IOP Conf. Series: Materials Science and Engineering **392** (2018) 062104 doi:10.1088/1757-899X/392/6/062104

Study on interpolation of soil grain-size sparse distribution

Zhang Yuxin[1] and Lu Yudong[1]

[1] School of Environmental Science and Engineering, Chang'an University, 126, Yanta Road, Xi'an 710054, Shaan Xi, China

52816880@qq.com

Abstract. This study compares the cubic spline interpolation method and several kinds of cumulative distribution curve fitting equations of soil particles in order to reduce the calculation error of the permeability caused by the interpolation error of the grain-size distribution to a certain extent. The soil samples are selected from UNSODA database for interpolation verification, and the result shows the reliability of the interpolation conversion to a some degree. In addition, methods have different adaptability to various soil textures, and the appropriate one can be selected through statistical methods. So interpolation methods are selected for different soil textures of HWSD in the research area. And the known particle size distribution of the interpolation samples has a limited influence on the interpolation results.

1. Introduction

In the SWAT (Soil and Water Assessment Tool) modeling process, the soil database may be built using the HWSD (Harmonized World Soil Database) soil database. Since Chinese soil texture classification(0.002 mm, 0.0625 mm, and 2 mm) is FAO90 in HWSD[1], while the soil texture classification is USDA(0.002 mm, 0.05 mm, and 2 mm) for the tool SPAW Hydrology[2] used to calculate soil parameters in the SWAT soil database, most studies use cubic spline interpolation method to obtain the final grain-size distribution. However, permeability is one of the important soil parameters. Due to its sensitivity to grain-size distribution, it is required to select interpolation methods carefully.

There are many researches on soil grain-size distribution model, which obtain satisfactory results[3-8]. If the interpolation results obtained by this method are better than the cubic spline interpolation method, the accuracy of soil parameters can be improved. In order to verify the suitability of different methods for different soil textures, based on the soil particle data from the UNSODA, the methods are separately used to interpolate.

2. Necessity of the research

The sensitivity of saturated permeability of HWSD soil in the research area is examined within the range of (-0.04, 0.04) through taking the cubic spline interpolation results as the center and 0.01 steps. The result is shown in figure 1. The data in the figure show the difference of saturated permeability values calculated by SPAW Hydrology when the particle diameter difference of <50 μm is 1%. In addition, the sensitivity of sand and sandy loam is greater than that of other soil textures. Therefore, the interpolation method must be carefully selected.

3. Methods

The methods to be compared in the study include cubic spline interpolation and interpolation methods of curve fitting equations. There are many kinds of curve fitting equations, which are briefly introduced in this section.

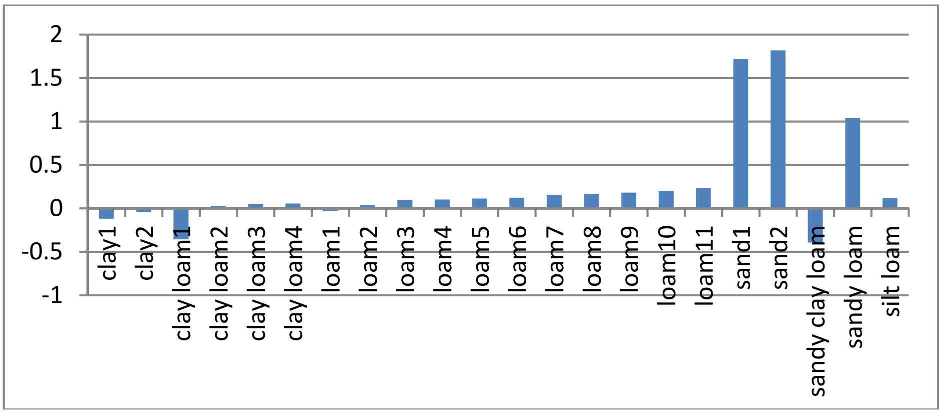

Figure 1. Saturated permeability sensitivity of 50 μm grain-size distribution in HWSD soil in the research area.

