Pertanika 3(2), 113–124 (1980)

A Pattern Recognition Approach to the Structure Identification of Soil Hydraulic Properties[1]

WAN SULAIMAN WAN HARUN and G. C. VANSTEENKISTE[2]
Department of Soil Science, Faculty of Agriculture, Universiti Pertanian Malaysia. Serdang, Selangor, Malaysia.

Key words: Pattern recognition; soil hydraulic properties; structure characterization.

RINGKASAN

Satu kaedah mengecam corak yang mudah diperihal dan digunakan untuk menentu struktur-struktur bagi sifat air tanah dan perkaitan antara difusiviti air tanah dengan kandungan air tanah pada sampel-sampel dari Siri Bungor. Model asymptotik didapati sesuai untuk sifat air tanah sementara fungsi eksponen sesuai untuk pertalian $D(\theta)-\theta$. Perbandingan dengan kaedah "Least Squares" menunjukkan bahawa cara mengecam corak ini boleh diguna sebagai satu lagi kaedah yang penting dalam penentuan atau pengenalan struktur-struktur model.

SUMMARY

A simple pattern recognition approach to the identification of model structures is described and used to identify structures for the soil water characteristic and soil water diffusivity, $D(\theta)$, versus water content relationships of disturbed samples for Bungor series. An asymptotic model and an exponential function were found to be suitable structures for the soil water characteristic and $D(\theta)-\theta$ relationship respectively. A comparison with the Least Squares Technique suggests the pattern recognition approach to be a useful alternative in identification of model structures.

INTRODUCTION

In the computer implementation of numerical solutions of the unsaturated soil water flow problem, tabulated values relating soil water diffusivity D, or hydraulic conductivity K, to volumetric water content θ and/or soil water pressure head, h in the range of interest are adequate and yield excellent results (Hanks and Bowers, 1962). However, in terms of storage and computational efficiency, simple empirical formulae are more desirable. Moreover, for derivation of closed analytical solutions, the formula representation of these relationships is a prerequisite.

Since the last three decades many empirical formulae for the hydraulic characteristics have been used. The unknown parameters of the various formulae or structures are generally determined by using some kind of best fitting technique in order to adjust them to the experimentally measured data of each particular soil or class of soils. As most of the structures involved are nonlinear, the least squares method, for example, applied to several of them can be a tedious and time consuming operation. For the same reason of nonlinearity, curve fitting technique based on the minimization of sum squares of deviation can be inadequate unless accompanied by some kind of weighting. Comparisons of goodness of fit among the different structures are at best subjective.

A recent approach to structure identification is that of pattern recognition proposed by Karplus (1972) and Saridis and Hofstadter (1974). Here, structures are considered to be different patterns and identification is regarded as a task of recognizing patterns using experience and current information. Certain details of the procedure have been worked out by Simundich (1975) and these have been extended and modified by Vansteenkiste, Bens and Spriet (1978a). The method suggested by the latter authors is still in its infancy and more exploratory work with various kinds of data and structures is required to establish it as a useful tool in structure

[1]Part of D. Agr. Sc. thesis submitted to the State University of Ghent.
[2]Professor, Dept. of Applied Mathematics and Biometry, Faculty of Agricultural Sciences, State University of Ghent.
Key to authors' names: Sulaiman, W.H.W, and Vansteenkiste, G.C.

characterization. This study aims at applying their technique in the identification of simple structures for the soil water characteristic and diffusivity which are necessary in solving the soil water flow equation.

THEORETICAL FRAMEWORK

General approach

The pattern recognition approach, in simple terms, involves a comparison of the pattern emerging from certain operations on the data set coming from the system to be modelled with an input library of patterns from known "candidate" models. A choice is then made among the candidates as to which structure is best adapted to the data.

The procedure is illustrated in Fig. 1. Two stages of operation are distinguished; first is the training of the classification algorithm (switches in position I) and second is the use of the classifier (switches in position II). In the first stage, a number of candidate models are proposed. A "feature extraction" procedure is performed on artificial data generated from these models. The resulting "feature space" is then classified into partitions corresponding to different models When the training is complete the classification algorithm or pattern recognizer is coupled to the data being investigated (second stage) and the most suitable model selected. The next step is then to identify the parameters of the chosen model.

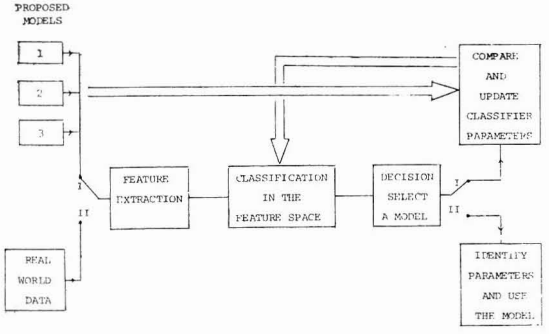

Fig. 1. Structure determination by the Pattern Recognition Approach of Vansteenkiste, Bens and Spriet (1978a).

Feature Extraction

Feature extraction is the most crucial part of the method. Features or characteristic expressions have to be defined, on the basis of which a choice among the different models will be made. No hard and fast rules exist in the choice of the features. However, they should be rather insensitive to noise besides having good discriminating properties. Very often one feature per proposed model is used, but sometimes more features can provide better discrimination. The different features form a so-called feature space in which most of the characterization operations are performed.

Consider the data set $D = \{ (t_i, x_i) \mid i = 1, 2, \ldots, n \}$ coming from the experiment or simulation of the experiment. Feature extraction can be regarded as a mapping of the set D of all possible data sets to a feature space:

$$\theta : D \rightarrow R^k : D \rightarrow f$$

where k is the number of features and f is a point in the feature space. The mapping θ can be single valued, i.e., a single point f is the image of a set D (Simundich, 1975), or multiple valued, i.e., more than one point f is the image of D, as in the current approach of Vansteenkiste, Bens and Spriet (1978a). The advantage of the multiple-valued mapping is that a poor or erroneous measurement will not destroy the information present in the other measurements.

