Evaluation of the Gray Model GM(1,1) Applied to Soil Particle Distribution

Cheng-Mau Wu
Research Center for Soil and Water Resources and Natural Disaster Prevention
National Yunlin Univ. of Science and Tech.
123, Section 3, University Rd.
Douliou, Yunlin County 64002, Taiwan, R.O.C.

Jet-Chau Wen*
Dep. of Safety Health and Environ. Engineering
National Yunlin Univ. of Science and Tech.
123, Section 3, University Rd.
Douliou, Yunlin County 64002, Taiwan, R.O.C.

Kou-Chiang Chang
Water Resources Agency
Ministry of Economic Affairs
9-12F, 41-3, Section 3, Hsin-Yi Rd.
Taipei 10651, Taiwan, R.O.C.

Particle size distribution (PSD) is a fundamental soil physical property. The conventional approaches for representing PSD use empirical models with two to four parameters. We developed an alternative way to predict PSD that differs from conventional approaches by using the gray model GM(1,1), which does not depend on the model shape as empirical approaches do. The performance of GM(1,1) was compared with Skaggs model by using four statistical criteria. From nine textures of soil samples in our study, the results reveal that the GM(1,1) is superior for making PSD predictions. The results show that for the overall textures, the GM(1,1) model makes better predictions than the Skaggs model except for sand. Therefore, the performance of the GM(1,1) is fairly reliable and efficient and is not affected by soil textures in general.

Abbreviations: AAE, accumulative absolute error; AGO, accumulated generating operator; CMF, cumulative mass fraction; Cc, curvature coefficient; Cu, uniformity coefficient; GM, gray model; MAPE, mean absolute percentage error; PSD, particle size distribution.

The PSD of soil is one of the fundamental soil physical properties. Because PSDs can be measured relatively easily and quickly, they have been widely regarded as the basis for estimating the hydraulic properties of soil, such as the water retention curve and saturated as well as unsaturated hydraulic conductivities (Gupta and Larson, 1979a,b; Arya and Paris, 1981; Arya et al., 1999; Hwang and Powers, 2003). A conventional particle size analysis involves the measurement of the mass fractions of clay, silt, and sand and the use of these fractions to find the textural class using a textural diagram, commonly in the form of a textural triangle (Gee and Bauder, 1986). A more complete description of a texture is obtained by defining a PSD function. Generally, PSDs are reported as cumulative distributions, and different functions have been proposed to fit experimental data. Several studies suggest that a PSD in soil shows an approximately lognormal distribution (Shirazi and Boersma, 1984; Campbell, 1985; Buchan, 1989). Buchan (1989) studied the applicability of lognormal models for PSDs and found that only half of the soils determined by the USDA textural triangle could be suitably described with the lognormal models. Buchan (1989) also investigated the effects of the number of particle size fractions that were measured on the shape of the cumulative mass fraction (CMF). The more complex the CMF was, the greater was the number of required model parameters. Buchan et al. (1993) compared five different lognormal models for experimental soil PSDs. All five models accounted for >90% of the variance in the PSD of most of the soils examined.

Hwang et al. (2002) evaluated seven parametric models (five lognormal models, the Gompertz model, and the Fredlund model) to find out which model gave a better fit for the PSD. These five lognormal models were previously studied by Buchan et al. (1993). The number of parameters for each model was from one to four. They concluded that the model with more parameters obtained a superior performance.

These researchers were concerned about model performance and accuracy. A general conclusion is that these models require more parameters to be accurate. In some cases, time and labor constraints may limit measurements to only the percentage of sand, silt, and clay (Skaggs et al., 2001). Skaggs et al. (2001) devised a method for estimating the soil PSD when only a few particle sizes are available, using a general logistic model. They presented a simple method in which the distribution was estimated from only the clay, silt, and fine sand mass fractions. This method was easy to use and the estimated PSD agreed reasonably well with the measured data.

Different from the above conventional approaches where PSD is described by an empirical model, the gray model that is free of a fixed model shape provides another aspect for PSD predictions. The gray system theory, originally proposed by Deng (1989), focuses on model uncertainty and information insufficiency in analyzing and understanding systems via research on conditional analysis, forecasting, and decision making (Guo et al., 2005). The gray system treats each variable as a gray quantity that changes within a given range. It differs from the deterministic model in that the gray model requires only a small amount of already known data to forecast future data. It avoids the inherent defects of the

Soil Sci. Soc. Am. J. 73:1775-1785
doi:10.2136/sssaj2007.0323
Received 6 Sept. 2007.
*Corresponding author (wenjc@yuntech.edu.tw).

conventional, large-sample, statistical method to estimate the behavior of a system with insufficient information (Guo, 2004).

The general gray model GM(N,H) uses an Nth order differential equation with H variables and is used to predict random and discrete data. When the first-order differential equation ($N = 1$) with one variable ($H = 1$) is specified, the gray model GM(1,1) is established. The gray model GM(1,1) has been successfully applied to various fields by the following researchers. Yu et al. (2001) applied the gray model to enhance rainfall and runoff relationships and to forecast runoff using a small amount of historical data. Tseng et al. (2001) proved that the GM(1,1) is insufficient for forecasting time series with seasonality and should be deseasonalized first before building a GM(1,1). Hsu (2003) applied the gray model to the global integrated circuit industry and concluded that the gray model is better suited to short-term predictions than mid- and long-term predictions. Mao and Chirwa (2006) used the GM(1,1) to estimate vehicular fatality risks. They applied the GM(1,1) to the United Kingdom and United States vehicular fatality data sets and showed that the GM(1,1) is a feasible, reliable, and highly efficient prediction method. Even though the GM(1,1) has been successfully applied to different fields, we did not find any related studies in soil physics.

In this study, we investigated the use of the GM(1,1) model as an alternative tool to conventional models for predicting PSD. This model was used to predict the PSD, using four statistics to compare the performance of the Skaggs model and the gray model GM(1,1). The relationship between the soil properties (the uniformity coefficient [Cu] and curvature coefficient [Cc]) and the predictive ability of both models were also investigated.

THEORETICAL BACKGROUND

The following briefly describes the theoretical background of both the Skaggs model (Skaggs et al., 2001) and the GM(1,1) model.

Skaggs Model

Skaggs et al. (2001) used the empirical model to describe the cumulative mass fraction of a PSD as

$$P(r)=\frac{1}{1+\left[1/P(r_0)-1\right]\exp\left(-uR^c\right)} \qquad [1]$$

where $R = (r - r_0)/r_0$ with $r \geq r_0 > 0$, $P(r)$ is the mass fraction of the soil particles with radii less than r, r_0 is the lower bound for the radii used in this model, and c and u are the model parameters. Equation [1] was evaluated using 125 soil samples and the following comments were made by Skaggs et al. (2001): (i) the model describes the distribution only for $r > r_0 > 0$, and the value of the distribution at r_0 must be specified, $P(r_0) > 0$; (ii) the model dictates $P(r_2) > P(r_1)$ for any $r_2 > r_1$, which may not be consistent with an exceptionally poorly graded soil; and (iii) the model predicts $P \to 1$ as $r \to \infty$, meaning that it cannot be guaranteed that $P \to 1$ at the upper limit of the soil material as it should.

The two unknown parameters u and c in Eq. [1] can be estimated by using the following expressions:

$$c = \alpha \ln\left(\frac{v}{w}\right) \text{ and } u = -v^{1-\beta}w^{\beta} \qquad [2]$$

where

$$v = \ln\left[\frac{1/P(r_1)-1}{1/P(r_0)-1}\right], \quad w = \ln\left[\frac{1/P(r_2)-1}{1/P(r_0)-1}\right] \qquad [3]$$

and

$$\alpha = \frac{1}{\ln\left[(r_1 - r_0)/(r_2 - r_0)\right]}, \quad \beta = \alpha \ln\left(\frac{r_1 - r_0}{r_0}\right) \qquad [4]$$

$$1 > P(r_2) > P(r_1) > P(r_0) > 0, \quad r_2 > r_1 > r_0 > 0 \qquad [5]$$

Thus we can use Eq. [1] to predict a PSD with parameters u and c being determined entirely by $P(r_0)$, $P(r_1)$, and $P(r_2)$ as specified in Eq. [2–5].

In Eq. [2], it can be seen that u is real by noting that since $v < 0$ and $w < 0$, then

$$u = -v^{1-\beta}w^{\beta} = |v|^{1-\beta}|w|^{\beta} \qquad [6]$$

where u is a positive real number for any real β.

