sciences, Beijing, China

* Jlliu@issas.ac.cn

OPEN ACCESS

Citation: Weipeng W, Jianli L, Bingzi Z, Jiabao Z, Xiaopeng L, Yifan Y (2015) Critical Evaluation of Particle Size Distribution Models Using Soil Data Obtained with a Laser Diffraction Method. PLoS ONE 10(4): e0125048. doi:10.1371/journal.pone.0125048

Academic Editor: Wenju Liang, Chinese Academy of Sciences, CHINA

Received: December 24, 2014

Accepted: March 10, 2015

Published: April 30, 2015

Copyright: © 2015 Weipeng et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All relevant data are within the paper and its Supporting Information files.

Funding: This study was funded by the National Basic Research Program of China (973 Program No. 2011CB100506)—LJL; National Natural Science Foundation of China (Contract No. 41171179)—LJL. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

Abstract

Mathematical descriptions of classical particle size distribution (PSD) data are often used to estimate soil hydraulic properties. Laser diffraction methods (LDM) now provide more detailed PSD measurements, but deriving a function to characterize the entire range of particle sizes is a major challenge. The aim of this study was to compare the performance of eighteen PSD functions for fitting LDM data sets from a wide range of soil textures. These models include five lognormal models, five logistic models, four van Genuchten models, two Fredlund models, a logarithmic model, and an Andersson model. The fits were evaluated using Akaike's information criterion (AIC), adjusted R^2 , and root-mean-square error (RMSE). The results indicated that the Fredlund models (FRED3 and FRED4) had the best performance for most of the soils studied, followed by one logistic growth function extension model (MLOG3) and three lognormal models (ONLG3, ORLG3, and SHCA3). The performance of most PSD models was better for soils with higher silt content and poorer for soils with higher clay and sand content. The FRED4 model best described the PSD of clay, silty clay, clay loam, silty clay loam, silty loam, loam, and sandy loam, whereas FRED3, MLOG3, ONLG3, ORLG3, and SHCA3 showed better performance for most soils studied.

Introduction

Modeling the size distribution of soil particles to obtain a continuous particle size distribution (PSD) curves is useful for understanding essential soil properties such as pore distribution, water retention, hydraulic conductivity, and thermal and adsorption properties [1–4]. However, conventional PSD analysis captures only a limited number of particle mass fractions. Therefore, to be able to use these discrete experimental PSD data to estimate other soil properties, many researchers have used parametric functions to extend the limited scope of PSD data.

Soil PSD is frequently assumed to follow a lognormal distribution [5, 6], although some soils do have a bimodal PSD [7]. Buchan and Hwang compared five lognormal models and certain other types of models in terms of their fit to experimental PSD data [8, 9]. Their results showed that the bimodal model gave a marginally better fit, but incorporates a sub-clay mode,

and the Jaky model gave a good fit to data for many soils, better than the standard lognormal model for 23 soils. The latter included models based on water-retention curves [10–12], the fragmentation model [13], power-law functions based on fractal geometry [13, 14], exponential functions [15], Gompertz equations [16], and a model estimating PSD from limited soil texture data [17].

Many researchers have relied upon a single PSD model at one time to represent a wide range of soil texture classes, although several studies have suggested that the performance of PSD models can be affected by soil texture [18, 9]. Buchan [5] proposed that lognormal models are adequate to describe only half the United States Department of Agriculture (USDA)'s soil texture triangle, and more complex models are needed to depict other soils such as sandy clay loam, sandy clay, and most clays. Rousseva [15] proposed and investigated the suitability of exponential and power-law distribution models to fit PSDs with diverse shapes and varying numbers of measured points. The results showed that model performance was affected by soil texture, but that these functions were a good fit for the PSD of both fine- and coarse-textured soils. Fredlund [10] pointed out that the suitability of their own model improved as the proportion of clay in soils increased. Contrasting results have been obtained for the fitting of some models. For instance, in some cases, the Skaggs model performed poorly for silty soils [17], whereas in other cases, the same model performed better when silt content was higher [18].

Soil PSD data in most databases are obtained via the sieve-pipette method (SPM) based on Stokes's law [19]. This traditional approach is a tedious and time-consuming procedure and does not adequately describe soil PSD, especially for the clay fraction ($<2\ \mu\text{m}$) [20]. As a faster alternative, laser diffraction method (LDM) techniques have been increasingly used recently because they require a much smaller sample and provide highly reproducible PSD results, encompassing a broader range of size classes than conventional methods [21–24]. The LDM was reported to underestimate the clay fraction and overestimate the silt fraction of soils compared to the SPM [25]. Some comparative studies of PSD functions with various underlying assumptions have been performed, mostly with sedimentation data sets [26, 8, 18, 9]. In a comparison of several PSD models with PSD data obtained using both LDM and SPM approaches, fractal and exponential functions fitted poorly, whereas the suitability of the Gompertz model increased with clay content for the LDM data sets [26]. In the same analysis, the Fredlund function provided very good fits with SPM-derived PSD, but not with the corresponding LDM data sets. The lognormal function showed better overall performance and provided very good fits with both SPM and LDM data sets. However, Buchan [5] found that a lognormal model could adequately describe only about half the United States Department of Agriculture (USDA)'s texture triangle.

Few comparisons have been performed of the fitting performance of PSD functions based on LDM data sets. Furthermore, the effect of texture on PSD model performance has not been fully investigated. Therefore, the objectives of this study were to (1) use a variety of models with different underlying assumptions to test the fit of varying PSD functions to laser-derived PSD of soils from across China, and (2) to determine whether soil texture significantly affects model performance. Three statistical indices were used to define the best models: the adjusted root-mean-square error (*RMSE*), the adjusted R^2 , and Akaike's information criterion (*AIC*).

Materials and Methods

Soil samples

i) No specific permissions were required for these sampling points; ii) these field studies did not involve endangered or protected species.