Classification of the Feature Space

During the training stage a "classifier" splits up the feature space into partitions in an optimal way, each subset corresponding to a cluster of points and hence to a candidate (proposed) model. To achieve this several simulation runs have to be made with each model. Feature points derived from these simulations are used as input to the classifier, whose parameters are then adjusted iteratively so as to give maximal correspondence between input and output, the latter being the various partitions or subsets of the feature space. It is, thus, evident that the more simulation runs with each of the candidate models and the more diverse these runs are, the better would the classification algorithm be.

For the discussion and development of the classifier one is referred to a number of handbooks (Nilsson, 1965; Young and Calvert, 1974). Kanal (1974) provides an excellent survey of the different methods available while Vansteenkiste, Bens and Spriet (1978) discuss the various problems of choosing a classifier algorithm. The latter authors emphasized that where it is deemed fit, as in a two or three-dimensional feature space, visual classification can be the most convenient method.

TABLE 1

Some empirical formulae used to represent the water content-pressure head relationship

Formula	Source	
Model 1: $S = \dfrac{\alpha}{\alpha + (\ln \lvert h \rvert)^{\beta}}$	Haverkamp *et al.* (1977)	(1)
Model 2: $S = \dfrac{\delta}{\delta + \lvert h \rvert^{\gamma}}$	Haverkamp *et al.* (1977)	(2)
Model 3: $S = \begin{cases} \left(\dfrac{h_a}{h}\right)^{\lambda} & , h < h_a \\ 1 & , h > h_a \end{cases}$	Brooks and Corey (1964)	(3)

S is the dimensionless water content given by $S = (\theta - \theta_r)/(\theta_s - \theta_r)$ where θ_s and θ_r are the saturation and residual water contents respectively.

STRUCTURE CHARACTERIZATION OF THE SOIL WATER CHARACTERISTIC

Candidate Models

In most instances, the soil water characteristic or the θ-h relationship yields an S-shaped curve, although it is not uncommon to find a relationship which exhibits two or more points of inflection. Some of the empirical relationships that have been used are listed in Table 1. These form the candidate models in the structure identification.

Feature Extraction

An examination of the candidate models reveals little except that the θ-h curve tends to be rather flat at the low and high ends. Although the first derivative has an important physical significance (this being the specific water capacity), it offers no immediate help in the characterization process. We thus resort to the parameters of the models for deriving features, since in the ideal cases, these parameters are invariant for each model. Their variability would be a measure of deviation from the model under consideration. In the present identification process only one parameter per proposed model is used to derive features.

Feature 1:

As h is negative, operations are made easier by using suction head $= -h$, so that Model 1 now reads

$$S = \frac{\alpha}{\alpha + (\ln \psi)^{\beta}} \tag{4}$$

Differentiating with respect to ψ we obtain

$$\frac{dS}{d\Psi} = -\frac{S^2\beta}{\alpha\Psi}(\ln \Psi)^{\beta-1} \tag{5}$$

Elimination of α from (4) and substituting it in (5) yields

$$\beta = -\frac{dS}{d\Psi}\frac{\Psi \ln \Psi}{S(1-S)} \text{ or } \beta = -\frac{dS}{dh}\frac{h \ln \lvert h \rvert}{S(1-S)} \tag{6}$$

Thus, for any point in the data set β may be obtained by the use of Eq. (6) and the central difference approximation $dS/dh = (S_{i+1} - S_{i-1})/2\triangle h$. Now, two random points, A and B with coordinates (h_A, S_A) and (h_B, S_B) respectively, are taken from the data set (simulated or experimental) and feature 1 or the first coordinate of the point corresponding to the chosen (A,B)-tuple is computed according to:

$$f_1 = \frac{\beta_A}{\beta_B} \tag{7}$$

Hence, f_1 provides us with a measure of the variability of parameter β. The procedure is repeated for different (A,B)-tuples providing a set of values corresponding to the projections on one of the axes of the feature space.

Feature 2:

The second feature is obtained by resorting to Model 2. Again, differentiating S with respect to h or Ψ and solving for γ yields

$$\gamma = \frac{-h}{S(1-S)}\frac{dS}{dh} \tag{8}$$

Following the same procedure as with the first, the second feature, f_2, is found by taking the ratio of the γ's at the same two points, A and B used earlier. Thus,

$$f_2 = \frac{\gamma_A}{\gamma_B} \tag{9}$$

Feature 3:

Model 3 is now used to derive feature 3 or the third component of the point in the feature space. Differentiating and solving for λ yield

$$\lambda = -\frac{h}{S} \cdot \frac{dS}{dh}, \quad h < h_a \tag{10}$$

Feature 3 is then calculated using points A and B

$$f_3 = \frac{\lambda_A}{\lambda_B}, \quad h < h_a \tag{11}$$

Complications arise for $h > h_a$ because now the function has a constant value (i.e., equal to 1) and, therefore, λ is zero. One way of overcoming this dilemma is to define an alternative feature f'_3 specific to this region according to

$$f'_3 = S_A/S_B, \quad h > h_a \tag{11a}$$

Note: If the absolute value of any feature is greater than unity, then the ratio is inverted, i.e., its reciprocal is taken instead. The reason for taking ratios for all features is thus clear, that is, to reduce them all to the same scale of -1 to $+1$.