Gray Model GM(1,1)

The gray model GM(N,H), which was formulated by Deng (1989), uses an Nth order differential equation with H variables and is used to predict random and discrete data. Gray predicting is applying the GM(1,1) to predict a characteristic value of the progressive change of a system behavior. Its essence is to consider a stochastic process or a stochastic variable as being gray, then using the GM(1,1) to deal with the progressive changes in these data.

The variable $X^{(0)}$ is a set of original mass fractions:

$$X^{(0)} = \left\{P^{(0)}(r_1), P^{(0)}(r_2), \ldots, P^{(0)}(r_n)\right\} \qquad [7]$$

where $P^{(0)}(r_i)$ is a mass fraction with a particle size r_i, i is an index of particle sequence, and n is the total number of sequences. Mao and Chirwa (2006) suggested that n must be ≥4 for making predictions more accurate. On the basis of the initial sequence $X^{(0)}$, a new sequence $X^{(1)}$ can be set up through an accumulated generating operator (AGO) to provide an intermediate sequence to build a model and to weaken the variation tendency, i.e.,

$$X^{(1)} = \left\{P^{(1)}(r_1), P^{(1)}(r_2), \ldots, P^{(1)}(r_n)\right\} \qquad [8]$$

where $P^{(1)}(r_i) = \sum_{i=1}^{k} P^{(0)}(r_i)$ for $k = 1, 2, \ldots, n$ and $X^{(1)}$ is 1 − AGO of $X^{(0)}$. The first-order differential equation of the gray model GM(1,1) is then

$$\frac{dX^{(1)}}{d\eta} + aX^{(1)} = b \qquad [9]$$

where a and b are parameters. Notice that Eq. [9] is a general differential equation, so the variable η can be either space or time. According to gray theory, the whitening of the gray derivatives for discrete data with a unit space interval ($\Delta\eta = 1$) is given by

$$\left.\frac{dX^{(1)}}{d\eta}\right|_{\eta=k} \approx \frac{P^{(1)}(r_k) - P^{(1)}(r_{k-\Delta\eta})}{\Delta\eta} = P^{(1)}(r_k) - P^{(1)}(r_{k-1}) = P^{(0)}(r_k) \qquad [10]$$

The variable $Z^{(1)}(k)$, which is the whitening value of $X^{(1)}|_{\eta=k}$, is defined as

$$X^{(1)}\big|_{\eta=k} \cong Z^{(1)}(k) = \frac{1}{2}\left[P^{(1)}(r_k) + P^{(1)}(r_{k-1})\right]; \ \forall k = 2,3,\ldots,n \qquad [11]$$

where $Z^{(1)}(k)$ is an average generating sequence of $X^{(1)}$.

By substituting Eq. [10] and [11] into Eq. [9] and writing the equation in a discrete differential form, we obtain

$$X^{(0)}(k) + aZ^{(1)}(k) = b \qquad [12]$$

and from Eq. [12], it is easy to get

$$\begin{bmatrix} P^{(0)}(r_2) \\ P^{(0)}(r_3) \\ \vdots \\ P^{(0)}(r_n) \end{bmatrix} = \begin{bmatrix} -Z^{(1)}(2) & 1 \\ -Z^{(1)}(3) & 1 \\ \vdots & \vdots \\ -Z^{(1)}(n) & 1 \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} \quad [13]$$

where a and b are the coefficients to be defined. We specify the matrices as

$$\mathbf{Y}_n = \begin{bmatrix} P^{(0)}(r_2) \\ \vdots \\ P^{(0)}(r_n) \end{bmatrix}, \ \mathbf{B} = \begin{bmatrix} -Z^{(1)}(2) & 1 \\ \vdots & \vdots \\ -Z^{(1)}(n) & 1 \end{bmatrix}, \text{ and } \mathbf{A} = \begin{bmatrix} a \\ b \end{bmatrix} \quad [14]$$

where $\mathbf{Y}_n$ is the constant vector; **B** is the accumulated matrix; and **A** is the coefficient matrix. Matrix **A** can be determined by solving the following matrix system:

$$\mathbf{A} = \left(\mathbf{B}^T\mathbf{B}\right)^{-1}\mathbf{B}^T\mathbf{Y}_n = [a,b]^T \quad [15]$$

Substituting **A** into Eq. [12], the approximation equation becomes

$$\hat{P}^{(1)}(r_{k+1}) = \left[P^{(0)}(r_1) - \frac{b}{a}\right]\exp(-ak) + \frac{b}{a} \quad [16]$$

where $\hat{P}^{(1)}(r_{k+1})$ is the predicted value of $P^{(1)}(r_{k+1})$ at particle radius r_{k+1}. After the completion of an inverse AGO on Eq. [16], $\hat{P}^{(0)}(r_{k+1})$, the predicted value of $P^{(0)}(r_{k+1})$ at particle radius (r_{k+1}), becomes available and, therefore,

$$\hat{P}^{(0)}(r_{k+1}) = \hat{P}^{(1)}(r_{k+1}) - \hat{P}^{(1)}(r_k) \quad [17]$$

Substituting Eq. [16] into Eq. [17], the predictive equation becomes

$$\hat{P}^{(0)}(r_{k+1}) = \left[P^{(0)}(r_1) - \frac{b}{a}\right]\left[1-\exp(a)\right]\exp(-ak) \quad [18]$$

The gray model described here was applied in developing the PSD prediction model in this study. Measured data from sites located upstream of the alluvial fan of the Cho Shui River were used to build a gray model and predict the system outputs. The measurement data of the CMF of the PSD were used in this study as the original series. Both the Skaggs and GM(1,1) models were applied to predict the PSD, and their performances were evaluated. Finally, the prediction accuracy and the performance of these two models were compared in this study.

Particle Size Distribution Estimation Using the Skaggs and GM(1,1) Models

The Skaggs model requires three cumulative mass fractions, $P(r_0)$, $P(r_1)$, and $P(r_2)$, with three specified particle sizes r_0, r_1, and r_2. The GM(1,1) needs four sequential cumulative mass fractions, $P(r_i)$, $P(r_{i+1})$, $P(r_{i+2})$, and $P(r_{i+3})$, to correspond with the already known particle sizes r_i, r_{i+1}, r_{i+2}, and r_{i+3}. Table 1 shows the inputs and outputs for both models. A numerical example is performed in the Appendix using the Skaggs and GM(1,1) models.

COMPARISON OF MODEL PERFORMANCE

Four statistical criteria were used for the comparison of model performance, namely, the mean square error (MSE), the mean absolute percentage error (MAPE), the accumulative absolute error (AAE), and the coefficient of determination (R^2).

Table 1. Comparisons of model inputs and outputs for the Skaggs and GM(1,1) models.

Model	Input	Output
Skaggs model	r_0, r_1, r_2, $P(r_0)$, $P(r_1)$, $P(r_2)$	$P(r)$
GM(1,1) model	$P(r_i)$, $P(r_{i+1})$, $P(r_{i+2})$, $P(r_{i+3})$	$P(r_{i+4})$

1. The MSE is defined as

$$\text{MSE} = \frac{1}{n}\sum_{k=1}^{n}\left[P^{(0)}(r_k) - \hat{P}^{(0)}(r_k)\right]^2 \quad [19]$$

where $\hat{P}^{(0)}(r_k)$ denotes the predicted value of mass fraction $P^{(0)}(r_k)$ at particle sequence index k, and n is the total number of predictions.

2. The MAPE is defined as

$$\text{MAPE} = \frac{1}{n}\sum_{k=1}^{n}\left|\frac{P^{(0)}(r_k) - \hat{P}^{(0)}(r_k)}{P^{(0)}(r_k)}\right| \quad [20]$$

3. The AAE is defined as

$$\text{AAE} = \sum_{k=1}^{n}\left|P^{(0)}(r_k) - \hat{P}^{(0)}(r_k)\right| \quad [21]$$

4. The R^2 is defined as

$$R^2 = \left\{\frac{n\sum P^{(0)}(r_k)\hat{P}^{(0)}(r_k) - \left[\sum P^{(0)}(r_k)\right]\left[\sum \hat{P}^{(0)}(r_k)\right]}{\sqrt{n\sum\left[P^{(0)}(r_k)\right]^2 - \left[\sum P^{(0)}(r_k)\right]^2}\sqrt{n\sum\left[\hat{P}^{(0)}(r_k)\right]^2 - \left[\sum \hat{P}^{(0)}(r_k)\right]^2}}\right\}^2 \quad [22]$$

It represents the measure of minimal discrepancies between measured and predicted data. For example, a model with a large R^2 may be more preferable than one with a small R^2.