Fig 1. Textural composition of soil samples.

doi:10.1371/journal.pone.0125048.g001

Samples of 1013 soils were collected from 13 provinces of China, representing a diversity of parent materials, climates, and times of soil formation and considerable variance in soil PSD, texture (Fig 1), and mineralogy. The sampling locations were as follows: 96 soil profiles were taken from Yellow River flood areas (Fengqiu, Henan), with samples collected every 10 cm in each profile from depths of 0–200 cm, for a sub-total of 960 samples; 6 samples were obtained from paddy soils in the Taihu Lake region (Changshu, Jiangsu); and 47 soil samples were obtained from 10 different provinces in China (Table 1). Samples were collected, air-dried, gently crushed, and dry-sieved through a 2-mm mesh to remove coarse fragments. Hydrogen peroxide at 30% concentration was used to remove the organic materials.

Apparatus and parameter settings for the LDM

The main illumination source for PSD analysis in a Beckman-Coulter LS-13 320 was a 5 mW laser diode (wavelength: 750 nm). The analyzer provided PSD in the form of 93 bins within the diameter range of 0.38–2000 μm (19 bins in the range of 0.38–2 μm and 74 bins in the range of 2–2000 μm). Because the LD analyzer required a very small subsample (<5 g) for the PSD

Table 1. General characteristics of the different types of soils studied.

Location	Classification ¹	Parent Material	SPM (%)			Clay mineral composition ²
			Clay	Silt	Sand	
Fengqiu, Henan	Durudepts	Yellow River alluvial	9.68–18.56	12.57–40.32	28.62–79.58	Quartz, Hydromica, FeldsparChlorite
Fengqiu, Henan	Haplustepts	Yellow River alluvial	4.26–25.36	9.28–53.35	20.82–86.48	Quartz, Hydromica, FeldsparChlorite
Fengqiu, Henan	Fragiudepts	Yellow River alluvial	5.86–29.66	6.27–55.15	25.82–89.48	Quartz, Hydromica, FeldsparChlorite
Changshu, Jiangsu	Aquanthepts	Lacustrine deposits	27.16	47.58	25.26	Quartz, Feldspar Hydromica
Changshu, Jiangsu	Aquanthepts	Lacustrine deposits	9.60	55.45	34.95	Quartz, Feldspar Hydromica
Changshu, Jiangsu	Aquanthepts	Lacustrine deposits	26.00	45.70	28.30	Quartz, Feldspar Hydromica
Changshu, Jiangsu	Aquanthepts	Lacustrine deposits	21.10	51.60	27.30	Quartz, Feldspar Hydromica
Changshu, Jiangsu	Aquanthepts	Lacustrine deposits	20.59	47.59	31.81	Quartz, Feldspar Hydromica
Changshu, Jiangsu	Aquanthepts	Lacustrine deposits	25.10	48.60	26.30	Quartz, Feldspar Hydromica
Nanjing, Jiangsu	Aquanthepts	Xia—Shu loess	27.46	46.94	25.60	illite, Kaolinite, Vermiculite
Nanjing, Jiangsu	Aquanthepts	Xia—Shu loess	18.27	46.91	34.82	illite, Kaolinite, Vermiculite
Yingtian, Jiangxi	Rhodudults	Red sandstone	39.97	27.17	32.86	Quartz, Montmorillonite, Vermiculite,
Yingtian, Jiangxi	Sulfudepts	Red sandstone	24.30	38.80	36.90	Quartz, Montmorillonite, Kaolinite
Yingtian, Jiangxi	Sulfudepts	Red sandstone	21.10	43.80	35.10	Quartz, Montmorillonite, Kaolinite
Yingtian, Jiangxi	Hapludults	Red sandstone	10.81	11.35	77.84	Quartz, Montmorillonite, Kaolinite
Shantou, Guangdong	Rhodudults	Granite	33.30	31.26	35.44	Quartz, Hydromica, Chlorite
Shenzhen, Guangdong	Hapludox	Granite	15.72	19.43	64.85	Quartz, Hydromica, Kaolinite
Binhai, Tianjin	Haplaquepts	Marine sediments	39.18	35.38	25.45	illite, Montmorillonite, Chlorite
Urumuqi, Xinjiang	Calciargids	Proluvium	18.72	31.83	49.45	illite, Chlorite, Kaolinite
Yanglin, Shanxi	Plagganthepts	Allochthonous soil	16.26	49.46	34.28	illite, Montmorillonite, Vermiculite
Beibei, Chongqing	Puplidepts	Jurassic Period Zisha Rock	29.97	34.95	35.08	Vermiculite, Kaolinite
Haikou, Hainan	Acraquox	Basalt	59.73	18.58	21.69	Kaolinite, Gibbsite, Quartz
Wencang, Hainan	Acrudox	Basalt	64.14	14.61	21.25	Kaolinite, Gibbsite, Quartz
Nehe, Heilongjiang	Argiaquolls	Proluvium	39.07	38.27	22.66	Montmorillonite, Hydromica, Feldspar
Binxiang, Heilongjiang	Frigudepts	Granitic gneiss	33.85	44.18	21.97	Montmorillonite, Hydromica, Feldspar
Shenyang, Liaoning	Natrudalfs	Quaternary Malan loess	20.05	32.12	47.82	Montmorillonite, Hydromica, Kaolinite
Kunming, Yunnan	Fragiudults	Argillaceous rock	52.89	19.04	28.08	Kaolinite, Gibbsite
Ningbo, Zhejiang	Haplaquepts	Marine sediments	11.40	54.30	34.30	illite, Kaolinite, Vermiculite

SPM = sieve-pipette method;

¹USDA = United States Department of Agriculture; NRCS = National Research Council Service;

Classification: the soils were classified according to the USDA, NRCS: Keys to Soil Taxonomy, Eleventh Edition.

