

Estimation of soil particle size distribution—From Katchinski to USDA scheme

Mao-Fen Li^{1, a}, Wei Wu^{2, b} and Hong-Bin Liu^{3, c}

¹Institute of Scientific and Technical Information, Chinese Academy of Tropical Agriculture Sciences, Danzhou 571737, China

²College of Computer and Information Science, Southwest University, Chongqing 400716, China

³College of Resources and Environment / Chongqing Key Laboratory of Digital Agriculture, Southwest University, Chongqing 400716, China

^amaofen.li@gmail.com, ^bweiwu_star@163.com, ^clbin@swu.edu.cn

Keywords: Particle size distribution; soil texture; Sichuan basin.

Abstract. For better use of soil particle size distribution (PSD) models and expand the use of large database held by national Soil Surveys (Katchinski's scheme) in China, 12 PSD models were proposed as the function of cumulative mass percentage of particles and diameter (mm) for 265 soil horizons (Katchinski's scheme) in a broader soil particle-size ranges, and together with 3 commonly used interpolations were validated to estimate the PSDs of USDA system especially particles <0.002mm and 0.02mm for other 49 soil horizons. The adjusted coefficient of determination (Adj. R²) and the Akaike's information criterion (AIC) were used to compare the quality of 12 numerical model fits. The ExpA and ExpD models with five parameters and Logistic model with four parameters showed relatively better performance. Correlation coefficient (r), absolute error (AE, %), and root mean square deviation (RMSD) were used to test the validation of the PSD models and interpolations on the soil particles <0.002 and 0.02mm, respectively. And it was noteworthy that the RMSDs yielded from interpolations were much larger than 12 models. Interpolations did not show any superiority towards numerical functions at particles <0.002 mm. Cubic spline interpolation performed worst with mean AE of 20% at particles <0.02 mm. This result suggested that the cubic spline interpolation could not use to estimate the soil particles <0.02 mm for Sichuan basin soil.

Introduction

Soil texture data (soil particle size distribution: PSD) of layers are indispensable and fundamental, for other soil data (e.g. soil saturation, saturated hydraulic conductivity) could be deduced from soil particle data. There are as yet no universally accepted classification schemes for soil texture ^[1] and soil data are rarely compatible across nations. For example, soil particle size fraction data in China are mainly classified by the Katschinski's scheme, whereas models such as DSSAT adopt the USDA system. The need to expand the use of the large database held by many national Soil Surveys is obvious.

The subject of indirect estimation of PSD from sparse particle size data that readily measured has been analyzed in the past with kinds of functions, such as weibull ^[2-4], Logarithm^[5], Exponential ^[1,6], power ^[1] and interpolations ^[7-8]. In China, interpolation algorithm was often applied to estimate PSD from one scheme to another. As a descriptor of soil particle distributions, several things are noteworthy about aforementioned methods. For example, most of the studies consider the soils with particles <2 mm while the diameter >2 mm is taken as an exception. However, soil particles >2 mm have shown ecological significance in the landscape ^[9], and soil particles >2 mm are widely spread in practice in hilly areas like southwest China. The soils with coarse particles (diameter >2 mm) have not yet been studied with PSD models by far. For instance, the measured cumulative particle-size <3mm was 100% in the work of Rousseva^[1]. So the reliability of these PSD models and methods needs to be identified in a broader soil particle-size ranges especially in the soils with particles >2 mm.

The previous results suggested that the performance of the PSD models depended more on the texture of the soil sample. Rousseva^[1] used two PSD models (exponential and power-law distribution

models) to test the effects of texture on their performance. He found that the exponential and power-law distribution models could be used to represent the PSDs of fine- and coarse-textured soils, respectively. Hwang^[4] reported that the performance of most PSD models improved with the increase in clay content in soils. All the textures mentioned above mean the sands, silts, and clays except the gravels (soils with particles >2 mm) in the texture scheme of USDA. Therefore, it is remained unknown that the soil particles >2 mm had any impacts on PSD models.