In this study, the results of both the Skaggs and GM(1,1) models were, in general, compared with each other using these four criteria.

MATERIALS AND METHODS

The Soil Samples

Soil samples were taken from three sampling sites, as shown in Fig. 1. They are all located upstream of the alluvial fan of the Cho Shui River, a gravel-covered area in central Taiwan. A total of 222 soil sam-

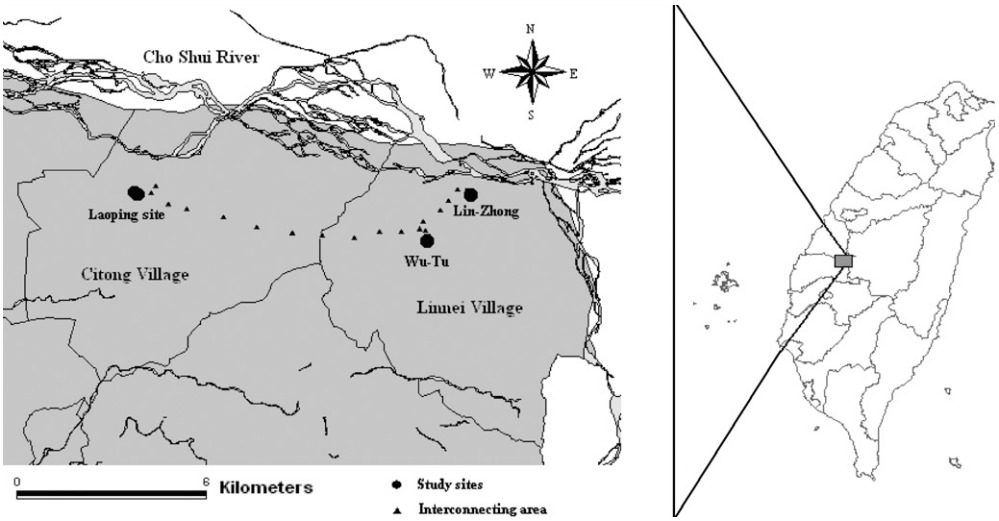


Fig. 1. Soil sample distributions across the study sites.

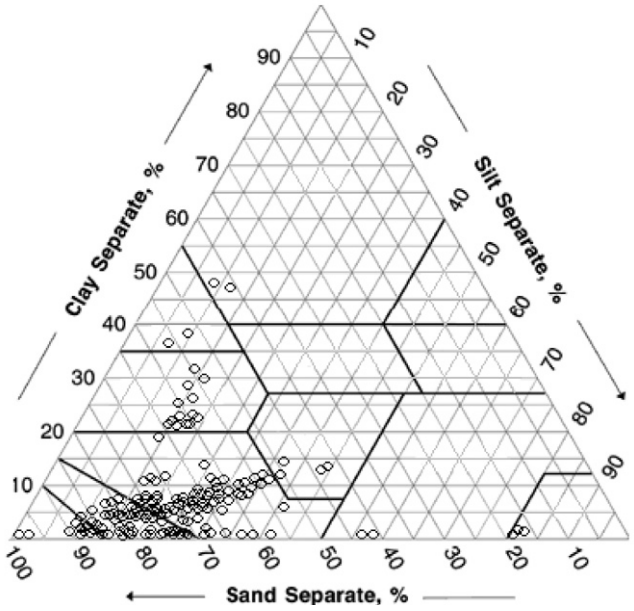


Fig. 2. Textural composition of the soil data set.

ples were taken from the sampling sites, of which 90 were from 30 boreholes in Laoping of Citong Village, 50 were from 25 boreholes located in Lin-Zhong of Linnei Village, 50 were from 25 boreholes in Wu-Tu of Linnei Village, and 32 soil samples were taken randomly from 17 scattered boreholes within the interconnecting areas.

Particle Size Analysis

The PSDs of the soil samples were determined using sieve analysis for particles >63 μm and the hydrometer analysis for particles <0.075 mm. The Cu and the Cc were derived from the following formulae:

$$\mathrm{Cu} = \frac{D_{60}}{D_{10}} \qquad [23]$$

and

$$\mathrm{Cc} = \frac{D_{30}^2}{D_{10} D_{60}} \qquad [24]$$

where D_{60}, D_{30}, and D_{10} represent the diameters corresponding to the percentage finer than 60, 30, and 10%, respectively.

The textural composition of the 222 soil samples was determined using the USDA classification system. The range of soil textures covered in this study is shown in Fig. 2. The 222 soil samples were classified into nine soil textures: loamy sand, sandy loam, sandy clay, loam, clay, sandy clay loam, sand, silt loam, and silt. Among the nine soil classes, sandy loam (43.24%), loamy sand (39.64%), sandy clay loam (6.30%), and sand (4.95%) were predominant (as shown in Table 2). To know how Cu and Cc affect the predictive ability of both models, they were determined for these four predominant soil classes.

Table 2. The soil texture classifications and their properties.

Soil texture	Samples	Percentage	Effective particle size	Uniformity coefficient	Curvature coefficient
	no.	%	mm		
Sand	11	4.95	0.0500 ~ 0.4900	1.75 ~ 16.90	0.84 ~ 3.00
Sandy clay	2	0.90	0.0150	86.67 ~ 153.33	1.24 ~ 2.48
Sandy clay loam	14	6.30	0.0020 ~ 0.0320	17.39 ~ 2000.00	0.31 ~ 7.54
Sandy loam	96	43.24	0.0020 ~ 0.0760	3.47 ~ 820.00	0.26 ~ 19.53
Loam	4	1.80	0.0015 ~ 0.0210	80.00 ~ 242.86	0.23 ~ 1.57
Clay	2	0.90	0.0057 ~ 0.0160	175.00 ~ 315.79	0.09 ~ 0.25
Loamy sand	88	39.64	0.0063 ~ 0.1000	6.00 ~ 181.82	0.14 ~ 16.53
Silt	3	1.35	0.0120 ~ 0.0150	0.033 ~ 2.530	0.459 ~ 0.990
Silt loam	2	0.90	0.0174 ~ 0.0197	5.00 ~ 15.98	0.240 ~ 0.570
Total samples	222				

RESULTS AND DISCUSSION

Samples of nine soil textures (shown in Table 2) were used to test both the Skaggs and GM(1,1) models. Figures 3a to 3i show the prediction results of both models for the different soil textures. As can be seen in Fig. 3b, 3e, and 3f, the predicted PSD curve of both models were in good agreement with the measured data. The GM(1,1) shows a performance superior to the Skaggs model in Fig. 3c, 3d, 3g, 3h, and 3i, while the Skaggs model shows better results than the GM(1,1) in Fig. 3a. The reason for this is that Eq. [1] of the Skaggs model is a logistic growth curve. When the measured PSD data closely follow a logistic type of function, the Skaggs model can obtain better results.

The performances of both models are presented in Table 3. The results show that the performance criteria of MSE and R^2 for the GM(1,1) model are superior to those of the Skaggs model except for sand.

According to the criteria MAPE and AAE, both models performed fairly well for the soil textures in general. Notice that italics represent better performance in Table 3. The comparisons reveal that, in general, the performance of the GM(1,1) is superior to the Skaggs model. As for the Skaggs model, poor estimates are exhibited for silt, as shown in Fig. 3h. This result is consistent with the conclusion of Skaggs et al. (2001).

Figures 4a to 4i show the predicted vs. measured CMFs of both models with their regression lines for the nine soil textures. The regression lines describe the minimized distance from the line to the data points of the individual models. The red dotted line is the 1:1 line. These figures show the relationship of predicted with measured CMFs for both the Skaggs and GM(1,1) models. Ideally, all the scatter points lying on the 1:1 line means that the predicted and measured CMF values are exactly the same. Alternatively, for individual regression lines that diverge from 1:1 line, e.g., Fig. 4c and 4d, one can easily tell that the CMF predicted by the GM(1,1) model coincides better with the 1:1 line than that predicted by the Skaggs model.