²A semi-quantitative result of soil mineral composition by X-ray diffraction patterns.

doi:10.1371/journal.pone.0125048.t001

determination by LDM, each sample was blended several times before testing to minimize the potential variability of the results caused by differences between subsamples. Soil samples (5 g, <2 mm) received the same pretreatment as for the SPM [19]; particles were then transferred to 100-mL cylinders containing 10 mL of a 0.5 mol L⁻¹ sodium hexametaphosphate (for soils with pH >7.5), sodium oxalate (soil pH = 6.5–7.5), or sodium hydroxide (soil pH <5.5) solution.

Table 2. Particle size distribution (PSD) functions used in the study.

Model N°.	Abbreviation ¹	Model type and Author	Equation ²	Parameters ³
1	JAKY1	Jacky [28]	$F(d) = \exp \{-1/p^2 [\ln(d/d_0)]^2\}$	p
2	LOGL2	Logarithmic [30]	$F(d) = a \ln d + b$	a, b
3	LOGI2	Logistic growth model [17]	$F(d) = 1/(1 + \alpha \exp(-\beta d))$	α, β
4	LOGN2	Simple lognormal [5]	$F_1(x) = (1 + \operatorname{erf}[(x - \mu_1)/\sigma_1 \sqrt{2}])/2 \quad (x \geq \mu_1)$ $F_2(x) = (1 - \operatorname{erf}[(x - \mu_1)/\sigma_1 \sqrt{2}])/2 \quad (x < \mu_1)$ $x = \ln(d)$	μ, σ
5	VANH2	Haverkamp—Parlange [11]	$F(d) = 1/[1 + (\frac{d}{d_0})]^{1-1/\beta}$	α, β
6	VANZ2	Modified van Genuchten [30–31]	$F(d) = d_{\min} + (1 - d_{\min}) [1 + (\alpha d)^{\beta}]^{-1+1/\beta}$	$\alpha, \beta, d_{\min}$
7	FRED3	Fredlund [10]	$F(d) = \frac{1}{\left\{ \ln \left[\exp(1) + \left(\frac{d}{d_0} \right)^{\alpha} \right] \right\}^m} \left\{ 1 - \left[\frac{\ln \left(1 + \frac{1}{\beta} \right)}{\ln \left(1 + \frac{d/d_0}{1 + d_{\min}} \right)} \right]^7 \right\}$	α, m, n
8	LOGI3	Logistic growth function extension model 1	$F(d) = \varepsilon/(1 + \alpha \exp(-\beta d))$	$\alpha, \beta, \varepsilon$
9	MLOGI3	Modified logistic growth model	$F(d) = 1/[1 + \alpha \exp(-\beta d^{\gamma})]$	$\alpha, \beta, \varepsilon$
10	ONLG3	Offset Non-renormalized Lognormal [8]	$G(x) = F(x) + c_1 x = \ln d F(x)$ defined by LOGN2 model.	μ, σ, c_1
11	ORLG3	Offset Renormalized Lognormal [8]	$G(x) = (1 - c_2)F(x) + c_2 x = \ln d F(x)$ defined by LOGN2 model.	μ, σ, c_2
12	SHCA3	Shiozawa-Campbell [3]	$G(x) = \varepsilon F_1(x) + (1 - \varepsilon)F_2(x) \quad F_1(x) = (1 + \operatorname{erf}[(x - \mu_1)/\sigma_1 \sqrt{2}])/2 \quad (x \geq \mu_1)$ $F_2(x) = (1 - \operatorname{erf}[(x - \mu_1)/\sigma_1 \sqrt{2}])/2 \quad (x < \mu_1) \quad x = \ln(d)$	σ_1, μ_1
13	VANZ3	van Genuchten, function extension model 1	$F(d) = \varepsilon + (1 - \varepsilon) [1 + (\alpha d)^{\beta}]^{-1+1/\beta}$	$\alpha, \beta, \varepsilon$
14	ANDE4	Andersson [32]	$F(d) = d_{f1} + \alpha \tan^{-1}(\beta \log \frac{d}{d_0})$	$\alpha, \beta, d_{f1}, d_0$
15	FRED4	Fredlund [10]	$F(d) = \frac{1}{\left\{ \ln \left[\exp(1) + \left(\frac{d}{d_0} \right)^{\alpha} \right] \right\}^m} \left\{ 1 - \left[\frac{\ln \left(1 + \frac{d/d_0}{1 + d_{\min}} \right)}{\ln \left(1 + \frac{d/d_0}{1 + d_{\min}} \right)} \right]^7 \right\}$	$\alpha, n, m, d_f, d_{\min}$
16	GOMP4	Gompertz [16]	$F(d) = \alpha + \gamma \exp\{-\exp[-\beta(d - \mu)]\}$	$\alpha, \beta, \gamma, \mu$
17	LOGI4	Logistic growth function extension model 2	$F(d) = \beta + (\alpha - \beta)/(1 + (d/\varepsilon)^{\gamma})$	$\alpha, \beta, \varepsilon, \gamma$
18	VANG4	van Genuchten, function extension model 2	$F(d) = \varepsilon + (\gamma - \varepsilon)[\gamma + (\alpha d)^{\beta}]^{-1+1/\beta}$	$\alpha, \beta, \varepsilon, \gamma$

¹The number in the model abbreviation represents the number of the function parameters.

² d , particle diameter in mm. Erf, error function

³ p : characterizes the stretching of the PSD curve; $d_0 = 2$ mm; d_f , number of fine particles; d_{ff} , number of d_0 ; $d_{\min} = 0.001$ mm. d_0 , diameter of the most frequent particle. a and b , fitting parameters; c_1, c_2 : offset parameters. α, β, m and n : shape parameters; μ , mean; σ , standard deviation; $d_{\min}$: minimum diameter; ε and γ fitting parameters. $\sigma_1 = 1.00$; $\mu_1 = -1.96$.

doi:10.1371/journal.pone.0125048.t002

The samples were left overnight to disperse, then transferred to fluid modules and diluted to 10% obscuration [27]. Each sample was subjected to a 60-sec optical measurement [2]. A soil particle refractive index of 1.50 and a soil particle absorption index of 0.01 were regulated for the Mie theory model [24].