Szechwan Basin is one of the four biggest basins of China, and it is also called Red Basin, because it holds the greatest concentration of purple soil, the most fertile natural soil in China. Therefore, the objectives of this study are (1) to propose and validate various PSD functions and interpolations for soils in Szechwan Basin; (2) to compare the performance of the PSD models and interpolations; (3) to explore the effects of soil texture (the size ranges) on the performance of the methods.

Data and methodology

Data used

Particle-size data used in this research focused on cultivated soil and the data were extracted from the Second National Soil Survey of China conducted in the 1980s in Sichuan basin. The fraction data were classified by Katschinski's schemes^[10]. Totally, there were 265 soil horizons from 32 soil series, including the representative soil types in Sichuan Basin: yellow soil (classified as Ferrallisols in USDA), purple soil (classified as Entisols in USDA), paddy soil (classified as Inceptisol/ Ultisol in USDA) and limestone soil (classified as calcareous soil in USDA). Value ranges of the measured particle-size distributions are presented in Table 1. The particle-size distribution of the soils encompasses a wide range of soil textures (Table 1). And these distributions, containing percentages of particles finer than 0.001, 0.005, 0.01, 0.05, 0.25, 1.0 and 3.0 mm, were used to test the suitability of PSD models studied in the current study.

Table 1 Value ranges of measured cumulative PSD of 265 horizons from 32 soil series.

Particle-size fraction mm	Mean %	Standard deviation %	Minimum %	Maximum %
<0.001	19.64	10.88	1.52	67.37
<0.005	38.47	14.74	2.80	74.68
<0.01	49.33	17.26	3.51	87.50
<0.05	72.55	21.52	4.65	100.00
<0.25	86.90	17.25	18.11	100.00
<1	92.60	15.24	23.69	100.00
<3	94.69	12.10	30.42	100.00

Table 2 Value ranges of measured cumulative PSD of 49 horizons from 14 soil series.

Particle-size fraction mm	Mean %	Standard deviation %	Minimum %	Maximum %
<0.001	21.48	10.76	1.01	42.41
<0.002	29.33	12.85	2.88	50.58
<0.005	40.40	15.39	8.05	71.50
<0.02	59.00	15.85	31.51	83.04
<0.05	73.52	14.80	47.46	92.75
<0.2	87.11	11.43	50.15	99.52
<0.25	90.15	9.52	59.08	99.99
<1	99.97	0.15	99.09	100.00
<2	100.00	0.00	99.99	100.00

Another set of data (49 horizons of 11 soil series from Sichuan basin), containing percentages of particles <0.001, 0.002, 0.005, 0.02, 0.05, 0.2, 0.25, 1.0 and 2mm was then used to validate the

studied PSD models. Table 2 gave the value ranges of the measured PSD used for the validation. This set of data contained all the possible size ranges accepted by the Katschinski and USDA systems. Percentages of particles <0.002 and 0.02 mm were validated respectively.

Methods of PSD transformation

Methods used in the current study included empirical functions and interpolation methods. In the present work, cumulative number of particles was related to diameter as most researchers did [1,4].

Table 3 listed twelve PSD functions needing to be tested in this work. Models with different number of fitting parameters and various structures were proposed and tested. Models 1-5 were proposed by published works whereas models 6-12 were the equations embedded in OriginPro 8.5. We assumed that the PSD follows the corresponding functions in Table 3. This approach was based mainly on the similarity between shapes of the PSD and the curves of these models. To find the values of the fitting parameters, all models were fit to experimental PSD data (Table 1) using OriginPro 8.5 as an iterative nonlinear regression procedure.

Commonly used interpolations including linear interpolation, cubic spline interpolation, piecewise cubic Hermite interpolation [8] were also adopted in this study. The interpolations were carried out in Matlab 7.12.

Table 3 PSD models used in this research (x was diameter (mm)).