The interesting results also show that the performance of the GM(1,1) model depends less on the soil texture. The reason for this is that for most logistic type of PSD models (like the Skaggs model), the PSD curve type is fixed. Once the parameters are obtained, the model is set permanently for making all other predictions. By contrast, the GM(1,1) model calculates its coefficient matrix independently for each prediction. Furthermore, there is no unique type of empirical model for describing the full range of soil textures. The results may also imply that there is a different characteristic between the GM (1,1) and Skaggs models [i.e., the Skaggs model is predominated by a logistic type of function, but the GM(1,1) model is not].

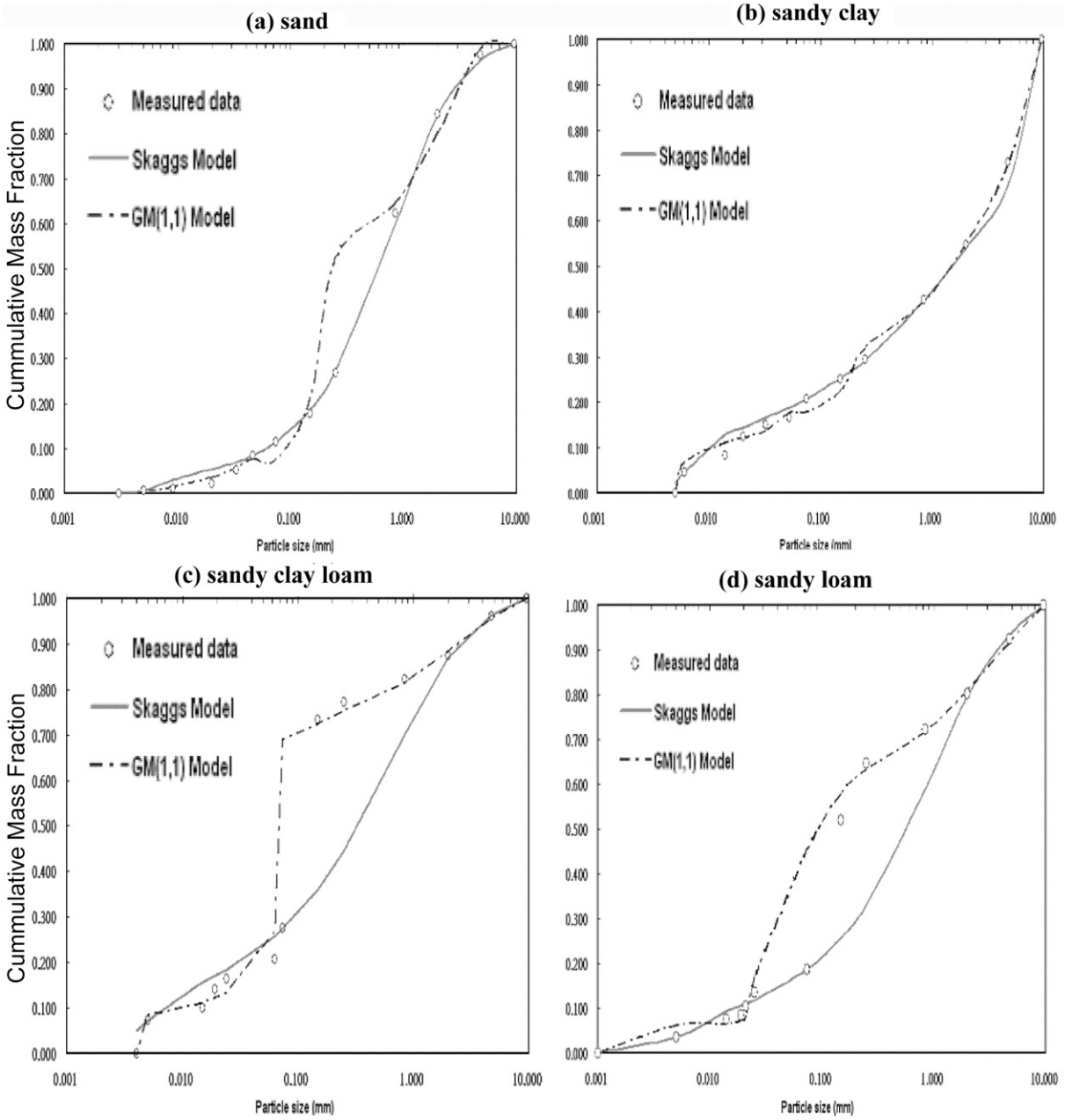


Fig. 3. Comparisons of the model predictions for (a) sand, (b) sandy clay, (c) sandy clay loam, (d) sandy loam, (e) loam, (f) clay, (g) loamy sand, (h) silt, and (i) silt loam soils.

Figure 5a presents the AAE of the Skaggs and GM(1,1) models for sand. Nine out of 11 soil samples show that the AAE of the Skaggs model is less than that of the GM(1,1) model, which means that the Skaggs model can more accurately predict sand estimates than the GM(1,1) model. Figures 5b to 5d show the AAE of the Skaggs and GM(1,1) models for sandy clay loam, sandy loam, and loamy sand. The GM (1,1) model has an AAE value less than the Skaggs model and, therefore, is better for predicting these soil textures.

The relationship between the Cu and AAE values of both models is presented in Fig. 5a to 5d. As can be seen, the AAE values for both the Skaggs and GM(1,1) models decrease when Cu increases. On the other hand, the relationship between the Cc and AAE values for both the Skaggs and GM(1,1) models shows a lot of fluctuations and the same trend, as shown in Fig. 6a to 6d.

CONCLUSIONS

From samples representing nine soil textures, a successful application of the gray model GM(1,1) to predict the PSD shows that this new prediction method is feasible, reliable, and highly efficient. It does not make assumptions about the shape of the curve, but deals directly with the original data. On the other hand, the Skaggs model needs to specify three bounds of radii (i.e., r_0, r_1, and r_2) and their related mass fractions. The bounds of radii should cover the range of the PSD in sequence from the smallest to the largest size to cover all possible sizes.

The Skaggs model itself is like an interpolation function. The GM(1,1), however, uses four successive continuous data chosen arbitrarily from the sequence of the original mass fractions to predict the following sequences. The model is not related to the particle sizes, which makes it easier to use the GM(1,1) than the Skaggs model.