Particle Size Distribution Functions

Eighteen PSD functions with different underlying assumptions were evaluated using PSD data derived from the procedures described above (Table 2). The five models were a Jaky function (JAKY1, with the number representing the quantity of model parameters) [28]; a simple lognormal (LOGN2) [5]; an offset non-renormalized lognormal model (ONLG3) [8]; an offset renormalized lognormal model (ORLG3) [8]; and a Shiozawa-Campbell model (SHCA3) all of which assume that the soil PSD follows a lognormal pattern [29]. In the JAKY1 model, p represented the PSD index that describes the stretching of the curve, and d_0 was the maximum diameter (see equations in Table 2). In LOGN2, ONLG3, ORLG3, and SHCA3, F_n was

the cumulative normal distribution function, μ was the mean, and σ was the standard deviation [8]. For the logarithmic function (LOGL2) as used by Zhuang to generate detailed PSD for fine-textured soils, a and b were fitting parameters [30]. The five logistic models a logistic growth model (LOGI2) [17], and three logistic growth model extension models, LOGI3, MLOGI3 and LOGI4, along with the Gompertz function (GOMP4) [16] were based on the assumption that the soil PSD followed a logarithmic growth function. Based on the LOGI2 model, the LOGI3, MLOGI3, and LOGI4 models were generated by respectively altering, adding, or reducing the parameters of the original LOGI2. The main procedure used for this extension was the inclusion of one or more adaptation coefficients to replace the constants in the original formula (see Eqs 1 through 4 below). GOMP4 was a logistic function represented by a closed-form equation, with α , β , γ , and μ being the shape parameters of the curve. The four van Genuchten type models [31] the Havrkamp-Parlang model (VANH2) [11], and the VANH2 extension models (VANZ2, VANZ3, and VANG4; see Eqs 5–8 below) were derived from the van Genuchten function that represents the water-retention characteristics of soils. This approach was based mainly on the similarity between PSD shapes and water-retention curves. The Fredlund functions (FRED3, FRED4) and the Andersson model were also based on the similarity between soil water-retention and PSD curves [10]. In the FRED3 and FRED4 models, α was the point of inflection of the curve, n was related to its steepest slope, m to its shape near the fine region, d_f was the number of fine particles, and d_m was the minimum particle size. Notably, FRED3 was a special case of FRED4 ($d_f = 0.01$). In the Andersson model, α , β , were shape parameters, and d_0 denoted the most frequent particle diameter corresponding to the cumulative percent d_{f1} [32, 33]. Because of the limited number of classes in the clay range, even with the LDM (0.5 and 1 μm), the bimodal model was not evaluated in this study. All models were fitted to experimental PSD data for the 1013 sampled soils using an iterative nonlinear regression procedure which found the values of the fitting parameters that gave the best fit between the model and the data. This procedure was performed using the modified nonlinear least-squares method (MNLS) [34].

The logistic growth model and its extensions

The logistic growth function [35] is widely used in studies involving forecasting of biotic populations and system reliability analysis. When used to depict soil PSD, the model is modified as follows:

$$F(d) = \frac{1}{1 + \alpha \exp(-\beta d)} \quad (1)$$

α, β : shape parameters ($\alpha > 0, \beta > 0$).

To improve the fitting ability of LOGI2, a fitting coefficient ε was used as a substitute for the LOGI2 numerator constant “1”; in this way, the logistic growth function extension model LOGI3 was obtained. This model can be expressed as follows:

$$F(d) = \frac{\varepsilon}{1 + \alpha \exp(-\beta d)} \quad (2)$$

α, β : shape parameters ($\alpha > 0, \beta > 0$); ε : fitting parameter.

Based on the LOGI2 model, the MLOG3 model was generated by adding an index parameter c to Equation [1] as follows:

$$F(d) = \frac{1}{1 + \alpha \exp(-\beta d^c)} \quad (3)$$

α, β : shape parameters ($\alpha > 0, \beta > 0$); c : fitting parameter.

Similarly, the LOGI4 model was obtained by adjusting the parameters of the MLOG3 model and rewriting Equation [3] as:

$$F(d) = \beta + \frac{\alpha - \beta}{1 + \left(\frac{d}{\epsilon}\right)^\gamma} \quad (4)$$

α, β : shape parameters ($\alpha > 0, \beta > 0$); ϵ, γ : fitting parameters.

The van Genuchten model and its extensions

The van Genuchten model (VANH2) was developed initially to depict soil moisture characteristic curves. To use it to depict soil PSD, the original function was modified as follows:

$$F(d) = \frac{1}{\left[1 + \frac{\alpha \beta}{d}\right]^{1+\frac{1}{\beta}}} \quad (5)$$

α, β : shape parameters.

Based on the VANH2 model, Zhuang modified the van Genuchten model and used the extension model (VANZ2) for soil PSD as follows [30]:

$$F(d) = d_{min} + (1 - d_{min})[1 + (\alpha d)^\beta]^{-1+\frac{1}{\beta}} \quad (6)$$

α, β : shape parameters; d_{min} : minimum diameter.

In addition, by substituting the fitting parameter ' ϵ ' for the d_{min} of Equation [6], the generic form of the VANH2 model can be obtained. This van Genuchten extension model was called VANZ3 and defined as follows:

$$F(d) = \epsilon + (1 - \epsilon)[1 + (\alpha d)^\beta]^{-1+\frac{1}{\beta}} \quad (7)$$

α, β : shape parameters; ϵ : fitting parameter.

Again, using a fitting parameter γ to replace the constant 1 of Equation [7], four-parameter van Genuchten model (VANG4) was obtained as follows:

$$F(d) = \epsilon + (\gamma - \epsilon)[\gamma + (\alpha d)^\beta]^{-1+\frac{1}{\beta}} \quad (8)$$

α, β : shape parameters; ϵ, γ : fitting parameters.