3.1. Cubic Spline Interpolation

The uniqueness of the cubic spline solution is mathematically well-proven, and only a brief description is given here.

$$S(X)=\begin{cases} S_1(x) & t_1 \le x \le t_2 \\ S_2(x) & t_2 \le x \le t_3 \\ \cdots & \cdots\cdots \\ S_{n-1}(x) & t_{n-1} \le x \le t_n \end{cases} \tag{1}$$

In equation (1), S_i denotes the cubic polynomial that will be used on the subinterval [t_i, t_{i+1}]. Since there are n nodes, n-1 subintervals, and different cubic polynomials in each subinterval. Each cubic polynomial has 4 coefficients, so there are 4(n-1) coefficients to be determined. Since each polynomial passes through two nodes, this gives 2(n-1) conditions. Again, because the first and second derivatives of the cubic spline function must be continuous over n-2 nodes, this gives 2(n-2) conditions, and there are still 2 constraints. To increase these two constraints, the most common method is also the one used by the function spline, using non-a-knot conditions. This condition forces the third derivatives of the first and second cubic polynomials to be equal. Do the same for the last and the penultimate cubic polynomial. So there are 2(n-1) +2(n-2) +2=4(n-1) conditions. Thus, a cubic spline solution is obtained[9]. Cubic spline interpolation is completed on Matlab.

3.2. Particle Curve Fitting Equation

The single parameter particle curve fitting equations and two-parameter parameter particle curve fitting equations are listed in table 1. Particle curve fitting equation interpolation is completed on Origin.

3.3. Method to compare the interpolations

The sample data of the known particle distribution is left with cumulative percentage data of 2 μm, 50 μm, x μm, and 2000 μm particles, where x is the smallest in grain-sizes greater than 50 μm. And the soil sample data of 2 μm, x μm, and 2000 μm is used for interpolating or fitting, the data of 50 μm is used to compare with the interpolated results. After calculating interpolated results of 50 μm through these methods for the soil samples, method effectiveness is examined by the absolute errors and relative errors of interpolations.

3.4. Method to select interpolations

The errors of various methods are classified according to the soil textures, and minimum, maximum, average, and standard deviation of the absolute or relative errors of various methods are calculated. And then, the method, of which average and standard deviation values of the error are least, is selected. If the two values of different methods are identical, the minimum and maximum values are examined and select the method with the least values as an interpolation method for the soil texture class.

Table 1. Table of particle curve fitting equations

Equation Type	Model Name	Particle Curve Fitting Equation	Parameters[a]
Single parameter model	Jaky[6, 8]	$F(d) = \exp\left[\frac{1}{p^2}\ln\left(\frac{d}{d_0}\right)^2\right]$	d_0 is often taken as 1.00; p is a parameter.
	S1P[8]	$F(d) = \frac{1}{1+\left(\frac{1}{F_0}-1\right)\mathrm{e}^{-kd}}$	k is the parameter.
Two-parameter model	LS[8]	$F(d) = F_0 + \frac{1-F_0}{\left[1+\left(\frac{\delta}{d}\right)^n\right]^{1-\frac{1}{n}}}$	δ and n are parameters.
	Weibull8]	$F(d) = 1-(1-F_0)\exp[-(kd)^n]$	k and n are parameters.
	Morgan[8]	$F(d) = 1-\frac{1-F_0}{1+(kd)^n}$	k and n are parameters.
	HP[4, 8]	$F(d) = \frac{1}{\left[1+\left(\frac{d_g}{d}\right)^n\right]^{1-\frac{1}{n}}}$	d_g and n are parameters.
	S2P[8]	$F(d) = \frac{1}{1+\left(\frac{1}{F_0}-1\right)\mathrm{e}^{-kd^n}}$	k and n are parameters.
	MVG[5, 7]	$F(d) = F_0 + \frac{1-F_0}{(1+(ad)^b)^{1-\frac{1}{b}}}$	a and b are parameters.

[a] *F(d)* is the cumulative mass percentage of soil particles with grain-size *d*; *d* is the grain-size of the soil; F_0 is the minimum cumulative mass percentage of the fraction.

4. Data

4.1. UNSODA

UNSODA (Unsaturated Soil Hydraulic Property Database) is a database of unsaturated soil hydraulic properties completed by U.S. National Risk Management Research Laboratory in 1996[10]. It collects data on soil physical properties such as water characteristic curves, hydraulic conductivity and soil water diffusion, grain-size distribution, bulk density, and organic matter content in 11 different textures soils from sand to clay. Of course, some are incomplete.

4.2. Soil samples

The grain-size data table in UNSODA stores a total of 790 soil sample particle data, including some samples with the same data. And research data is obtained from this data sheet. Select samples criteria:

1. There are 2 μm, 50 μm and 2000 μm sub-grades;
2. There are particle grades between 50 μm and 2000 μm.

Results of sample selecting are shown in table 2, and 244 soil samples are selected without duplicates. The table lists 11 textures of the soil samples and 6 smallest particle grades greater than 50 *μm*, since such data of samples are interpolation data and may affect the accuracy of interpolation methods.