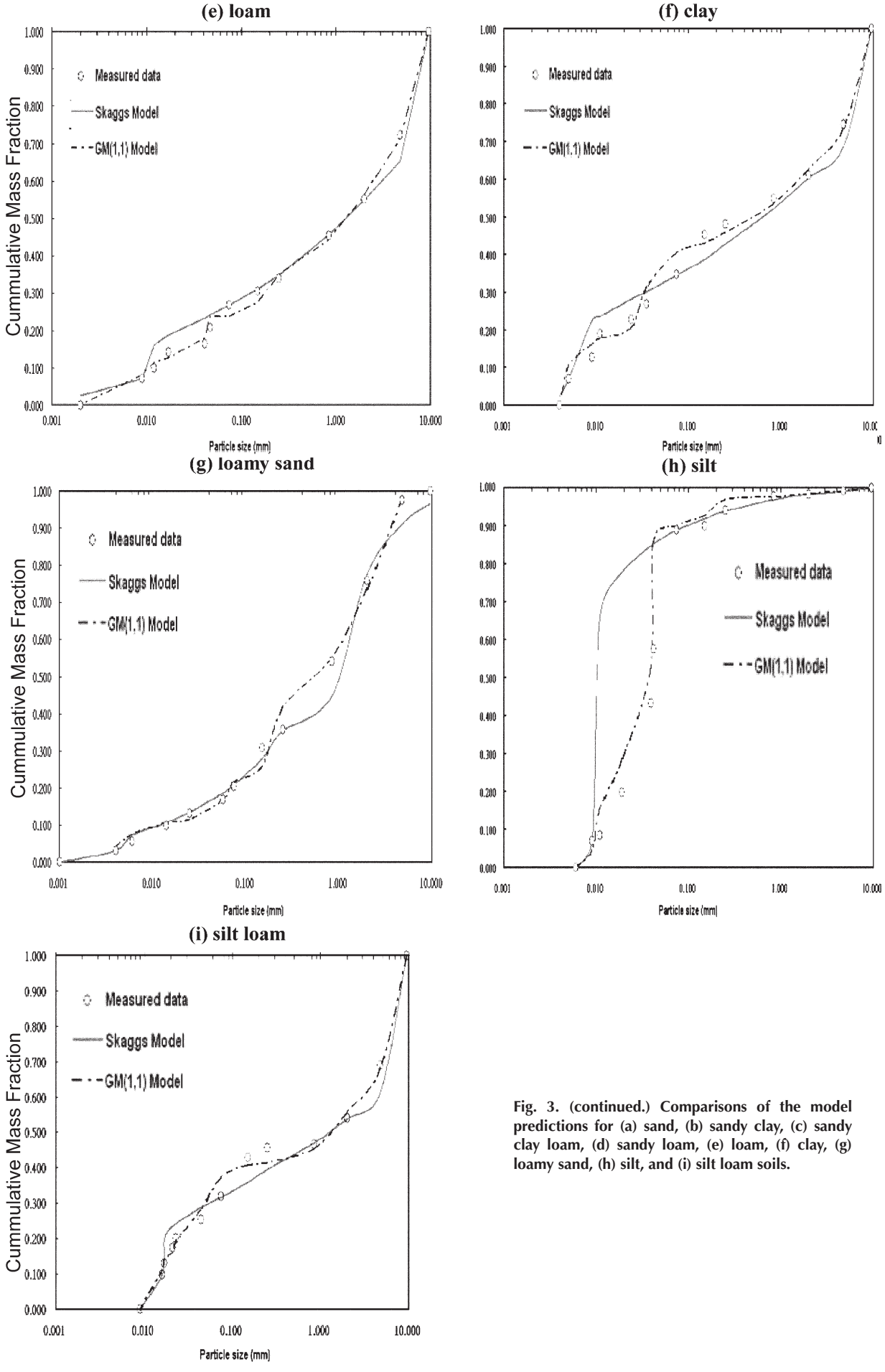


Fig. 3. (continued.) Comparisons of the model predictions for (a) sand, (b) sandy clay, (c) sandy clay loam, (d) sandy loam, (e) loam, (f) clay, (g) loamy sand, (h) silt, and (i) silt loam soils.

Table 3. Comparisons of the Skaggs and GM(1,1) models using the mean square error (MSE), mean absolute percentage error (MAPE), accumulative absolute error (AAE), and coefficient of determination (R^2). Italics represent better performance.

Soil texture	No. of samples	MSE		MAPE		AAE		R^2	
		Skaggs	GM(1,1)	Skaggs	GM(1,1)	Skaggs	GM(1,1)	Skaggs	GM(1,1)
		mm^2				mm			
Sand	11	*0.0002*	0.0065	0.3949	*0.3303*	*−0.0605*	−0.2446	*0.999*	0.962
Sandy clay	2	0.0005	*0.0003*	*0.0943*	0.1211	−0.0633	*−0.0099*	*0.996*	*0.996*
Sandy clay loam	14	0.0224	*0.0148*	*0.2022*	0.2207	0.6139	*−0.4297*	0.868	*0.910*
Sandy loam	96	0.0156	*0.0065*	*0.1732*	0.2780	0.6739	*−0.3262*	0.909	*0.959*
Loam	4	0.0015	*0.0003*	0.1561	*0.0866*	−0.1851	3.00×10^{-5}	0.989	*0.996*
Clay	2	0.0023	*0.0008*	0.1593	*0.1254*	*−0.0445*	−0.0855	0.978	*0.991*
Loamy sand	88	0.0012	*0.0006*	*0.0804*	0.1201	*0.1487*	0.9656	0.992	*0.993*
Silt	3	0.0768	*0.0087*	1.0261	*0.2229*	−1.8585	*−0.5747*	0.670	*0.986*
Silt loam	2	0.0020	*0.0014*	0.1366	*0.0908*	0.0043	*−0.0151*	0.948	0.948

The performance of both the Skaggs and GM(1,1) models were evaluated using four different criteria (i.e., MSE, MAPE, AAE, and R^2). Overall, the results show that the prediction accuracy for the PSD of the GM(1,1) is better than the Skaggs model for most soil textures. The results also show that when the soil texture is silt, the Skaggs model yields a large error, which is consistent with the conclusion of Skaggs et al. (2001). On the contrary, the performance of the GM(1,1) is quite accurate and is not affected by the soil texture in general.

APPENDIX: CALCULATION EXAMPLE

The following example calculation was performed using both the Skaggs model and the GM(1,1) model based on the theoretical background above.

Prediction Using the Skaggs Model

To use Skaggs model, three radii of soil particles have to be specied (i.e., r_0, r_1, and r_2). As shown in Table A1, in this study we used r_0 = 0.003 mm, r_1 = 0.074 mm (the particle size of a no. 200 sieve), and r_2 = 2.0 mm (the boundary between sand and gravel). The related mass fractions can be specified as $P(r_0) = 0.034$, $P(r_1) = 0.215$, and $P(r_2) = 0.813$.