Model Comparison and Fitting Techniques

Several approaches have been reported for choosing the most suitable model. The simplest approach is to find the model that minimizes the discrepancy between observed and predicted data. However, it must be recognized that as the number of fitted parameters increases, the fitting performance generally improves. This occurs at the cost of a corresponding increase in the probability of over parameterization. The PSD models considered here required between one and four fitting parameters. To minimize the risk of "over fitting" models with too many parameters, a better approach was to define the optimum model as the one that fitted the data well with the least number of fitting parameters, all other conditions being equal. Several

researchers have used this kind of criterion to select the best model. *AIC* [36] has been frequently used to identify the best predictive function for soil moisture characteristics because it can be determined indirectly from other easily measured soil properties [37], and to select the best soil hydraulic function [9, 18, 26].

The correspondence of model predictions with experimental data was ascertained using the adjusted root mean square error (*RMSE*) and R^2 because of the unequal number of parameters in the PSD functions:

$$RMSE (adjusted) = \sqrt{\frac{SSE}{n-p}}, \quad (9)$$

$$R^2 (adjusted) = 1 - \left[\frac{SSE/(n-p)}{SST/(n-1)} \right], \quad (10)$$

where *SSE* is the sum of squared errors, *SST* the total sum of squares, *n* the total number of PSD data points, *P* the number of model parameters, and

$$SSE = \sum_{i=1}^n (Y_m - Y_c)^2, \quad (11)$$

Y_m : measured; Y_c : estimated.

AIC values were used to compare the various models. According to this criterion, the model with the smallest *AIC* value was selected as the best soil description. *AIC* imposes a penalty for additional fitting parameters, which was calculated as:

$$AIC = N \ln(SSE) + 2P. \quad (12)$$

Theory for verifying the precision of the LDM PSD

As pointed out by Miller and Schaetzl [38], the cumulative bin difference (*CBD*)^(Eq.13) plus one standard deviation (*SD*) can be used as a precision threshold (*PH*)^(Eq.14) to verify the measurement variability among the subsamples. In this way, the data from the first two subsamples were used as a starting point to ascertain the normal amount of intra-sample variation. The *CBD* was larger when many bins had small differences, as in the LDM analysis which provided a PSD with 117 bins. The *CBD* was calculated by adding the absolute differences between the two subsamples for all bins. The definition of the *CBD* was as follows:

$$CBD = \sum_{k=1}^n |a_k - b_k|, \quad (13)$$

where *a* is the first subsample value, *b* is the second subsample value, and *n* is the number of bins, and that of *PH* was:

$$PH = \frac{\sum_1^m CBD}{m} + SD, \quad (14)$$

where *PH* is the precision threshold, *m* is the number of samples, and *SD* is the standard deviation for *CBD*s of all samples.

In the present study, *PH* was used to identify outliers. When a set of subsample pairs had a *CBD* exceeding 1 *SD* plus the mean *CBD* of all samples, the precision or repeatability of the PSD results were considered insufficient for this type of analysis. In these cases, additional

subsamples were measured using the same procedure as for the previous subsamples until every sample had a pair of subsamples with a *CBD* less than the *PH*, which meant that some samples had to undergo more than two analyses to obtain an eligible *CBD*. The samples chosen for additional measurements were only those that continued to have a *CBD* greater than *PH*. For those samples which had been measured several times, the mean of the two subsample values with the least *CBD* was used to represent the PSD for each sample.

Results and Discussion

LDM provides a rapid solution for obtaining soil PSD, especially when analyzing a large number of soil samples. However, there are disadvantages associated with the use of LDM for PSD measurements over the traditional SPM. In this study, the LDM underestimated the clay fraction by as much as 35% and overestimated the silt fraction by about 27% compared with SPM (data not shown). This is consistent with results obtained elsewhere [22, 23, 25]. These differences are attributable to the heterogeneity of soil particle density and the deviation of particle shapes from sphericity. Sedimentation methods assume a single particle density, which is a major source of error, whereas LDM is independent of particle density. Deviation from sphericity, however, is the major factor that has disparate effects on both methods. In the case of LDM, an irregularly shaped soil particle reflects a cross-sectional area greater than that of a sphere of equivalent volume [25]. Therefore, particles tend to be assigned to larger size fractions of the PSD, thereby resulting in a smaller clay fraction. On the other hand, nonspherical particles in the pipette method have longer settling times than their equivalent spheres, resulting in overestimation of the clay fraction.

Definition of the optimum model

Given the inherent limitation of PSD based on SPM—that the soil PSD becomes characterized by limited particle size data—this study verified the performance of the mathematical models using SPM PSD results with a different data set. Each soil sample's six classes (2, 20, 50, 125, 250, and 1000 μm) of the LDM PSD data were extracted to construct the comparison LDM PSD set. Results for model fitting of the PSD functions to the data sets are shown in Figs 2 and 3. Overall, *AIC* values ranged from -642.432 to 154.635 (in a 93-class PSD) and from -0.382 to -87.286 (in a six-class PSD data). The adjusted *RMSE* obtained for regression of estimated on measured cumulative volume fractions ranged from 0 to 0.233 (for 93-class PSD data) and from 0 to 0.205 (for six-class PSD data). With adjustments to R^2 values among models and between groups, the differences were somewhat less than what had been noted for *AIC* values and adjusted *RMSE* values (S1 and S2 Datasets).

Low *AIC*, low adjusted *RMSE*, and high adjusted R^2 indicate better correspondence of a model with experimental data, and vice versa. Based on the performance of eighteen models with the 93-class PSD data (Fig 2), the models were ranked in order from best to worst fit, according to the comparison criteria: FRED4 > MLOG3 > FRED3 > VANH2 > SHCA3 > ORLG3 > ONLG3 > ANDE4 > LOGI4 > LOGN2 > VANZ3 > VANG4 > VANZ2 > GOMP4 > LOGI2 > LOGI3 > JAKY1 > LogL2. Correspondingly, for the comparison data set, the sequence of model suitability was: FRED4 > FRED3 > MLOG3 > ORLG3 > SHCA3 > LOGI4 > VANZ2 > VANG4 > VANH2 > ONLG3 > LOGN2 > VANZ3 > ANDE4 > GOMP4 > LOGI2 > LOGI3 > JAKY1 > LogL2.