Model No.	Model name	Model	Parameters
1	Weibull (W, modified from Hwang, 2004)	$F(D) = 100 * (c + (1 - c) * (1 - \exp(-a * D^b)))$ where $D = (x - x_{min}) / (x_{max} - x_{min})$	a, b, c ($x_{max}=3$ mm, $x_{min}=0.001$ mm)
2	Logarithm (L, Zhuang et al., 2001)	$F(x) = a * \ln(x) + b$	a, b
3	Exponential (E, Gimenez et al., 2001)	$F(x) = c * x^{-b}$	c, b
4	Closed-form exponential (F1, Rousseva, 1997)	$F(d) = a + (100 - a)d^n (1 + b * \exp(m * d))^{-1}$ where $d = \log_{10}(x * 1000)$	a, b, m, n
5	Closed-form power (F2, Rousseva, 1997)	$F(d) = a + (100 - a)d^n (1 + b * (d^m))^{-1}$ where $d = \log_{10}(x * 1000)$	a, b, m, n
6	Exponential growth (ExpG)	$F(x) = a * \exp(\frac{x}{t}) + b$	a, b, t
7	Logistic (variant of the Hill equation (Barlowa & Blake, 1989))	$F(x) = a + \frac{b - a}{1 + (x/c)^p}$	a, b, c, p
8	Exponential associate (ExpA)	$F(x) = a * (1 - \exp(\frac{-x}{t})) + b * (1 - \exp(\frac{-x}{m})) + c$	a, b, c, t, m
9	Exponential decay (ExpD)	$F(x) = a * \exp(\frac{-x}{t}) + b * \exp(\frac{-x}{m}) + c$	a, b, c, t, m
10	Three-parameter exponential (Exp3P)	$F(x) = \exp(a + \frac{b}{x + c})$	a, b, c
11	Rational (R4)	$F(x) = c + \frac{b}{x + a}$	a, b, c
12	Exponential within OriginPro8.5 (Exp)	$F(x) = b + a * \exp(c * x)$	a, b, c

Model responses exploration

The adjusted coefficient of determination (Adj.R²) was used as a relative measure of the goodness-of-fit of the proposed PSD models for each soil horizon in Table 1. In addition, Akaike's information criterion (AIC)^[11] that imposes penalties for additional fitting parameters was adopted to compare the quality of model fits. A model with a lower value of AIC is considered to be a better model (Akaike, 1974). Correlation coefficient (r), absolute error (AE, %), and root mean square

deviation (RMSD)^[12] were used to test the validation of the PSD models and interpolations on the soil particles <0.002 and 0.02mm based on the soils in Table 2. Correlation coefficients were calculated to examine the relationships among measured and estimated PSD by the proposed methods.

Results and discussion

Identification of PSD models

Fig.1 identified the suitability of the tested models to estimate percentages of particles < 0.001, 0.005, 0.01, 0.05, 0.25, 1.0 and 3.0 mm. Values of Adj.R² for the PSD models tested for all soils ranged from 0.65 to 1.0. Models with two parameters (L and E) and model F2 (with four parameters) showed lowest Adj.R² values, indicating that they were less accurate than others. Logistic and F1 models with four parameters performed similarly better, and ExpA and ExpD models with five parameters had the highest Adj.R² values. The model of W fitted relatively well, as well as Exp3P model. The performances of models of ExpG, R4 and Exp were similar and showed relatively bad fitting. However, results of AIC and Adj.R² were not always identical. For example, the F2 model revealed relatively bad fitting in Adj.R², but its performance as determined by AIC was good. The ExpA model and ExpD model with five parameters showed relatively better performance than other models due to AIC values. Overall, the ExpA and ExpD models with five fitting parameters showed better performance, according to the Adj.R² and AIC values.