Table 2. Table of selected soil samples

Soil Texture	Minimum grain fraction of soil texture greater than 50 *μm*						Total
	75 *μm*	100 *μm*	105 *μm*	106 *μm*	125 *μm*	200 *μm*	
Clay		2	2				4
Clay loam		4		1	1		6
Loam		12			2		14
Loamy sand	1	24	1	5	9		40
Sand	13	46		1	21	2	83
Sandy clay					2		2
Sandy clay loam	1	1	3	1	13	1	20
Sandy loam	1	15	1	2	10	1	30
Silt		1					1
Silt loam		28	11				39
Silty clay loam	1	2	2				5
Total	17	135	20	10	58	4	244

5. Results and Discussion

Through calculating, results of 9 interpolation methods for 244 soil samples are obtained. And the absolute error and relative error between each result and the actual value are calculated. In order to show the distribution, the minimum absolute error and the minimum relative error of each sample are drawn to a scatter plot, as shown in figure 2, and the quantitative proportion of the minimum absolute error <0.02 reached 70% and the minimum relative error <0.085 reached 63.9%, as shown in table 3.

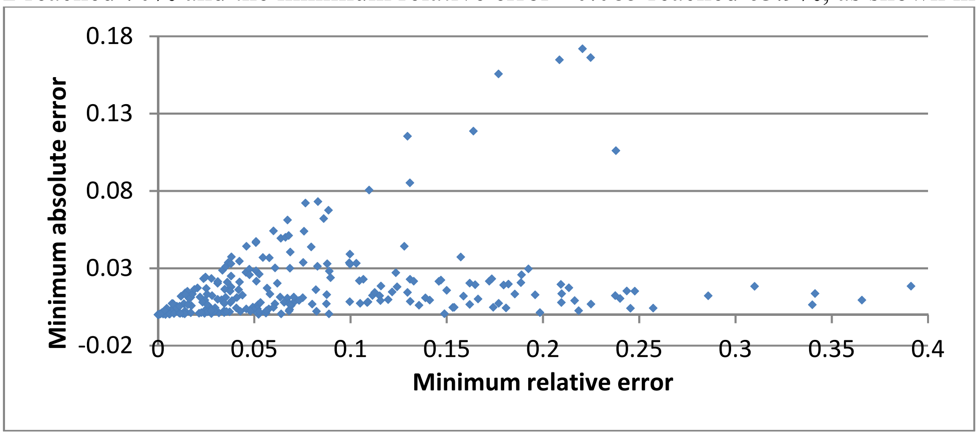

Figure 2. Distribution of minimum relative errors and minimum absolute errors.

Table 3. Statistics table of distribution of minimum relative errors and minimum absolute errors

Minimum relative error	Minimum absolute error	Total	Percentage
<0.085	<0.02	116	47.54%
>0.085	<0.02	55	22.54%
<0.085	>0.02	40	16.39%
>0.085	>0.02	33	13.52%
Total		244	100%

Selected interpolation methods for different soil textures are based on the statistical results of absolute errors or relative errors. The results of absolute and relative errors are close to each other. In practice, the statistics of

IOP Conf. Series: Materials Science and Engineering **392** (2018) 062104 doi:10.1088/1757-899X/392/6/062104

the errors of two kinds are calculated. Here only the statistical results of absolute errors are shown. The evaluation standard of curve fitting degree usually uses related coefficient (R^2). However, this study uses the curve fitting equation to obtain the interpolation, the amount of data used is very small, resulting in R^2 being too large and limiting its reference value. Therefore, R^2 is not applied to analyse the interpolation results. Here, the minimum, maximum, mean, and standard deviation of absolute errors are used as selection criteria. The minimum and maximum values reflect the range of absolute errors. The average value reflects the degree to which the overall value is close to the true value. Deviation reflects the stability of the method. The smaller four statistical values are, the better the method.

Select the interpolation method for the study area HWSD soil type. First, select soil samples according to the texture type, a total of 7 kinds, and classify them. Then, calculate respectively the minimum, maximum, average and standard deviation of the absolute errors of various methods of soil texture. Finally, the method with the smallest average and standard deviation of the absolute error is selected. If there are methods with the same value, examine the minimum and maximum values and select the smallest method as the interpolation method for the soil texture type. Line charts of the calculation data and the final selected methods for various texture types are shown in figure 3.