Since computation of the second set of parameters involves very little extra effort, three more features can be easily defined. These are

$$f_4 = \frac{\alpha_A}{\alpha_B}, \quad f_5 = \frac{\delta_A}{\delta_B}, \quad f_6 = \frac{h_{a_A}}{h_{a_B}} \tag{12}$$

Training

The feature extraction procedure just described is performed on each model using simulated data. For each model and for a simulation run, two random points A and B are taken and the features f_1, f_2 and f_3 or f'_3 computed according to the procedure outlined. This gives a point in the feature space belonging to the given model. The procedure is repeated for as many points as desired (the more points the better), after which more simulations are performed and feature points computed likewise. In this study five simulation runs per model are carried out. The parameter values used for the different models are shown in Table 2. To make the training more realistic a 2% of full scale random error (hence 0.01 cm^3/cm^3 water content) is introduced to the simulated data.

For convenience and ease of visual comparison the features are now plotted two by two. Figures 2(a) and (b) show feature points for Model 1 and Model 2 in the $f_1 - f_2$ feature space, and for Models 1 and 3 in the $f_1 - f_3$ feature space, respectively. Some overlapping of feature points is observed especially in Fig. 2(b). This is due to the contamination of the 2% error. Had no error been introduced one would expect to obtain two narrow bands of clusters, one along the line $f_1 = 1$, consisting of points from Model 1 and the other along the line $f_2 = 1$ or $f_3 = 1$ for feature points from Model 2 or Model 3, respectively. Nevertheless,

TABLE 2

Parameter values used to generate data needed in the classification of the feature space for the different models of soil water characteristic

Simulation run	Model 1		Model 2		Model 3	
	α	β	δ	γ	h_a	λ
1	50	2.4	1×10^2	2.2	−5	0.21
2	100	2.6	5×10^1	3.0	−10	0.23
3	250	3.2	1×10^3	3.4	−25	0.30
4	500	3.8	1×10^4	2.8	−50	0.28
5	1,000	3.6	1×10^6	2.5	−100	0.32

classification is still possible. One may choose a linear or a nonlinear splitting of the feature space. In these instances linear splitting is performed whereby the feature space is divided into two in such a way that there is a minimum number of non-conforming feature points for each model. An approximately 45° line, extending from the origin, as shown, is found satisfactory in both cases. Thus, in Fig. 2(a) the subspace below the dashed line is assigned to Model 1 and that above the line to Model 2. Likewise, in Fig. 2(b), the lower subspace belongs to Model 1 while the upper to Model 3.

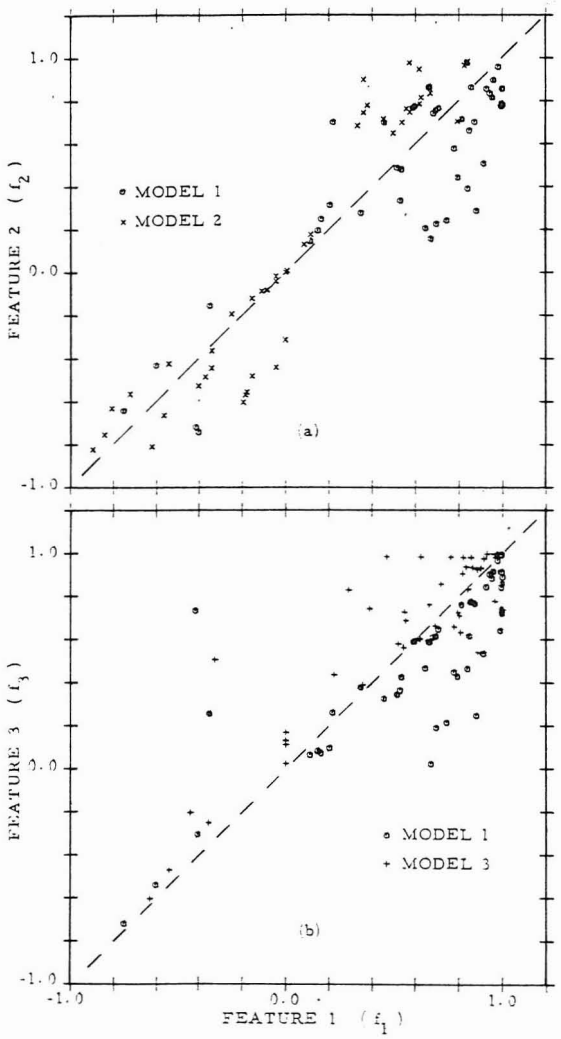

Fig. 2. Feature space of the various candidate models for the soil water characteristic: (a) Model 1 and Model 2, (b) Model 1 and Model 3.

Identification of the structure for the soil water characteristic

The test cases consist of the water content versus pressure head data for disturbed samples from each of four horizons from two profiles of the Bungor series (Typic Paleudult) (Wan Sulaiman, 1979). These are shown in Fig. 3 where the asymptotic model (Model 1) has been fitted to data from each horizon.

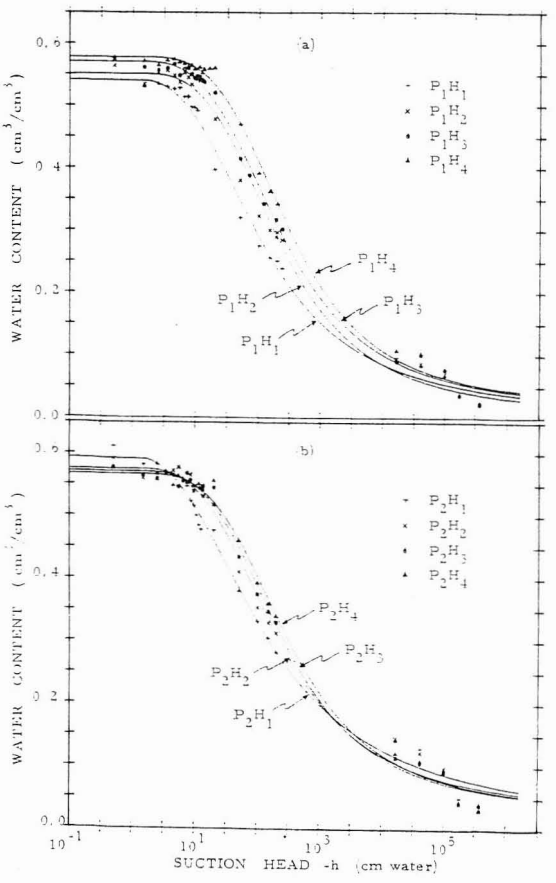

Fig. 3. Water content versus suction head for different horizons of (a) Profile 1 and (b) Profile 2 of Bungor Series. P with subscript refers to profile number while H with subscript refers to the horizon number. Solid lines are asymptotic functions (Model 1) fitted through experimental points.