The lowest adjusted *RMSE* values for both the 93-class and the 6-class PSD data, indicating best fit, were obtained with the FRED4 function, followed by the MLOGI3 or FRED3 functions. Comparatively, the worst fits were obtained with the LOGL2 function, followed by the JAKY1, LOGI3, LOGI2, and GOMP4 functions. ONLG3 sometimes showed similar fitting

Fig 2. Performance of eighteen mathematical functions fitted to ninety-three class particle size distributions (PSD) based on soil data obtained with laser diffraction methods (LDM), in terms of Akaike's information criterion (AIC), adjusted R^2 , and adjusted root mean square error (RMSE).

doi:10.1371/journal.pone.0125048.g002

performance to the ORLG3 and SHCA3 functions. As expected, the PSD functions other than JAKY1 and LOGL2 gave relatively better fits with the 93-bin PSD data compared to those with the comparison set of 6-bin PSD data. These differences in fitting ability may originate from the difference of data density between the two groups' PSD models; in other words for the 6-bin PSD data, the comparison set's PSD curve had longer between-point distances than the curve for soils with the ninety-three bin PSD data. This difference may render it impractical to construct a soil PSD accumulation distribution curve with an "actual shape". Fortunately, the relative performance among the different models in the two groups was basically consistent, which means that the experimental results for the two data groups can be used to analyze the actual effect of using these PSD models to derive continuous soil PSD curves.

An illustration of the comparative fit of the various models to three soil samples of varying textures is shown in Fig 4. It was surprising to find that other than JAKY1 and LOGL2, the models showed no distinct differences in their predicted values at each measurement point for sandy soil (Fig 4.1). This seemed to occur because the PSD has a simpler shape when sand content is high, which can be represented easily by various models. Moreover, this study found that the major difference among the models generally occurred for clay and silt fractions, which could be due to the multivariate shape of the cumulative PSD curve in the size range for such particles. This contrasted with the results of Hwang [9], who found based on SPM PSD

Fig 3. Performance of eighteen mathematical functions fitted to six-class (2, 20, 50, 125, 250 and 1000 μ m) particle size distributions (PSD) based on soil data obtained with the laser diffraction method (LDM), in terms of Akaike's information criterion (AIC), adjusted R^2 , and adjusted root mean square error (RMSE).

doi:10.1371/journal.pone.0125048.g003

data that the major discrepancy in model fitting performance among various PSD models occurred for the silt fraction of soils. The results presented here may be more comprehensive because a larger proportion of the PSD data came from clay fractions.

The LOGN2, MLOG3, LOGI4, VANH2, VANZ2, VANZ3, VANG4, ORLG3, SHCA3, and ANDE4 models consistently provided relatively good fits to all three samples, despite the differences in soil textures (Fig 4). This suggested that these models are suitable for describing detailed soil PSD. The FRED3 and FRED4 (Fig 4.8 and 4.9) models performed poorly for loam and clay soil, which seems to contradict the earlier results that both FRED3 and FRED4 performed better than other models in the RMSE and AIC analyses. Therefore, it seems that the PSD models that showed good performance for overall soil PSD were not always the most suitable models for individual soil texture classes. The LOGL2 models performed poorly for almost all soils (Fig 4.1–4.3), and therefore these models should be used with caution. Other PSD models had relatively poor performance for at least one soil type, e.g., JAKY1 performed well only for loamy soil, and GOMP4 was suitable only for the sand sample.

Again, as the number of model parameters increased, the extension models LOGI3, MLOG3, and LOGI4 all had better fits than that of the primary two-parameter LOGI2 model. Similarly, the extension models VANZ3 and VANG4 both had smaller AIC values than the original VANZ2 model. However with a series of lower AIC values, the three-parameter

Fig 4. Comparative fit of eighteen mathematical functions to particle size distribution data of different soil samples (n = 3). Sand (Henan soil, 0–20 cm), Loam (Henan soil, 0–20 cm), clay (Hainan soil, 0–20 cm).

doi:10.1371/journal.pone.0125048.g004

MLOG3 performed better than the four-parameter LOGI4, illustrating that introducing more parameters can appear to improve model fit, even when the model is not necessarily better.

Effect of soil texture on model performance

To examine further the effect of soil texture on model performance, trends in the variation of model AIC values were compared. According to the model comparison criteria used here, with increasing in the clay fraction of soils, all models except for VANH2 had poorer fits (Fig 5). Although the rise in AIC values was quite limited for some models, i.e., JAKY1, LOGI2, ONLNG3, ORLG3, and SHCA3, these trends provided an indication of the relative performance of PSD models with varying clay contents. It was also found that the fits of PSD models (except for LOGI2 and VANG4) declined with increasing sand content. Furthermore, the results presented here indicate that with higher sand content (40%), the AIC values for most models had a wider discrete range (Figs 6–8). On this basis, it can be noted that model performance deteriorated as the clay and sand proportions of soils increased, in contrast to previous reports. Hwang found that models performed better as clay content increased [9]. Bah came to similar conclusions after examining seven PSD models with 55 soil samples measured using LDM [26]. The difference in the results presented here may be due to the different data acquisition method and the PSD data density. Usually, with conventional SPM, soil PSD classification data are limited to the resolution of clay size classes, but based on the LDM, it is possible to obtain 19 bins of particle classification data in the clay fraction, which helps the models to depict the soil PSD characteristic distribution curve more accurately. Another contributing factor could be the relatively small and limited sample sizes that Bah used, which covered only two textural

Fig 5. Box plots showing model performance with varying clay content of soils.

doi:10.1371/journal.pone.0125048.g005

Fig 6. Box plots showing model performance with varying sand content of soils.

doi:10.1371/journal.pone.0125048.g006

classes of clay. The results presented here may be more representative because a larger, more comprehensive soil classification data set was used.