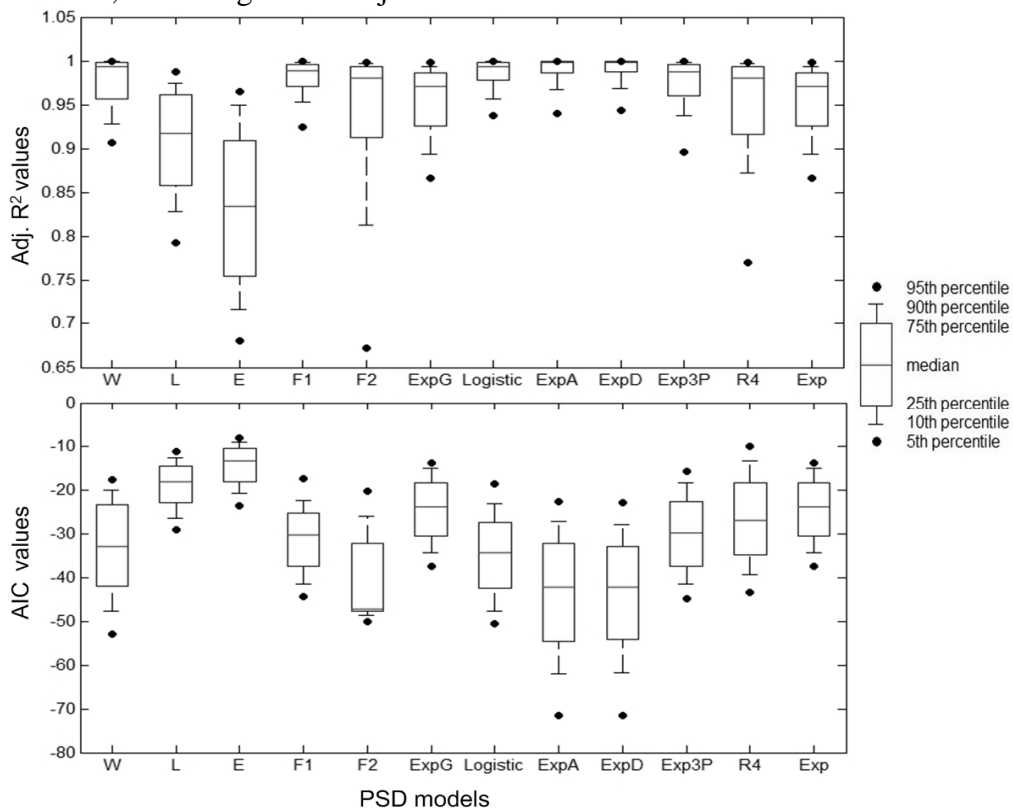


Fig.1 Box plot for Adj.R² and AIC percentiles as criteria for assessing the performance of twelve models for soils in Table 1.

Impact of soil texture on PSD models

Fig.2 identifies the suitability of the studied models to predict percentages of particles <0.001, 0.005, 0.01, 0.05, 0.25, 1 and 3 mm. And correlations between estimated and measured values were also shown. Obviously, models of ExpD and ExpA and Logistic model described cumulative PSD of diverse texture better, while L model and E model overestimated the percentages of particles <0.001, 3 mm and underestimated particles <0.05, 0.25mm, respectively. W model overestimated the percentages of particles <0.005, 1 mm but underestimated particles <0.05, 0.25 mm. ExpG and Exp models overestimated the percentages of particles <0.001, 0.05, 0.25 mm but underestimated those <0.01, 3 mm. F2 model overestimated the percentage of particles <3 mm and F1 model overestimated

a little that <1 mm. Other models of Logistic, R4 and Exp3P described cumulative PSD of all the texture relatively well. The performance of a PSD model may vary with particle sizes of soils [4]. In this study, only ExpA, ExpD, Exp3P, F1 and R4 models showed better fitting at higher clay (<0.001 mm) content (20%), indicating that these models should be used for soils with clay (<0.001 mm) content over 20%.