Data for each soil horizon is now subjected to the feature extraction procedure and plots of f_2 against f_1 and f_3 against f_1 are prepared. In all cases the choice between Model 1 and Model 2 can be readily made since most of the feature

points fall in the space belonging to Model 1. However, discrimination between Model 1 and Model 3 is rather difficult in most of the test cases. Figures 4 (a) through (d) illustrate some of the features of the analysis. Figures 4 (a) & (b) clearly indicate the superiority of Model 1 to the other two. Figure 4 (c), on the other hand, appears slightly to favour Model 3 to Model 1, while Fig. 4 (d) indicates the reverse; in both cases, the distinction is not very convincing. Overall, the asymptotic model (Model 1) fits best data from three soil horizons, namely, P_1H_1, P_2H_1 and P_2H_4 while the Brooks and Corey model (Model 3) is best for two horizons, P_1H_2 and P_2H_2. For horizons P_1H_3, P_1H_4 and P_2H_3, both Models 1 and 3 are equally adapted. However, in view of the fact that the soils are texturally similar, one should expect a predominance of one particular model. Further attempts at classification using features 4 and 6 fail to resolve the ambiguity.

The lack of success can be attributed in part, to the inadequacy of the chosen features in discriminating the suitable model; the importance of finding suitable features has already been emphasized earlier. A certain degree of experience and insight is necessary in deriving good features for classification. Another reason is that the present identification has been performed in only two dimensions using one feature per model. By increasing the number of features it should be possible to obtain a more convincing result.

Parameter Identification

Least squares parameter identification is performed with all the three models so as to provide a comparison between the least squares technique and the pattern recognition approach. Results of the least squares fit are presented in Table 3. The sum of squares of deviations (SSD) for Model 2 are markedly higher than those of either Model 1 or Model 3, thus in accord with the pattern recognition results. As with the pattern recognition method, there is no complete dominance of a particular model. One significant feature, however, emerges and that is, the least squares method and the pattern recognition technique can give conflicting results as observed in the case of horizon P_1H_1. The most obvious reason for this is that a few bad observations can increase the SSD significantly, yet these same observations will have little impact on the overall pattern from the whole data set in the pattern recognition approach. This augurs well for the latter technique.

STRUCTURE CHARACTERIZATION OF THE SOIL WATER DIFFUSIVITY

Candidate Models

Early work with diffusivity (Childs and Collis-George, 1950; Klute, 1952) left the functional form of $D(\theta)$ completely general. An exponential function of the form $D(\theta) = \alpha.\exp(\beta\theta)$ was suggested by Gardner and Mayhugh (1958), which since then has been widely used. This can also be written as

$$D(\theta) = D_{min} e^{\beta(\theta - \theta_{min})} \qquad (13)$$

Ahuja and Swartzendruber (1972) suggested a power function of the form

$$D(\theta) = \frac{a\theta^n}{(\theta_s - \theta)^{n/5}} \qquad (14)$$

which yields infinite diffusivity at saturation.

A third model is proposed herein. From the commonly used relationship $K = K_s S^n$ where K is conductivity and subscript s refers to saturation, and the Brooks and Corey model (Eq. 3) for $h < h_a$, from which the above conductivity relationship is derived, we obtain via $D(\theta) = K/(d\theta/dh)$, a diffusivity structure of the form

$$D(\theta) = b\left(\frac{\theta - \theta_r}{\theta_s - \theta_r}\right)^m \qquad (15)$$

where $b = -\frac{K_s h_a}{\lambda(\theta_s - \theta_r)}$ and $m = n - 1 - \frac{1}{\lambda}$

Equation (15) is also a power function but unlike (14), the diffusivity at saturation is finite and equal to b.

The three models represented by (13), (14) and (15), hereby called Model 1, Model 2 and Model 3 respectively, are now considered as the candidate models for the diffusivity functions of the soils under investigation.

Feature Extraction

In contrast to the method employed for the soil water characteristic, here both parameters of each model are used to derive a single feature per model. The feature is defined as the sum of the ratios of each parameter evaluated for a random pair of points from the data set. Thus,

$$\text{Feature 1, } f_1 = \frac{\beta_A}{\beta_B} + \frac{D_{min_A}}{D_{min_B}} \qquad (16)$$