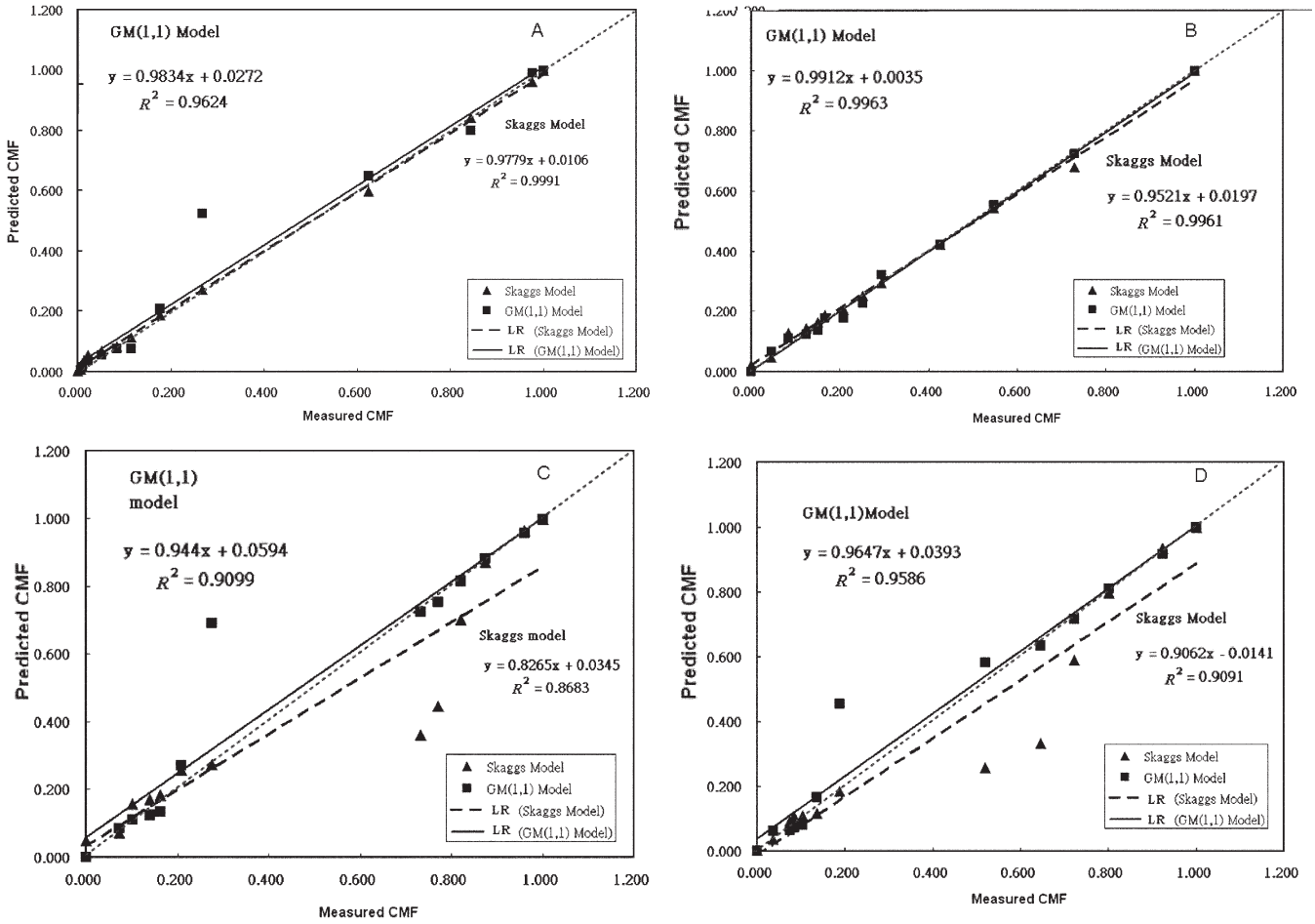
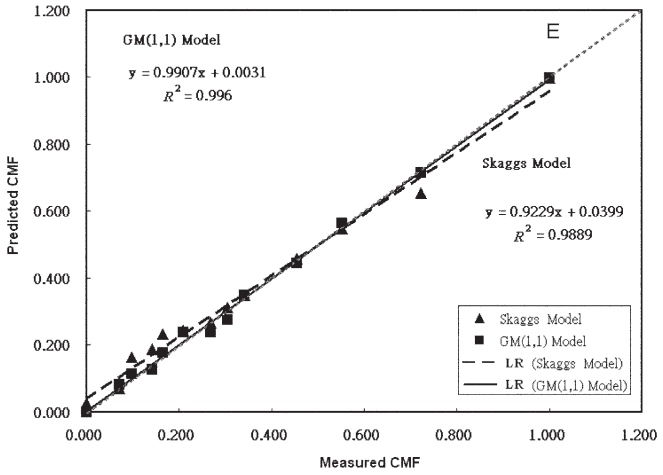
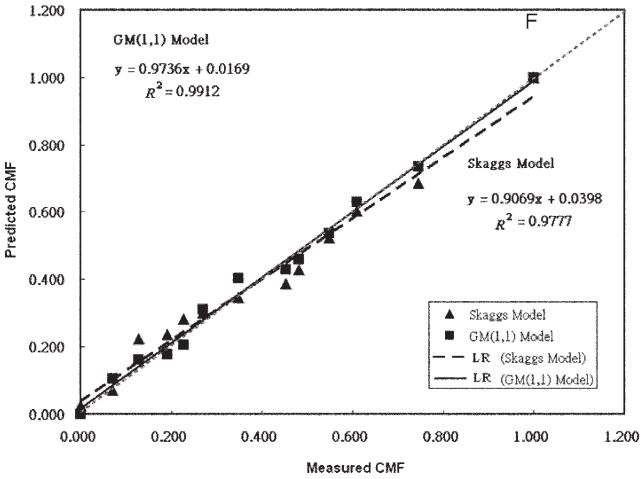
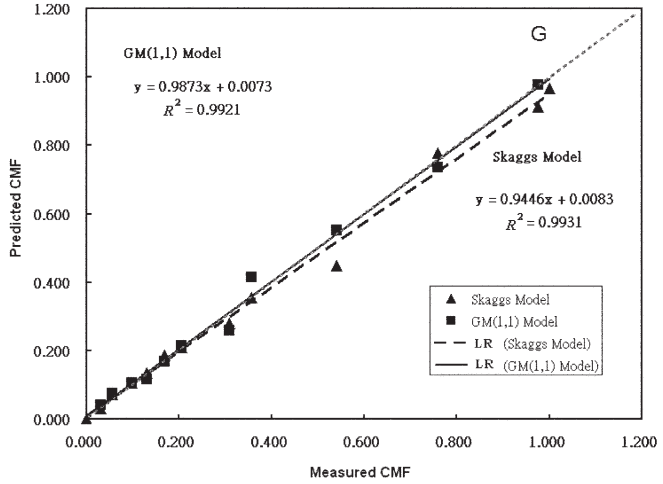
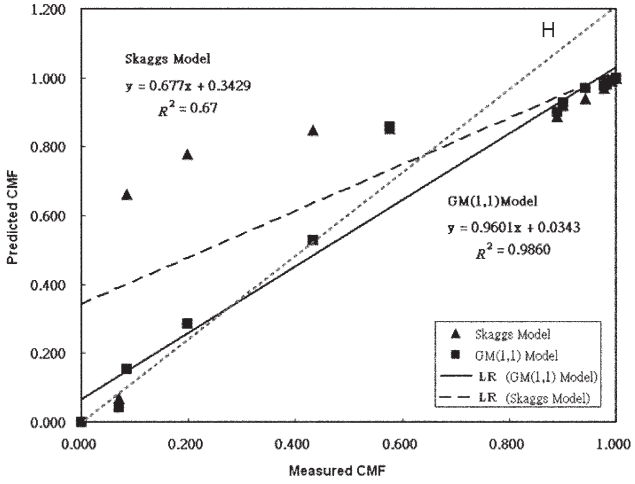


Fig. 4. The predicted cumulative mass fraction (CMF) vs. measured CMF for (a) sand, (b) sandy clay, (c) sandy clay loam, (d) sandy loam, (e) loam, (f) clay, (g) loamy sand, (h) silt, and (i) silt loam soils; the red dotted line lies on a 45° angle.









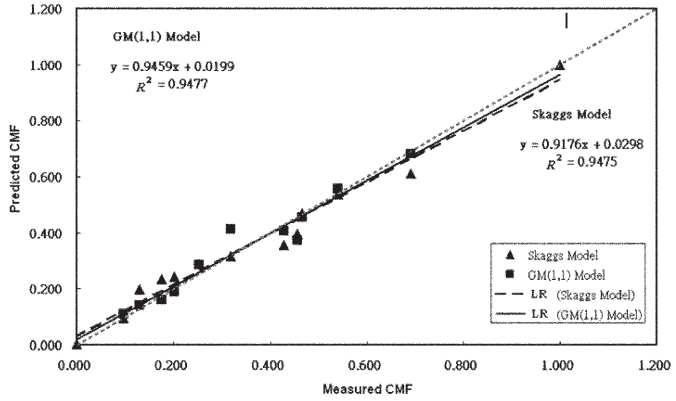


Fig. 4. (continued.)The predicted cumulative mass fraction (CMF) vs. measured CMF for (a) sand, (b) sandy clay, (c) sandy clay loam, (d) sandy loam, (e) loam, (f) clay, (g) loamy sand, (h) silt, and (i) silt loam soils; the red dotted line lies on a 45° angle.

By substituting these three radii and their related mass fractions into Eq. [4], [3], and [2] sequentially, we obtain

$$\alpha = \frac{1}{\ln\left[(r_1 - r_0)/(r_2 - r_0)\right]} = \frac{1}{\ln\left[(0.074 - 0.003)/(2.0 - 0.003)\right]} = -0.30$$

$$\beta = \alpha \ln\left(\frac{r_1 - r_0}{r_0}\right) = -0.3 \ln\left(\frac{0.074 - 0.003}{0.003}\right) = -0.949$$

$$v = \ln\left\{\frac{\left[1/P(r_1)\right]-1}{\left[1/P(r_0)\right]-1}\right\} = \ln\left[\frac{(1/0.215)-1}{(1/0.034)-1}\right] = -2.052$$

and

$$w = \ln\left\{\frac{\left[1/P(r_2)\right]-1}{\left[1/P(r_0)\right]-1}\right\} = \ln\left[\frac{(1/0.813)-1}{(1/0.034)-1}\right] = -4.816$$

The model parameters *u* and *c* can be obtained as follows:

$$c = \alpha \ln\left(\frac{v}{w}\right) = -0.3 \ln\left(\frac{-2.052}{-4.816}\right) = 0.256$$

$$u = -v^{1-\beta} w^{\beta} = \left|-2.052\right|^{1+0.949} \left|-4.816\right|^{-0.949} = 0.913$$

By substituting parameters *u* and *c* into Eq. [1], the prediction equation becomes

$$P(r) = \frac{1}{1 + \left\{\left[1/P(r_0)\right]-1\right\} \exp\left(-uR^c\right)}$$

$$= \frac{1}{1 + \left[(1/0.034)-1\right] \exp\left\{-0.913\left[(r/0.003)-1\right]^{0.256}\right\}}$$