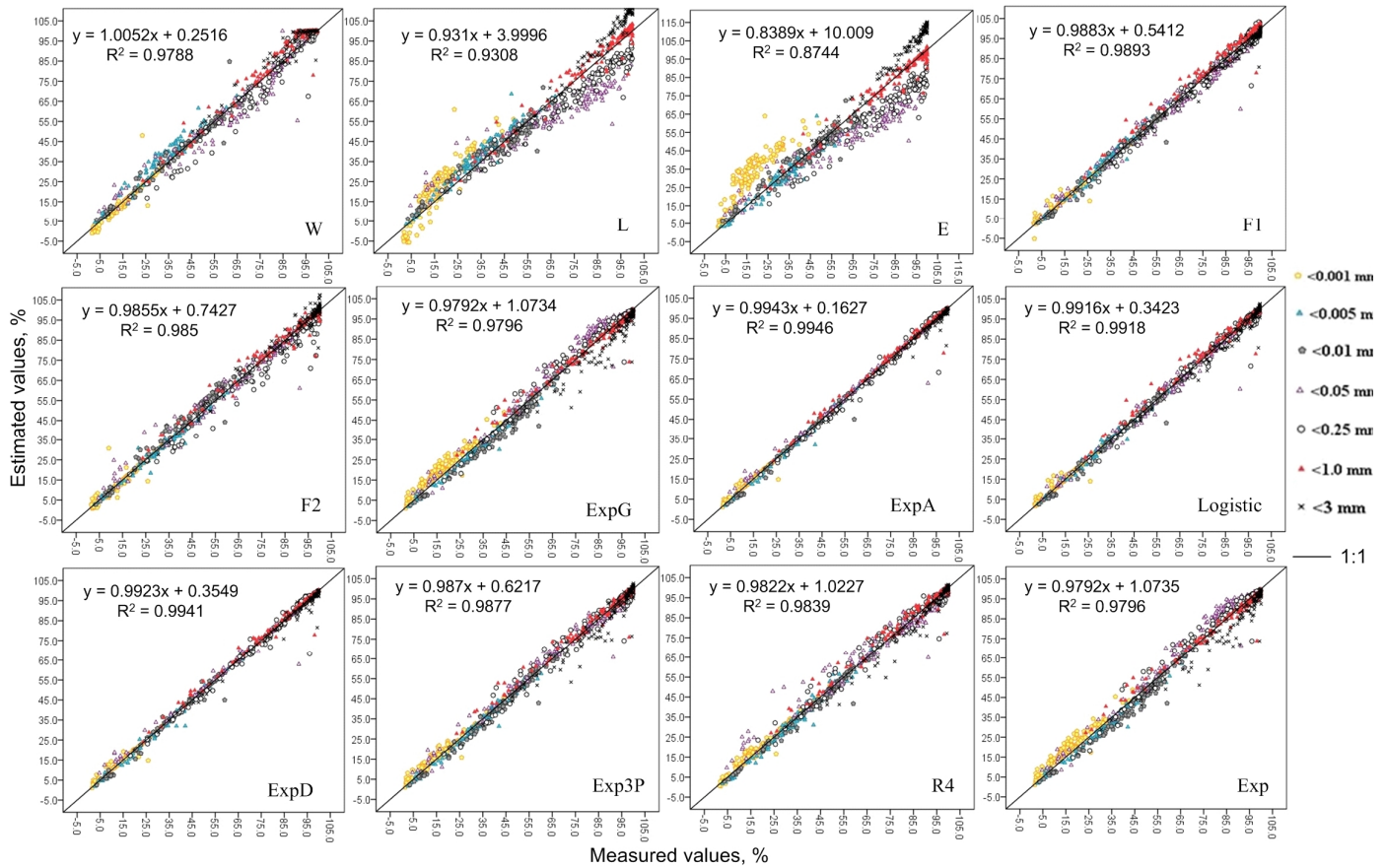


Fig.2 Percentages estimated by various models used in this study VS. measured percentages of particles <0.001, 0.005, 0.01, 0.05, 0.25, 1, 3 mm with soils in Sichuan basin.

Table 4 showed the results of the suitability of studied models and interpolations based on soils in Table 2. The discrepancy in terms of RMSD between the measurements and the estimates derived from studied methods was in the range of 0.02-0.25 for particles <0.002mm and 0.03-0.97 for particles <0.02mm, respectively. And it was noteworthy that the RMSDs yielded from interpolations were much larger than models in Table 3 both at particles <0.002 and 0.02mm. While the correlation coefficient (r) between measured and estimated mass percentage changed between 0.913 (F2 model) and 0.990 (R4 model) for particles <0.002mm and between 0.88 (Cubic interpolation) and 0.987 (W model) for particles finer than 0.02 mm, respectively. Mean absolute error (MAE, %) of studied methods at particles finer than 0.002 mm varied from 1.58 % (Logistic model) to 7.53% (E model) and MAEs at particles finer than 0.02 mm varied from 2.09% (W model) to 21.11% (Cubic interpolation). All the three statistic error indices suggested that the interpolations did not show advantage, especially Cubic linear interpolation, in estimating PSDs for soils in Sichuan Basin.