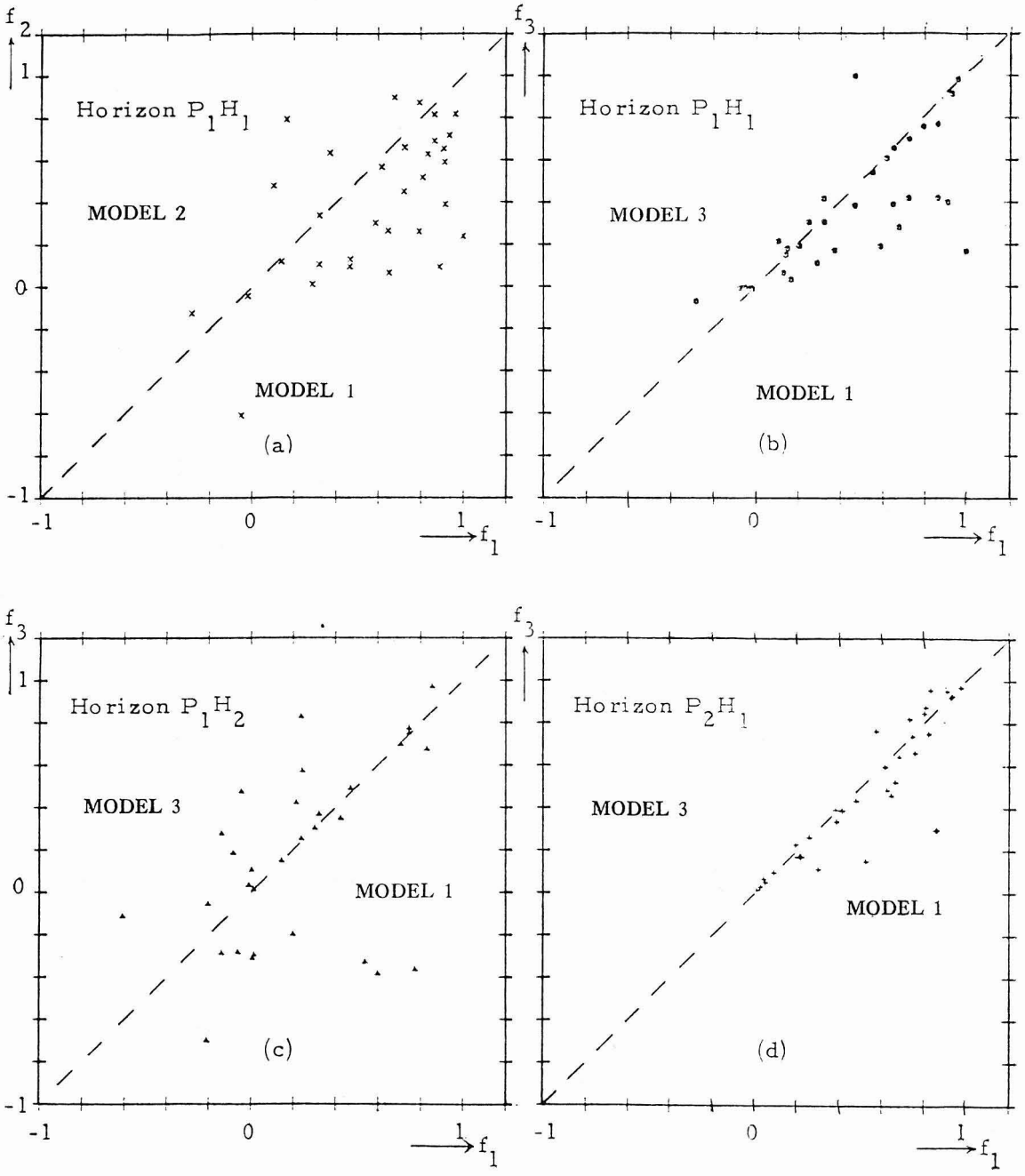

Fig. 4. Feature space containing feature points for various test cases in the structure identification of the soil water characteristic.

$$\text{Feature 2, } f_3 = \frac{n_A}{n_B} + \frac{s_A}{s_B}$$

$$\text{and Feature 3, } f_3 = \frac{m_A}{m_B} + \frac{b_A}{b_B}$$

The various parameters for points A and B are computed in the same manner as described earlier. These are:

$$\beta = \left\{ \frac{1}{D(\theta-\theta_r)} \frac{dD}{d\theta} \right\}^{\frac{1}{2}}, \quad D_{min} = De^{-\beta(\theta-\theta_r)}$$

$$n = 5\theta \left(\frac{\theta_s-\theta}{5\theta_s-4\theta} \right) \frac{dD}{d\theta}, \quad a = \frac{D(\theta_s-\theta)^{n/5}}{\theta^n} \qquad (17)$$

$$m = \frac{\theta-\theta_r}{D} \frac{dD}{d\theta}, \quad b = D \left(\frac{\theta-\theta_r}{\theta_s-\theta_r} \right)^{-m}$$

Whenever the ratio between parameter values at the two points exceeds unity, its reciprocal is taken instead. In this way, all the features remain within the limits of -2 and $+2$.

Training

Five simulation runs are performed with different combinations of parameter values. Since error in the $D(\theta)$ determination is rather large, a 10% relative error is added randomly to the simulated data. The parameter values used for the different simulation runs are indicated in Table 4.

Figures 5 (a) & (b) show the plots of feature points for the classification between Model 1 and Model 2 and between Model 1 and Model 3. Again some overlapping is observed, nevertheless, the splitting of the feature space into two subsets corresponding to the respective models is quite evident. (Note: none of the features took negative values in the training operation).

Identification of the Structure for Soil Water Diffusivity

The soil water diffusivity-water content relationships of various horizons of the two profiles of the Bungor series, whose structure are to be identified are shown in Fig. 6. Using data from each horizon, features are calculated according to (16) and (17) and plots of feature points, f_1-f_2 and f_1-f_3 are prepared. Figures 7 (a) and (b) illustrate, respectively, the plots obtained with P_1H_1 and P_2H_1. From these figures it is clear that the exponential function fits the data best. A similar trend is also observed with all other horizons.