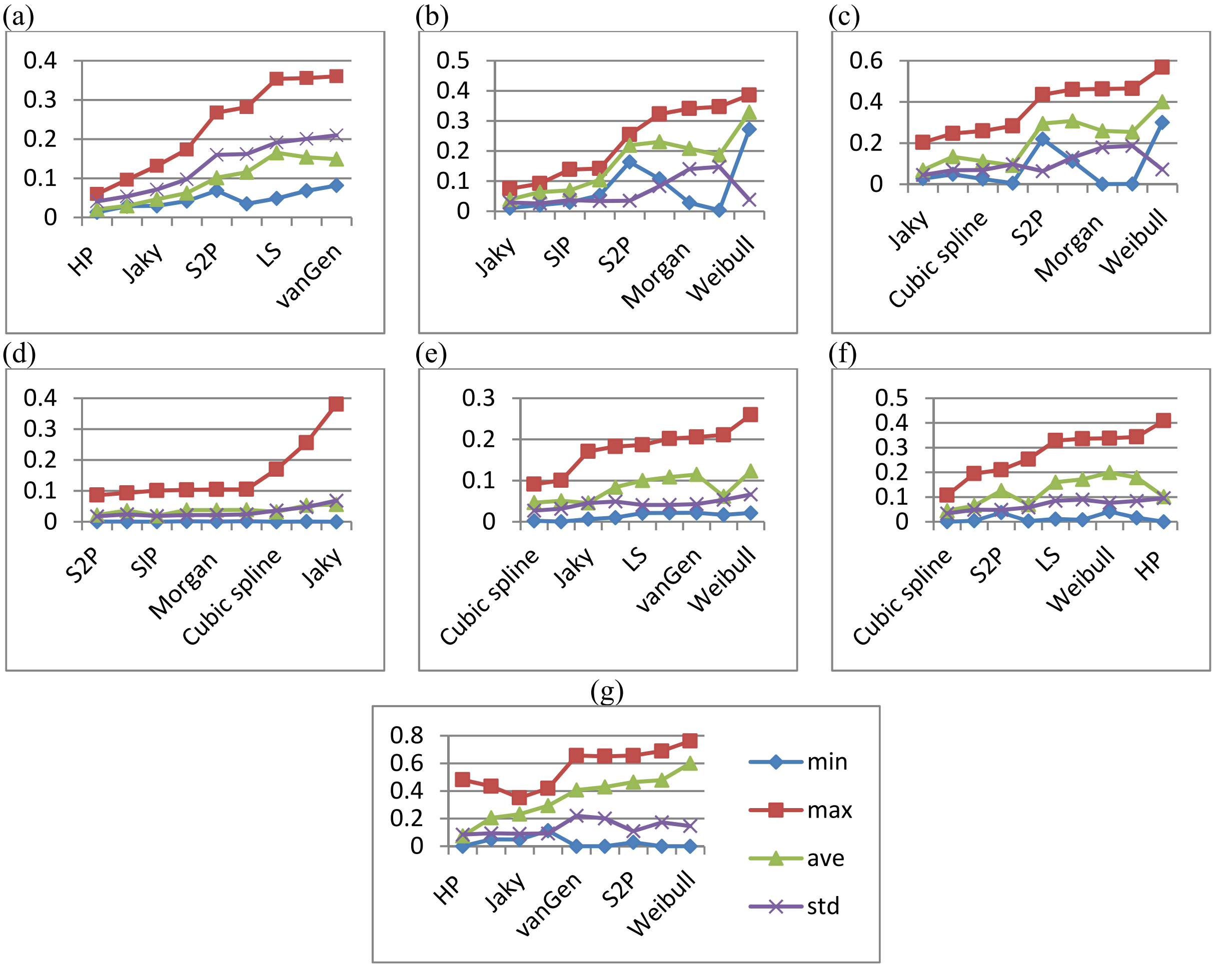

Figure 3. Line charts of statistical data of interpolation absolute errors for textures.
(a) Line chart for clay, HP is the selected method.
(b) Line chart for clay loam, Jaky is the selected method.
(c) Line chart for loam, Jaky is the selected method.
(d) Line chart for sand, S2P is the selected method.
(e) Line chart for sandy clay loam, Cubic spline is the selected method.
(f) Line chart for sandy loam, Cubic spline is the selected method.
(g) Line chart for silt loam, HP is the selected method.