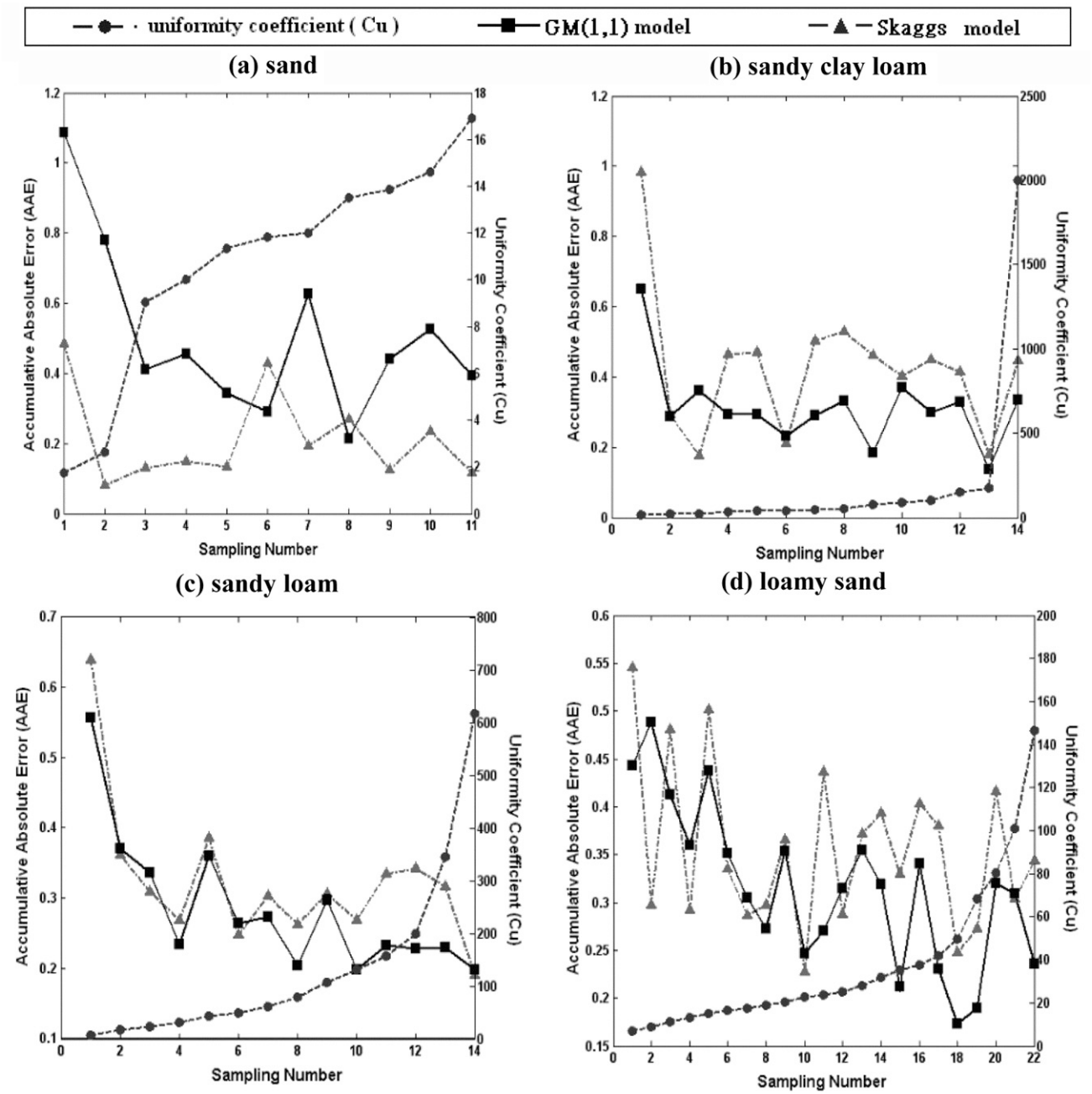


Fig. 5. Relationship between the uniformity coefficient and the accumulative absolute error of the model predictions for (a) sand, (b) sandy clay loam, (c) sandy loam, and (d) loamy sand soils.

When particle size $r = 0.250$, the related mass fraction is predicted as

$$P(0.25) = \frac{1}{1 + \left[\left(1/0.034\right) - 1\right]\exp\left\{-0.913\left[\left(0.250/0.003\right) - 1\right]^{0.256}\right\}} = 0.372$$

The remaining mass fractions $P(r_1)$ can be predicted accordingly. The prediction results are shown in Table A1 and Fig. A1.

Prediction Using the GM(1,1) Model

For the GM(1,1) model, four particle sizes are required. In this study, we chose the first four particle sizes r (mm) in Table A1 (i.e., $r_i = 9.53$ mm, $r_{i+2} = 4.76$ mm, $r_{i+3} = 2.0$ mm, and $r_{i+4} = 0.84$ mm) as the original particle size sequence. The resulting CMF sequence is

$$X^{(0)} = \left\{P^{(0)}(r_1), P^{(0)}(r_2), ..., P^{(0)}(r_4)\right\} = \{1.0,\ 0.959,\ 0.813,\ 0.683\}$$

From the original CMF sequence, the procedures used for the GM(1,1) prediction are

Step 1. On the basis of $X^{(0)}$, the AGO sequence $X^{(1)}$ is

$$X^{(1)} = \{1,\ 1+0.959,\ 1+0.959+0.813,\ 1+0.959+0.813+0.683\}$$
$$= \{1,\ 1.959,\ 2.772,\ 3.455\}$$

Step 2. The average generating sequence is

$$Z^{(1)} = \left\{\frac{1+1.959}{2},\ \frac{1.959+2.772}{2},\ \frac{2.772+3.455}{2}\right\}$$
$$= \{1.479,\ 2.3655,\ 3.1135\}$$

Step 3. The gray parameters a and b are determined by solving Eq. [13]:

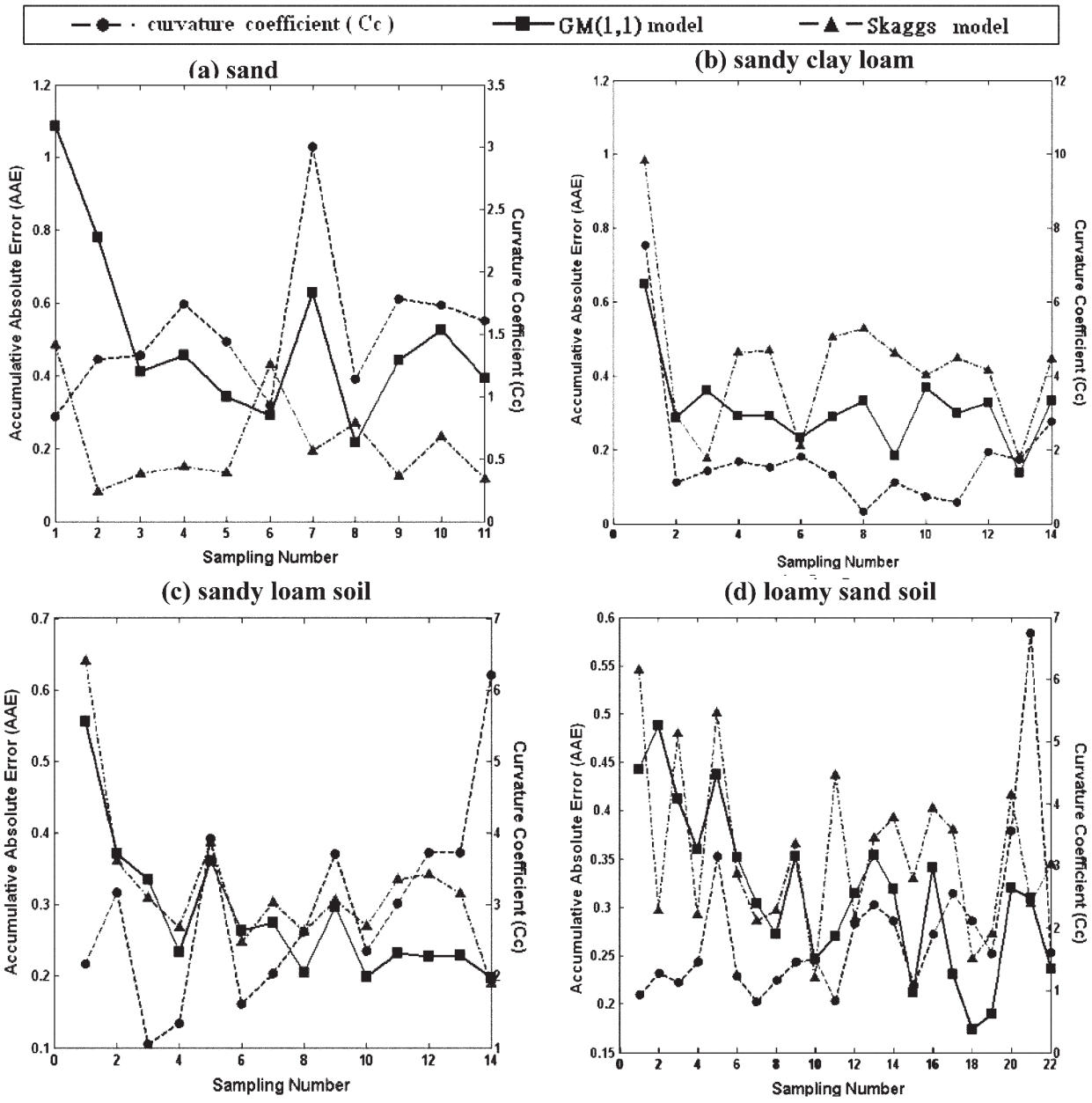


Fig. 6. Relationship between the curvature coefficient and the accumulative absolute error of the model predictions for (a) sand, (b) sandy clay loam, (c) sandy loam, and (d) loamy sand soils.