Exponential functions are now fitted to the data from the various horizons and the results presented in Table 5. Even though high correlation coefficients are obtained in all cases it is

TABLE 3

Results of parameter identification of Model 1, 2 and 3 by the least squares method

Soil horizon	θ_s	θ_r	Model 1 α	Model 1 β	Model 1 SSD$\times 10^2$	Model 2 δ	Model 2 γ	Model 2 SSD$\times 10^2$	Model 3 h_a	Model 3 λ	Model 3 SSD$\times 10^2$
PROFILE 1											
P_1H_1	0.54	0.01	75.61	2.764	1.796	85.5	0.932	9.169	− 6.03	0.242	1.123
P_1H_2	0.55	0.01	321.3	3.429	2.327	176.2	0.986	8.701	−12.05	0.249	1.252
P_1H_3	0.57	0.02	289.3	3.342	1.314	135.3	0.925	5.978	−15.05	0.263	2.843
P_1H_4	0.575	0.025	667.0	3.718	1.313	154.4	0.895	5.220	−22.40	0.249	2.144
PROFILE 2											
P_2H_2	0.59	0.02	45.65	2.389	1.923	32.4	0.700	9.484	− 4.87	0.209	2.225
P_2H_2	0.57	0.03	236.3	3.242	2.817	111.6	0.898	9.056	− 9.90	0.235	2.617
P_2H_3	0.57	0.03	394.1	3.477	1.488	105.6	0.843	5.633	−17.70	0.271	2.580
P_2H_4	0.565	0.03	551.2	3.596	1.474	98.6	0.790	5.745	−23.35	0.274	2.078

SSD = Sum of squares of deviation $= \sum_{i=1}^{n} (S_i - S_i^m)^2$ where superscript m refers to measured value.

TABLE 4
Parameter values used to generate data needed in the classification of the feature space for the different diffusivity models

Simulation run	Model 1		Model 2		Model 3	
	Dmin	β	a	n	b	m
1	1×10^{-4}	60.0	15.0	2.0	5.0	−6.0
2	1×10^{-5}	50.0	12.0	4.0	0.0001	4.0
3	1×10^{-6}	30.0	4.0	3.5	0.01	10.0
4	1×10^{-7}	20.0	1.0	4.0	0.1	7.0
5	1×10^{-8}	40.0	8.0	3.0	1.0	−2.0

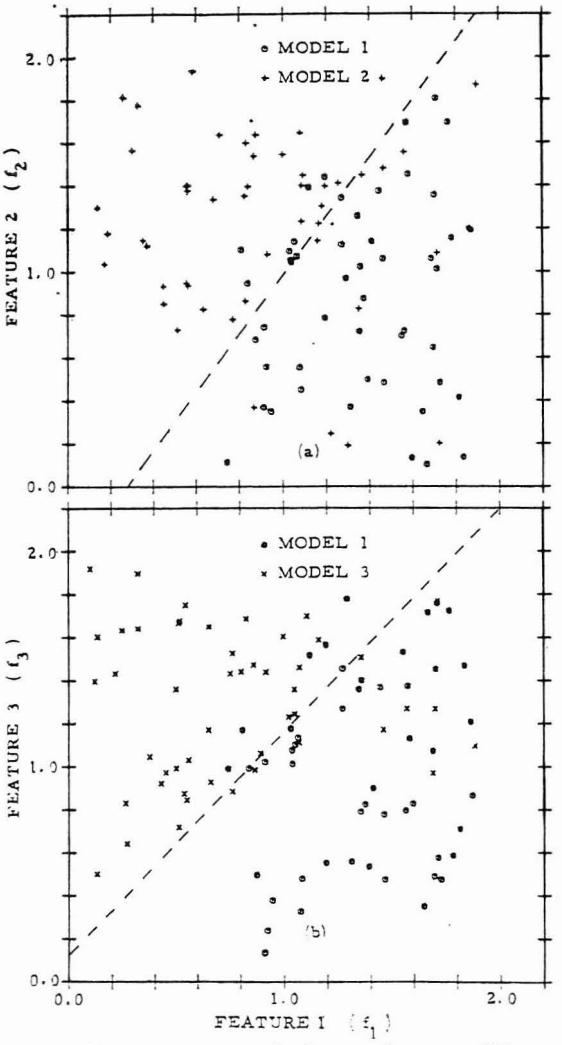

Fig. 5. Feature space of the various candidate models for the soil water diffusivity. (a) Model 1 and Model 2. (b) Model 1 and Model 3.

found that the derived functions underpredict D(θ) values at and near saturation by as much as 30%. This definitely will affect the accuracy of the solutions to the water flow equation. For improved accuracy it would be preferable to use

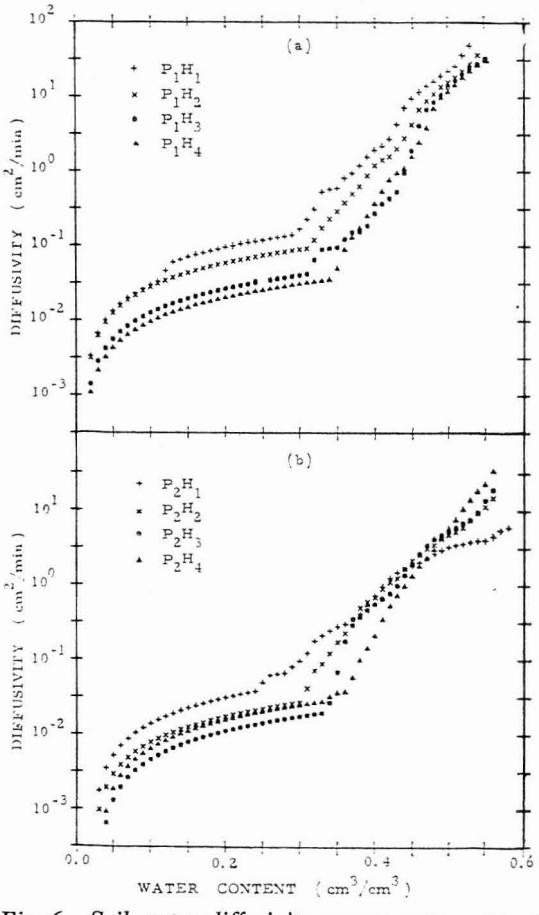

Fig. 6. Soil water diffusivity versus water content for different horizons of (a) Profile 1 and (b) Profile 2 of Bungor Series. P with subscript refers to profile number while H with subscript refers to the horizon number.