In addition, the x-values of these soil samples data are different as shown in table 1, and the effect on the interpolation results is shown by making a scatter plot. As is shown in figure 4, the value of x ranges from 75 μm to 200 μm. Regardless of x, there are soil samples whose interpolation results are closer to the actual measured values. Therefore, it can be guessed that the influence of x on the interpolation result is limited.

6. Conclusions

1. Saturation permeability through SPAW Hydrology calculation is sensitive to soil particle data, so it is necessary to carefully select the interpolation method when the soil texture classification is converted.

2. Based on the UNSODA soil samples, 9 interpolation methods were performed. The results show that the quantitative proportion of the minimum absolute error <0.02 reached 70% and the minimum relative error <0.085 reached 63.9%. To a certain extent, the reliability of the interpolation conversion results is demonstrated.

3. Diverse methods have different adaptability to various soil textures, and the appropriate one can be selected through statistical methods.

4. The known particle size distribution of the samples has limited influence on the interpolation results.

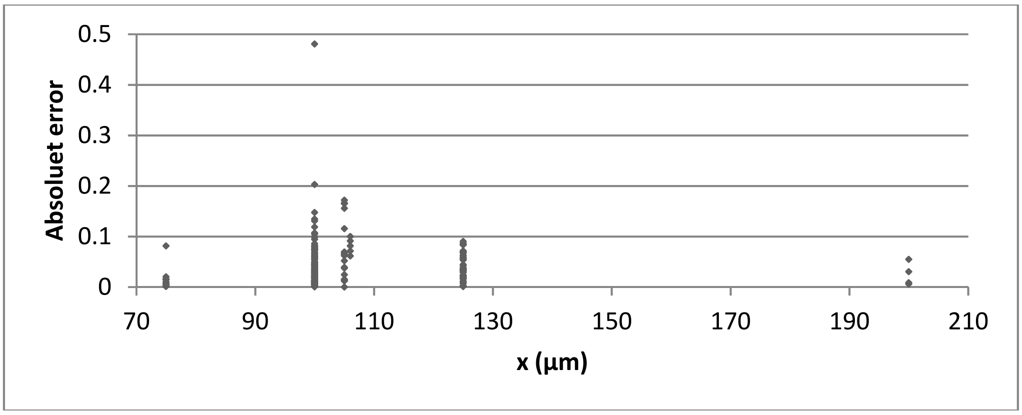

Figure 4. Distribution of absolute error values on x.

Acknowledgments

This work was financially supported by the fund of Key projects of National Natural Science Foundation of China (Grant No. 41630634)

References

[1] Nachtergaele F, Velthuizen H V, Verelst L and Wiberg D 2012 *Harmonized World Soil Database (version 1.2)* p 14

[2] SPAW Hydrology. Available: http://hydrolab.arsusda.gov/SPAW/index.htm

[3] Fredlund M D, Wilson G W, and Fredlund D G 2000 An equation to represent grain-size distribution *Canadian Geotechnical Journal* **37** 817-27

[4] Haverkamp, Parlange, and Y. J 1986 Predicting the Water-Retention Curve From Particle-Size Distribution: 1. Sandy Soils Without Organic Matter1 *Soil Science* **142** 325-39

[5] Jie Z, Yan J, and Tsuyoshi M 2001 Estimating Water Retention Characteristic From Soil Particle-Size Distribution Using A Non-Similar Media Concept *Soil Science* **166** 308-21

[6] Sang I H, Lee K P, Dong S L, and Powers S E 2002 Models for Estimating Soil Particle-Size Distributions *Soil Science Society of America Journal* **66** 1143-50

[7] Zhao Lixiao 2007 Research on the Prediction Model of Soil Water Characteristic Curve (Nanjin: Hohai University) p 27

IOP Conf. Series: Materials Science and Engineering **392** (2018) 062104 doi:10.1088/1757-899X/392/6/062104

[8] Du Zewen 2016 Calculation Methods of Soil-water Characteristics Curve (Xi'an: Chang'an University) p 26-27
[9] Cai Yongming, Zhang Keli and Li Shuangcai1 2003 Study On The Conversion Of Different Soils Texture *Acta Pedologica Sinica* **40(4)** 511-7
[10] Nemes A, Schaap M, and Leij F The UNSODA Unsaturated Soil Hydraulic Property Database. Version 2.0. Available: http://www.ars.usda.gov/research/software/