Model performance also varied greatly with silt content (Fig 7). As the silt fraction increased, the models (except for VANG4) had smaller AIC values and AIC box areas, indicating that the fit of the model improved with increasing silt proportion. Meanwhile, the performance of some models was basically stable, regardless of soil composition, i.e., VANZ2, ONLG3, ORLG3, SHCA3, and VANZ3. These models can therefore be expected to be robust for use under conditions where a large number of samples of different origins, profiles, and landscapes must be analyzed.

AIC analysis of models for soils of different clay, silt, and sand contents revealed that the performance of a PSD model can be affected by soil texture. Results indicated that FRED3, MLOG3, and four lognormal models (LOGN2, ONLG3, ORLG3, and SHCA3) performed consistently well for all kinds of soil texture in this study (Fig 8). The FRED4 model had the best fit overall, accounting for most of the smallest AIC values for all soils with different textures, except for sand and loamy sand soils. Other studies have previously established that FRED4 is highly suitable for modeling all soil texture classes [26, 8, 9]. The difference in the present results with FRED4, i.e., the finding that it did not perform well for all soil classes, was probably related to the fact that the LDM underestimated the clay fraction. ANDE4 performed well for loam and sandy loam soils. JAKY1, LOGL2, the two logistic models (LOGI2 and LOGI3), and the three van Genuchten models (VANH2, VANZ2, and VANZ3) showed poorer performance for most texture classes. GOMP4 also performed poorly for most soil types; however, for sand and loamy sand, it achieved the second most suitable fit with the PSD data, after SHCA3. LOGI4 seemed to be more appropriate for silty loam, clay, clay loam, loam, and silty clay loam. The performance of VANG4 was only relatively better in silty loam, clay, clay loam, sand, and sandy loam soils.

Fig 7. Box plots showing model performance with varying silt content of soils.

doi:10.1371/journal.pone.0125048.g007

Conclusions

The performance of eighteen models in describing the particle size distribution (PSD) of 1013 soil samples has been compared, with the following results. (1) The results of model fitting (AIC , adjusted R^2 , and $RMSE$) indicated that the FRED4 model provided the best fit for most samples (highest adjusted R^2 , lowest SSE, and very low AIC), (2) The performance of the PSD models was obviously affected by soil texture. Model fit declined as the clay and sand contents of soils increased, except for the VANH2 model. (3) The FRED4 and FRED3 models are preferable for clay, clay loam, loam, loamy sand, sandy clay loam, sandy loam, and silty clay soils. The MLOG3, ONLG3, ORLG3 and SHCA3 models can be used to describe the PSD of a full range of Chinese soils, whereas the remaining models had better performance on one or a few types of soil. (4) This research has referred to previous studies of soil PSD functions and to preliminary studies of several PSD function extension models. Note that these results were based on factors such as small sample size, a wide range of sampling sites, larger differences among soil types, organic matter content, mineral composition, and processes and soil development environment, which led to the relative lowly simulation precision of these PSD models. There is a need to evaluate more alternative models with richer, more comprehensive experimental data, obtained with the LDM from a wider range of soil textures encompassing a greater geographic area, to understand soil particle distributions more completely.

However, the method described here can be expected to be useful under certain conditions. For example, where a larger number of samples of different origins, profiles, and landscapes need to be analyzed, a small number of samples for each soil type can be extracted and analyzed by the SPM and LDM, following which a set of regressive equations can be established to improve the accuracy of LDM PSD results. In this way, error ranges for the LDM can be obtained, and the differences between LDM and SPM can be calibrated. This would provide a basis for choosing a suitable PSD model to estimate soil hydraulic properties.

Fig 8. Box plots showing AIC for eighteen particle size distribution functions across different soil texture classes.

doi:10.1371/journal.pone.0125048.g008

Supporting Information

S1 Dataset. The RMSE, R2, SSE and AIC for LDM data.
(XLS)

S2 Dataset. The RMSE, R2, SSE and AIC for PM Data.
(XLS)

Acknowledgments

We wish to thank editors for his constructive comments and anonymous reviewers for their insightful comments and suggestions.

Author Contributions

Conceived and designed the experiments: LJL WWP. Performed the experiments: WWP. Analyzed the data: WWP YYF LXP. Contributed reagents/materials/analysis tools: LJL ZBZ ZJB. Wrote the paper: WWP LJL.