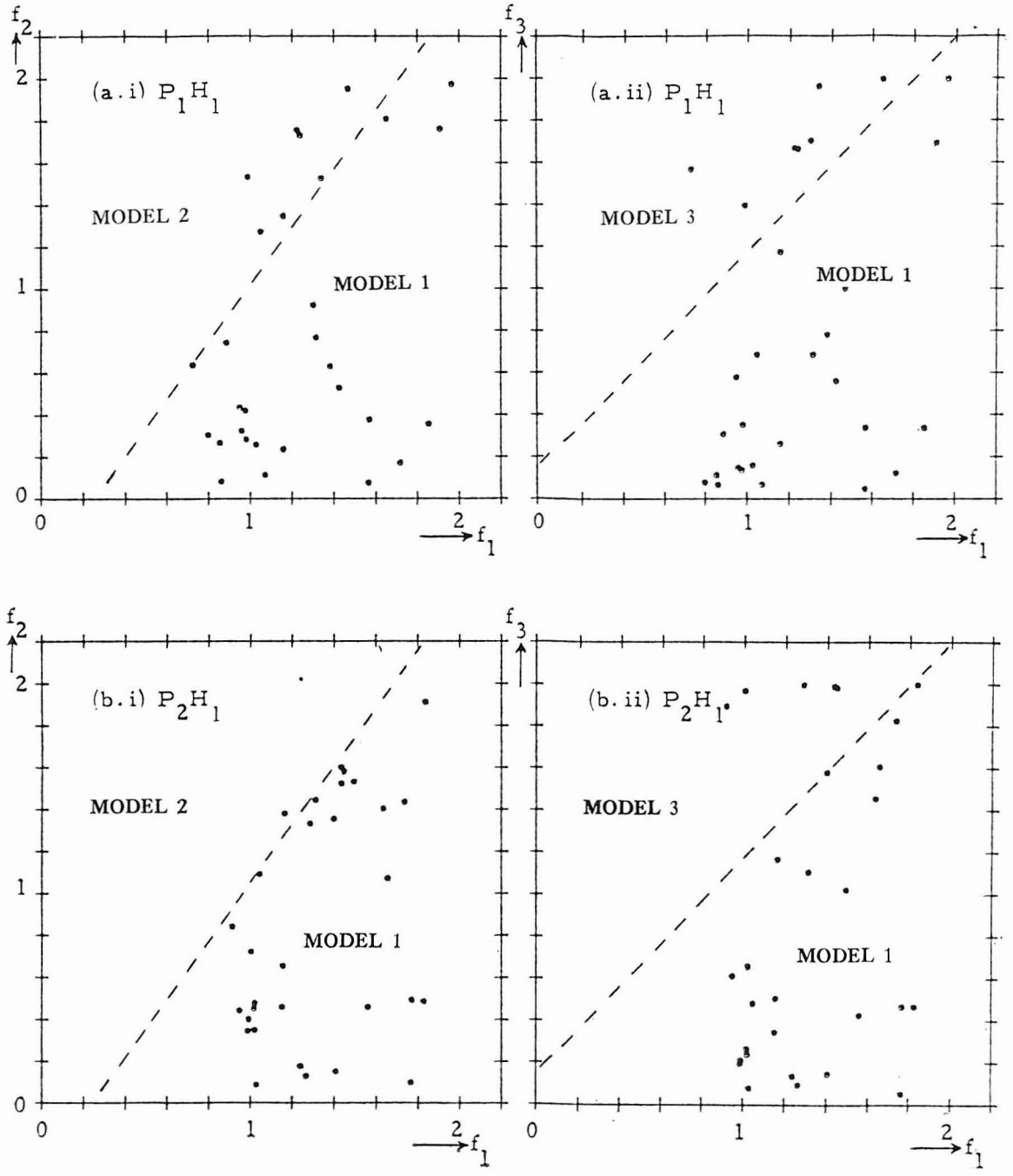

Fig. 7. Feature space containing feature points for various test cases in the structure identification of soil water diffusivity.

TABLE 5

Soil water diffusivity $D(\theta)$, expressed as a unique and piecewise exponential function. R is the correlation coefficient for the linear regression $\ln D = \ln D_{min} - \beta(\theta - \theta_{min})$

Soil horizon	Unique Function		Piecewise Function	
	Diffusivity function (cm^2/min)	R	Diffusivity function (cm^2/min)	R
PROFILE 1				
P_1H_1	$D = 2.928 \times 10^{-3} \exp[11.473(\theta-0.01)]$	0.975	$D = \begin{cases} 2.468 \times 10^{-3} \exp[13.345(\theta-0.01)], & 0.01 < \theta < 0.356 \\ 7.186 \times 10^{-5} \exp[23.553(\theta-0.01)], & 0.356 < \theta < 0.54 \end{cases}$	0.945 0.996
P_1H_2	$D = 1.157 \times 10^{-3} \exp[15.311(\theta-0.01)]$	0.960	$D = \begin{cases} 2.750 \times 10^{-3} \exp[4.761(\theta-0.01)], & 0.01 < \theta < 0.335 \\ 1.459 \times 10^{-5} \exp[25.881(\theta-0.01)], & 0.335 < \theta < 0.55 \end{cases}$	0.923 0.998
P_1H_3	$D = 4.389 \times 10^{-4} \exp[16.08\ (\theta-0.02)]$	0.940	$D = \begin{cases} 1.086 \times 10^{-3} \exp[11.288(\theta-0.02)], & 0.02 < \theta < 0.352 \\ 6.354 \times 10^{-6} \exp[26.310(\theta-0.02)], & 0.352 < \theta < 0.57 \end{cases}$	0.927 0.973
P_1H_4	$D = 3.074 \times 10^{-4} \exp[16.592(\theta-0.02)]$	0.940	$D = \begin{cases} 7.972 \times 10^{-4} \exp[11.747(\theta-0.02)], & 0.02 < \theta < 0.352 \\ 4.689 \times 10^{-6} \exp[26.779(\theta-0.02)], & 0.352 < \theta < 0.56 \end{cases}$	0.929 0.976
PROFILE 2				
P_2H_1	$D = 9.467 \times 10^{-4} \exp[14.054(\theta-0.02)]$	0.991		
P_2H_2	$D = 3.277 \times 10^{-4} \exp[16.780(\theta-0.02)]$	0.936	$D = \begin{cases} 6.096 \times 10^{-4} \exp[14.012(\theta-0.02)], & 0.02 < \theta < 0.357 \\ 4.718 \times 10^{-5} \exp[21.600(\theta-0.02)], & 0.357 < \theta < 0.57 \end{cases}$	0.934 0.989
P_2H_3	$D = 1.765 \times 10^{-4} \exp[18.235(\theta-0.03)]$	0.963	$D = \begin{cases} 5.198 \times 10^{-4} \exp[10.869(\theta-0.03)], & 0.03 < \theta < 0.32 \\ 4.321 \times 10^{-6} \exp[27.408(\theta-0.03)], & 0.32 < \theta < 0.57 \end{cases}$	0.926 0.981
P_2H_4	$D = 2.730 \times 10^{-4} \exp[16.692(\theta-0.03)]$	0.942	$D = \begin{cases} 7.802 \times 10^{-4} \exp[10.111(\theta-0.03)], & 0.03 < \theta < 0.356 \\ 9.024 \times 10^{-7} \exp[30.744(\theta-0.03)], & 0.356 < \theta < 0.57 \end{cases}$	0.924 0.993