Fig.3 showed the percentiles of absolute error (AE, %) at particles <0.002 and 0.02 mm, respectively. Models of Logistic, ExpA and ExpD showed smaller AE at particles <0.002 mm, <6%. Whereas models of E and F2 indicated much larger AE, maximum values of AE were about 14%. At particles <0.002 mm, interpolations did not show any superiority towards functions. Whereas at particles <0.02 mm, piecewise cubic Hermite interpolation had lowest AE, followed by models of W, Logistic, F1, ExpA and ExpD.

Table 4 Validation results of the studied models at particles <0.002 and 0.02mm.

	Diameter (mm)	Models														
		W	L	E	F1	F2	ExpG	Logistic	ExpA	ExpD	Exp3P	R4	Exp	Linear	Cubic	Hermite
MAE	<0.002	2.91	2.43	7.53	2.32	5.04	3.45	1.58	1.99	1.99	2.95	2.54	3.45	3.27	2.96	2.65
	<0.02	2.09	5.22	5.99	2.41	3.84	5.93	2.17	2.79	2.80	3.66	6.34	5.93	7.61	21.11	2.56
RMSD	<0.002	0.04	0.03	0.08	0.03	0.06	0.04	0.02	0.02	0.02	0.04	0.03	0.04	0.23	0.24	0.25
	<0.02	0.03	0.06	0.07	0.03	0.05	0.07	0.03	0.04	0.04	0.05	0.08	0.07	0.49	0.97	0.61
r	<0.002	0.98	0.99	0.97	0.97	0.91	0.95	0.99	0.99	0.99	0.97	0.99	0.95	0.97	0.97	0.98
	<0.02	0.99	0.97	0.97	0.98	0.96	0.98	0.99	0.97	0.97	0.98	0.95	0.98	0.98	0.88	0.98

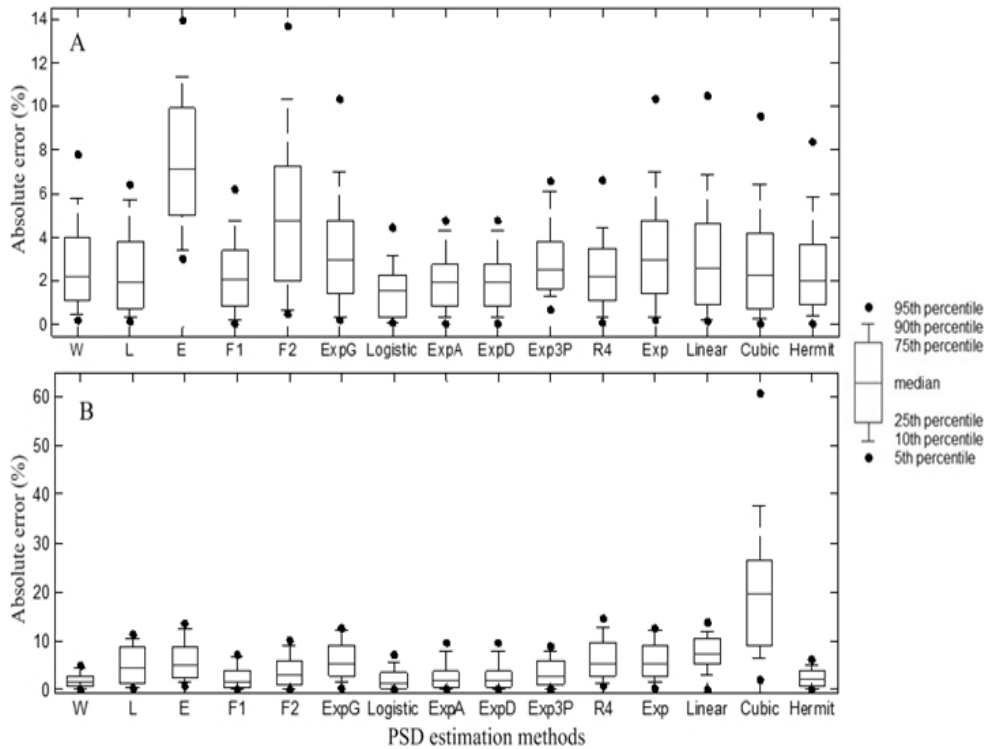


Fig.3 Box plot for absolute error (%) percentiles as criterions for assessing the performance of the studied methods for soils in Table 2 at particles < 0.002 mm (A) and <0.02 mm (B), respectively.