References

1. Hajnos M, Lipiec J, Swieboda R, Sokolowska Z, Witkowska-Walczak B (2006) Complete characterization of pore size distribution of tilled and orchard soil using water retention curve, mercury porosimetry, nitrogen adsorption, and water desorption methods. *Geoderma*. 135: 307–314.
2. Ryzak M, Bieganski A (2011) Methodological aspects of determining soil particle-size distribution using the laser diffraction method. *J Plant Nutr Soil Sci*. 174:624–633.
3. Shiozawa S, Campbell GS (1992) Calculation of mean particle diameter and standard deviation from sand, silt, and clay fractions: Shiozawa, S; Campbell, G S *Soil Sci V152, N6, Dec 1991, P427-431. Int J Rock Mech Min*. 29:211.
4. Slawinski C, Walczak TR, Skierucha W (2006) Error analysis of water conductivity coefficient measurement by instantaneous profiles method. *Int Agrophys*. 20:55–61.
5. Buchan GD (1989) Applicability of the Simple Lognormal Model to Particle-Size Distribution in Soils. *Soil Sci* 147:155–161.
6. Hwang SI, Hong SP (2006) Estimating relative hydraulic conductivity from lognormally distributed particle-size data. *Geoderma*. 133:421–430.
7. Walker PH, Chittleborough DJ (1986) Development of particle-size distributions in some alfisols of southeastern Australia. *Soil Sci Soc Am J* 50:394–400.
8. Buchan GD, Grewal KS, Robson AB (1993) Improved Models of Particle-Size Distribution: An Illustration of Model Comparison Techniques. *Soil Sci Soc Am J* 57: 901–908.
9. Hwang SI, Lee KP, Lee DS, Powers SE (2002) Models for Estimating Soil Particle-Size Distributions. *Soil Sci Soc Am J*. 66:1143–1150.
10. Fredlund MD, Fredlund DG, Wilson GW (2000) An equation to represent grain-size distribution. *Can Geotech J*. 65:638–648.
11. Haverkamp R, Parlange JY (1986) Predicting the Water-Retention Curve from Particle-Size Distribution: 1. Sandy Soils Without Organic Matter. *Soil Sc*. 142:325–339.
12. Smettem K, Gregory P (1996) The relation between soil water retention and particle size distribution parameters for some predominantly sandy Western Australian soils. *Soil Res*. 34:695–708.
13. Bittelli M, Campbell GS, Flury M (1999) Characterization of particle-size distribution in soils with a fragmentation model. *Soil Sci Soc Am J* 63:782–788.
14. Filgueira RR, Fournier LL, Cerisola CI, Gelati P, Garcia MG (2006) Particle-size distribution in soils: A critical study of the fractal model validation. *Geoderma* 134: 327–334.
15. Rousseeva SS (1997) Data transformations between soil texture schemes. *Eur J Soil Sci*. 48:749–758.
16. Nemes A, Wösten JHM, Lilly A, Oude Voshaar JH (1999) Evaluation of different procedures to interpolate particle-size distributions to achieve compatibility within soil databases. *Geoderma*. 90:187–202.
17. Skaggs TH, Arya LM, Shouse PJ, Mohanty BP (2001) Estimating Particle-Size Distribution from Limited Soil Texture Data. *Soil Sci Soc Am J*. 65: 1038–1044.

18. Hwang SI (2004) Effect of texture on the performance of soil particle-size distribution models. *Geoderma*. 123:363–371.
19. Gee GW, Bauder JW (1986) Particle-size analysis. In: Klute A. (ed.) *Methods of soil analysis*, Part 1. 2nd edn. Agron. Monogr. 9. ASA and SSSA, Madison, WI, pp 383–411.
20. Gee GW, Bauder JW (1979) Particle-size analysis by hydrometer-simplified method for routine textural analysis and a sensitivity test of measurement parameters. *Soil Sci Soc Am J*. 43(5): 1004–1007.
21. Arriaga FJ, Lowery B, Mays MD (2006) A fast method for determining soil particle size distribution using a laser instrument. *Soil Sci*. 171:663–674.
22. Eshel G, Levy GJ, Mingelgrin U, Singer MJ (2004) Critical evaluation of the use of laser diffraction for particle-size distribution analysis. *Soil Sci Soc Am J*. 68:736–743.
23. Konert M, Vandenberghe JEF (1997) Comparison of laser grain size analysis with pipette and sieve analysis: a solution for the underestimation of the clay fraction. *Sedimentology*. 44:523–535.
24. Wang WP, Liu JL, Zhang JB, Li XP, Cheng YN, Xin WW, et al. (2013) Evaluation of Laser Diffraction Analysis of Particle Size Distribution of Typical Soils in China and Comparison with the Sieve-Pipette Method. *Soil Sci*. 178:194–204.
25. Pieri L, Bittelli M, Pisa PR (2006) Laser diffraction, transmission electron microscopy and image analysis to evaluate a bimodal Gaussian model for particle size distribution in soils. *Geoderma*. 135: 118–132.
26. Bah AR, Kravchuk O, Kirchhof G (2009) Fitting Performance of Particle-size Distribution Models on Data Derived by Conventional and Laser Diffraction Techniques. *Soil Sci Soc Am J*. 73:1101–1107.
27. Buurman P, Pape T, Reijneveld JA, de Jong F, van Gelder E (2001) Laser-diffraction and pipette-method grain sizing of Dutch sediments: correlations for fine fractions of marine, fluvial, and loess samples. *Neth J Geosci*. 80:49–57.
28. Jaky J (1944). *Soil mechanics*. (In Hungarian), Egyetemi Nyomda, Budapest.
29. Shiozawa S, Campbell GS (1991) On the calculation of mean particle diameter and standard-deviation from sand, silt and clay fractions. *Soil Sci*. 152:427–431.
30. Zhuang J, Jin Y, Miyazaki T (2001) Estimating water retention characteristic from soil particle-size distribution using a non-similar media concept. *Soil Sci*. 166: 308–321.
31. van Genuchten M Th (1980) A Closed-form Equation for Predicting the Hydraulic Conductivity of Unsaturated Soils1. *Soil Sci Soc Am J*. 44: 892–898.
32. Andersson S (1990) Markfysikaliska undersokningar i odlad jord. In: XXVI. Om mineraljordens och mullens rumsutfyllande egenskaper. En teoretisk studie. (In Swedish), Swedish University of agricultural sciences, Uppsala, pp 7.
33. Jauhainen M (2004) Relationships of particle size distribution curve, soil water retention curve and unsaturated hydraulic conductivity and their implications on water balance of forested and agricultural hillslopes. Helsinki University of Technology, Water Resources Publications, Espoo, pp 165.
34. Press WH, Teukolsky SA, Vetterling WT (1992) *Numerical Recipes in Fortran 77*. Cambridge, London
35. Thornley JHM (1990) A new formulation of the logistic growth equation and its application to leaf area growth. *Ann Bot*. 66:309–311.
36. Carrera J, Neuman SP (1986) Estimation of aquifer parameters under transient and steady-state conditions. 1. Maximum-likelihood method incorporating prior information. *Water Resour Res*. 22:199–210.
37. Minasny B, McBratney AB, Bristow KL (1999) Comparison of different approaches to the development of pedotransfer functions for water-retention curves. *Geoderma*. 93:225–253.
38. Miller BA, Schaetzl RJ (2012) Precision of Soil Particle Size Analysis using Laser Diffractometry. *Soil Science Society of America Journal*. 76(5): 1719–1727.