a piecewise exponential function instead. As shown in Table 5, in the region of higher water content the correlation coefficient for the latter is much higher than for the unique exponential function; soil P_2H_1 is an exception in which a unique function is already satisfactory, hence a partition is unnecessary.

CONCLUSIONS

The pattern recognition approach has been applied to the structure characterization of two soil properties, namely, the soil moisture characteristic and the soil water diffusivity. While the method was able to delineate the most suitable model for the diffusivity (i.e. an exponential function) from three possible candidate models, classification in a two-dimensional feature space achieved limited success in the structure identification of the soil moisture characteristic. Two of the candidates, an asymptotic model (Model 1) and the Brooks and Corey model (Model 3) were equally adapted to the data. Better discrimination could probably be obtained by using better and/or more features, the latter entailing classification in multi-dimensional space.

The training stage of the procedure admittedly entails heavy computations; however, once this stage is complete, processing of each data set requires minimal effort. In fact, it took less than 0.5 hour of computation time on the CDC 1700 (32K) to process the diffusivity data and to plot the graphs of f_1-f_2 and f_1-f_3 for all the 8 test cases. Together with the fact that the technique can give different results from the least squares method leads to the conclusion that the pattern recognition approach provides us with a useful technique in the identification of model structures of poorly defined systems.

ACKNOWLEDGEMENTS

The granting of study leave by the Universiti Pertanian Malaysia and the financial support of the Algemeen Bestuur van de Outwikkelingssamenwerking (A.B.O.S.) for the study are gratefully acknowledged.

REFERENCES

AHUJA, L.R. and SWARTZENDRUBER, D. (1972): An improved form of soil-water diffusivity function. *Soil Sci. Soc. Amer. Proc.* **36,** 9–14.

BROOKS, R.H. and COREY, A.T. (1964): Hydraulic Properties of porous media. *Hydrol. Pap.* **3.** Colo. State Univ. Fort Collins.

CHILDS, E.C. and COLLIS-GEORGE, N. (1950): The permeability of porous materials. *Proc. Roy. Soc.* **201***A*, 392–405.

GARDNER, W.R. and MAYHUGH, M.S. (1958): Solution and tests of the diffusion equation for the movement of water in soil. *Soil Sci. Soc. Amer. Proc.* **22,** 197–201.

HAVERKAMP, R., VAUCLIN, M., TOUMA, J., WIERENGA, P.J. and VACHAUD, G. (1977): A comparison of numerical simulation models for one-dimensional infiltration. *Soil Sci. Soc. Amer. Proc.* **41,** 285–294.

KANAL, L. (1974): Patterns in pattern recognition 1968–1974. *IEEE Trans. on information theory.* November, IT 20.

KARPLUS, W.J. (1972): System identification and simulation, a pattern recognition approach. *Proc. Fall Joint Compt. Conf.* 385–392.

KLUTE, A. (1952): A numerical method for solving the flow equation for water in unsaturated materials. *Soil Sci.* **73,** 105–116.

NILSSON, N.J. (1965): Learning machines – Foundations of trainable pattern-classifying systems. New York. McGraw Hill.

SARIDIS, G.N. and HOFSTADTER, E.F. (1974): A pattern recognition approach to the classification of non-linear systems. *IEEE Trans. on Systems, Man and Cybernatics* SMC-4, 362–370.

SIMUNDICH, T.M. (1975): System characterization: a pattern recognition approach. Ph.D. Thesis. UCLA. School of Engineering and Applied Sciences. Los Angeles.

VANSTEENKISTE, G.C., BENS, J. and SPRIET, J. (1978): Design of a linear classifier using simulation techniques. *Proc. United Kingdom Simulation Conf.*, May (in press).

VANSTEENKISTE, G.C., BENS, J., and SPRIET, J. (1978a): Structure characterization for system modeling in uncertain environments. *Proc. Symp. on System Simulation Methodology, Rehovot, Israel.* Aug. 1978.

WAN SULAIMAN WAN HARUN (1979): Simulation of water movement in the Bungor Series (Typic Paleudult) in relation to soil erosion. Doctorate Thesis (unpubl.). State Univ. Gent.

YOUNG, T.Y. and CALVERT, T.W. (1974): Classification, estimation and pattern recognition. New York. American Elsevier.

(Received 1 August 1980)