Conclusions

Twelve numerical models were proposed as the function of cumulative mass percentage of particles and diameter (mm) from Soil Survey of China in Sichuan basin, to estimate 0.002, 0.02 and 2.0 mm (convert the Katchinski’s scheme to USDA system). The ExpA model and ExpD model with five parameters, as well as Logistic model with four parameters, showed relatively better performance according to Adj.R² and AIC values. By comparing the measured and estimated percentages of particles <0.001, 0.005, 0.01, 0.05, 0.25, 1 and 3 mm, the proposed models of ExpD, ExpA and Logistic described cumulative PSD of diverse texture better.

Three commonly used interpolations were tested together with the twelve models. In terms of RMSD, MAE, and correlation coefficient (r), studied models and interpolations had been tested at particles <0.002mm and 0.02mm, respectively. And it was noteworthy that the RMSDs yielded from interpolations were much larger than models both at particles <0.002mm 0.02mm. Interpolations did not show any superiority towards numerical functions at particles <0.002mm. Cubic spline interpolation performed worst with MAE of 20% at particles <0.02 mm. This result suggested that the cubic spline interpolation could not use to estimate the soil particles <0.02 mm for Sichuan basin.

Acknowledgements

This work was financially supported by Hainan special project of research and demonstration on Farmland improvement key technology (HNGDzy201503) and Central Public-interest Scientific Institution Basal Research Fund for Chinese Academy of Tropical Agricultural Sciences (No. 1630072017004).

References

- [1] Rousseva S.S., (1997) Data transformations between soil texture schemes. *European Journal of Soil Science*, 48, 749-758.
- [2] Zobeck T.M., Gill T.E., Popham T.W., (1999) A two-parameter Weibull function to describe airborne dust particle size distribution. *Earth Surface and Land forms*. 24,943-955.
- [3] Zobeck T.M., Popham T.W., Skidmore E.L., Lamb J.A., Merrill S.D., Lindstrom M.J., Mokma D.L., Yoder R.E., (2003) Aggregate-mean diameter and wind-erodible soil predictions using dry Aggregate-size distribution. *Soil Sci. Soc. Am. J.*. 67,425-436.
- [4] Hwang S.I., (2004) Effect of texture on the performance of soil particle-size distribution models. *Geoderma*, 123, 363-371.
- [5] 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.
- [6] Gimenez, D., Rawls, W.J., Pachepsky, Y., Watt, J.P.C., (2001) Prediction of a pore distribution factor from soil textural and mechanical parameters. *Soil Sci.* 166, 79–88.
- [7] Cai, Y., Zhang, K., Li, S., (2003) Study on the conversion of different soils texture. *Acta pedologica sinica*. 40(3), 511-517.
- [8] Xie Y., Chen X., Wang Z., Zhang Q., (2009) Optimum interpolation technique for soil texture conversion. *Journal of Irrigation and Drainage*. 28(3), 50-57. (In chinese)
- [9] Cai, Y., Zhang, K., Li, S., (2003) Study on the conversion of different soils texture. *Acta pedologica sinica*. 40(3), 511-517.
- [10] Katschinski, N.A. (1956) Die Mechanische Bodenanalyse und die Klassifikation der Böden nach ihrer Mechanischen Zusammensetzung. *Rapports au Sixième Congrès International dela Science du Sol*, Paris, B, 321-327..
- [11] Akaike, H. (1973) Information theory and an extension of the maximum likelihood principle. In: Petrox, B., Csáki, F. (Eds.), *Symposium on Information Theory*, Akadémiai Kiadó, Budapest, Hungary, 267 pp.
- [12] Braga, R.P., Cardoso, M.J., Coelho, J.P. (2008). Crop model based decision support for maize (*Zea mays L.*) silage production in Portugal. *Europ. J. Agronomy* 28, 224–233.