$$\begin{bmatrix} P^{(0)}(r_2) \\ P^{(0)}(r_3) \\ \vdots \\ P^{(0)}(r_n) \end{bmatrix} = \begin{bmatrix} -Z^{(1)}(2) & 1 \\ -Z^{(1)}(3) & 1 \\ \vdots & \vdots \\ -Z^{(1)}(n) & n \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix}$$

that is,

$$\begin{bmatrix} 0.959 \\ 0.813 \\ 0.683 \end{bmatrix} = \begin{bmatrix} -1.479 & 1 \\ -2.365 & 1 \\ -3.1135 & 1 \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix}$$

and therefore,

$$\begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} 0.1687 \\ 1.2097 \end{bmatrix}$$

Step 4. Based on Eq. [18], the prediction equation becomes

$$\hat{P}^{(0)}(r_{k+1}) = \left[P^{(0)}(r_1) - \frac{b}{a}\right][1 - \exp(a)]\exp(-ak)$$

When $k = 5$, the predicted value for $\hat{P}^{(0)}(r_5)$ is

$$\hat{P}^{(0)}(r_5) = \left[P^{(0)}(r_1) - \frac{1.2097}{0.1687}\right][1 - \exp(0.1687)]\exp[-0.1687(5)] = 0.577$$

Step 5. Since $\hat{P}^{(0)}(r_5)$ has been obtained, the next predicted variable [i.e., $\hat{P}^{(0)}(r_6)$] can be estimated sequentially by using the new sequence:

$$X^{(0)} = \{P^{(0)}(r_2), P^{(0)}(r_3), ..., P^{(0)}(r_5)\} = \{0.959,\ 0.813,\ 0.683, 0.577\}$$

Step 6. Steps 1 to 5 are repeated to determine the remaining predicted variables.

The results predicted by the GM(1,1) model are presented in Table A1 and Fig. A1.

Table A1. Examples of particle size distribution predicted using the Skaggs model and GM(1,1) models.

Data	Cumulative mass fraction												
	9.530 mm	4.760 mm	*2.000* mm †	0.840 mm	0.250 mm	0.149 mm	*0.074 mm*	0.03 mm	0.014 mm	0.011 mm	0.008 mm	*0.003 mm*	0.001 mm
Measured data	1.000‡	0.959‡	0.813‡§	0.683‡	0.527	0.449	0.215§	0.168	0.129	0.082	0.056	0.034§	0.000
Skaggs	1.000	0.933	0.813	0.622	0.372	0.292	0.215	0.143	0.112	0.102	0.091	0.034	0.016
GM(1,1)	1.000	0.959	0.813	0.683	0.577	0.436	0.353	0.183	0.100	0.081	0.062	0.035	0.023

†The italic numbers represent the particle size used for Skaggs model predictions.

‡ Used for the GM(1,1) model predictions.

§ Used for the Skaggs model predictions.

ACKNOWLEDGMENTS

We acknowledge that the National Science Council sponsored this project under Project no. NSC-96-2221-E-224-029 and NSC-95-2221-E-224-005. We would also like to acknowledge Teamrat Ghezzehei, who made valuable suggestions that made it possible for the publication of this manuscript.

REFERENCES

Arya, L.M., F.J. Leij, M.Th. van Genuchten, and P.J. Shouse. 1999. Scaling parameter to predict the soil water characteristic from particle-size distribution data. Soil Sci. Soc. Am. J. 63:510–519.

Arya, L.M., and J.F. Paris. 1981. A physicoempirical model to predict soil moisture characteristics from particle-size distribution and bulk density data. Soil Sci. Soc. Am. J. 45:1023–1030.

Buchan, G.D. 1989. Applicability of the simple lognormal model to particle size distribution in soils. Soil Sci. 147:155–161.

Buchan, G.D., K.S. Grewal, and A.B. Robson. 1993. Improved models of particle-size distribution: An illustration of model comparison techniques. Soil Sci. Soc. Am. J. 57:901–908.

Campbell, G.S. 1985. Soil physics with BASIC: Transport models for soil–plant systems. Elsevier, Amsterdam.

Deng, J.-L. 1989. The quality of the GM(1,1) model. p. 205–215. *In* The course of the grey system theory. Huazhong Univ. of Science and Technology, Wuhan, China.

Gee, G.W., and J.W. Bauder. 1986. Particle-size analysis. p. 383–411. *In* A. Klute (ed.) Methods of soil analysis. Part 1. 2nd ed. Agron. Monogr. 9. ASA and SSSA, Madison, WI.

Guo, R. 2004. Interval-valued fuzzy set modelling of system reliability. p. 157–164. *In* T. Dohi and W.Y. Yun (ed.) Advanced reliability modeling. Proc. Asian Int. Worksh. (AIWARM 2004), Hiroshima, Japan. 26–27 Aug. 2004. World Scientific Publ. Co., Singapore.

Guo, Z., X. Song, and J. Ye. 2005. A Verhulst model on time series error corrected for port throughput forecasting. J. East. Asia Soc. Transp. Stud. 6:881–891.

Gupta, S.C., and W.E. Larson. 1979a. Estimating soil-water retention characteristics from particle size distribution, organic matter percent, and bulk density. Water Resour. Res. 15:1633–1635.

Gupta, S.C., and W.E. Larson. 1979b. A model for predicting packing density of soils using particle size distribution. Soil Sci. Soc. Am. J. 43:758–764.

Hsu, L.-C. 2003. Applying the grey prediction model to the global integrated circuit industry. Technol. Forecast. Social Change 70:563–574.

Hwang, S.I., K.P. Lee, D.S. Lee, and S.E. Powers. 2002. Models for estimating soil particle-size distributions. Soil Sci. Soc. Am. J. 66:1143–1150.

Hwang, S.I., and S.E. Powers. 2003. Using particle-size distribution models to estimate soil hydraulic properties. Soil Sci. Soc. Am. J. 67:1103–1112.

Mao, M., and E.C. Chirwa. 2006. Application of grey model GM(1,1) to vehicle fatality risk estimation. Technol. Forecast. Social Change 73:588–605.

Shirazi, M.A., and L. Boersma. 1984. A unifying quantitative analysis of scale. Soil Sci. Soc. Am. J. 48:142–147.

Skaggs, T.H., L.M. Arya, P.J. Shouse, and B.P. Mohanty. 2001. Estimating particle size distribution from limited soil texture data. Soil Sci. Soc. Am. J. 65:1038–1044.

Tseng, F.-M., H.-C. Yu, and G.-H. Tzeng. 2001. Applied hybrid grey model to forecast seasonal time series. Technol. Forecast. Social Change 67:291–302.

Yu, P.-S., C.J. Chen, S.-J. Chen, and S.-C. Lin. 2001. Application of grey model toward runoff forecasting. J. Am. Water Resour. Assoc. 37:151–166.

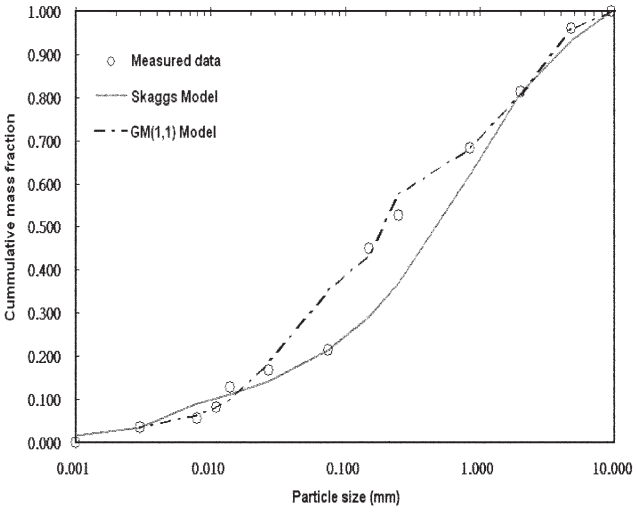


Fig. A1. Comparison of predicted results using the Skaggs and GM(1,1) models.