0038-075X/95/1603-0218$03.00/0
Soil Science

September 1995
Vol. 160, No. 3
Printed in U.S.A.

PRECISION OF PREDICTING SOIL SALINITY BASED ON VEGETATION CATEGORIES OF ABANDONED LANDS

TIBOR TÓTH,[1] SATOSHI MATSUMOTO,[2] RENZHAO MAO,[3] AND YANFENG YIN[3]

Abandoned saline lands on the Huang-Huai-Hai Plain of China are occasionally used for crops. Our objective was to determine how precisely the values of soil variables, such as ion concentrations, pH, and penetration resistance, could be predicted to fall into a particular range of values based on the presence of specific categories within the semi-natural vegetation.

The correlation between sets of soil (mostly chemical) and plant (cover percentage) variables in a 100 × 220-m plot was examined by calculating canonical variates. The correlation between the first canonical variates of the two sets was 0.88, and indicated the presence of a strong relationship between soil and vegetation.

The semi-natural vegetation of the site was classified either as dominated by *Phragmites communis* or by *Imperata cylindrica* plant species. In this plot, these two categories represent overlapping ranges of several soil variables. On average, the PHRAGMITES category occurs at a higher elevation and has a greater salt concentration and smaller penetration resistance and pH than does the IMPERATA category. Based on our observations, leaching does not affect the concentration of bicarbonate ions in the surface soil layers. However, as the concentration of other ions decreases and the ratio of bicarbonate to other ions increases, pH and penetration resistance increase in some patches.

The degree of separability of different soil variables by the vegetation categories was inferred from (i) the significance of the difference of their means using the Mann-Whitney test and (ii) from the correlation between the original variables and the linear discriminant function of the two vegetation categories based on the soil variables. The precision of predicting ranges of individual soil variables from vegetation categories was determined by one-dimensional thresholding and ranged from 67 to 84%. Misclassification by vegetation categories into the range assigned to the other vegetation category occurred at least twice as often at the borderlines between the vegetation categories than inside each category. Vegetation categories are considered an easy and economic way of determining the likely success of cropping in saline areas.

The semi-natural vegetation (i.e., the constant vegetation established from naturally occurring plants after ceasing cultivation for several years) of saline soil is in near-equilibrium with the soil variables. Within each area the occurrence and abundance of the particular species are governed by abiotic (such as soil and relief) and biotic (competition among plants) factors. The established vegetation is in a probabilistic relationship with the soil variables. The formalization of these probabilistic relationships may be used to predict soil variables both at the level of an individual plant and at that of a vegetation category. Because extreme soil salt concentration and pH are important factors ecologically, salt-affected types of soil are especially suitable for soil-vegetation correlation studies (Tóth and Rajkai 1994). Consequently, natural vegetation is often used worldwide for mapping salt-affected soils (Shantz 1911; Kearney et al. 1914; Shantz and Piemeisel 1924; Ballenegger 1929; Sigmond 1927; Magyar 1928; Bodrogközy 1965; Zonneveld 1976). This use of vegetation as an indicator of soil properties is rapid and economic, and it is even more effective if the precision with which the distinguished vegetation categories can predict the ranges of different soil variables is known. The present aim is to determine the degree of that precision.

Much of the vast area of saline soil in the Huang-Huai-Hai Plain of Northern China is planted only infrequently. From time to time, demographic pressure necessitates the cultivation for crop production of these abandoned lands. Local farmers select areas for cultivation based on elevation and the dominant plant species. This study focuses on various aspects of predicting the ranges of selected soil variables using vegetation categories. Another study describes the results of using individual plant cover percentage data and other field-measured variables in the quantitative prediction of selected soil variables on the same site (Tóth et al. 1994).

The use of vegetation for predicting ranges of soil variables is promising when the sets of vegetation and of soil variables show close correlation. The correlation between vegetation species and soil properties was determined by canonical correlation analysis. Differences in predictability between soil variables were concluded from the Mann-Whitney test and from discriminant analysis. The precision of predicting separate ranges of soil variables was expressed as a percentage in cross-tabular form.

[1] Research Institute for Soil Science and Agricultural Chemistry of the Hungarian Academy of Sciences, MTA TAKI H-1022, Budapest II, Herman O. út 15, Hungary. Dr. Tóth is corresponding author.

[2] Tokyo University, Tokyo, Japan.

[3] Shijiazhuang Institute of Agricultural Modernization, Shijiazhuang, China.

Received April 21, 1994; accepted May 18, 1995.

RESEARCH AREA

The study was carried out in Wangsu, Hebei Province, China (lying at 38'25"N and 117'16"E). The area is 60 km from the Bohai Bay (Yellow Sea). The parent material of the soil is stratified transported loess. The soil was characterized by the description of a profile and was classified as pale meadow soil according to the Chinese genetic classification, which corresponds to a Haplaquoll in the USDA classification system. The dominant cations in the area are Na^+ and Mg^{2+}, and the dominant anions are Cl^- and SO_4^{2-} (Table 2). The origin of the saline soils of the area is related to the recent formation of the alluvial plain associated with the retreat of the Yellow Sea, the transport of salts by rivers, and the present fluctuation in the groundwater level. The study site is 7 m above sea level. The site and the surrounding 5000-ha area are used by local authorities as a flood water reservoir area to collect for later use some of the river floods that arise during the heavy rains of June and July or November and December. This use began at the end of the 1970s. Recent records of the area show that it was cropped about five times during the last half century. Within the site there were abandoned channels and dams as well as areas of cropland. According to observations during the past 2 years, farmers tend to crop the same patches with winter wheat and cotton. The local maps (1:10,000) of soil salt concentration are based on the 0 to 20-cm layer and distinguish the following classes: nonsaline, (<0.1% salt, with a wheat yield reduction of less than 10%), slightly saline (0.1–0.2% salt, with a 10–30% yield reduction of wheat), medium saline (0.2–0.4% salt, with a 30–50% yield reduction of wheat), severely saline (0.4–0.6% salt, with more than 50% yield reduction of wheat), and very severely saline (>0.6% salt, with no wheat growth).

PLANT COVER

In the abandoned salt-affected lands covered by semi-natural vegetation the plants are the same as at the seashore. The dominant plants found in the area are generally widespread in temperate-subtropical zones. *Imperata cylindrica* is known to be a weed of many continents (Girard and Isarwa 1990). *Phragmites communis* is certainly one of the most common vascular plants, and *Imperata cylindrica* and *Aeluropus littoralis* have been reported by Chapman (1960) to be growing on drying banks of slightly saline marshes in southern Iraq. The potential of the plants to indicate patches with lower salt concentration and, therefore, to be more suitable land for wheat is also known to the farmers. They prefer places covered with *Agropyron dahuricus* and *Imperata cylindrica* and do not like to sow wheat in areas covered mostly with *Aeluropus littoralis*, *Scorzonera albicaulis* or *Suaeda salsa*.

MATERIALS AND METHODS

Experimental layout and field measurements

A simplified version of a comprehensive map of the land use and semi-natural and weed vegetation on the 100 × 220-m site is shown in Fig. 1.

We chose a block or quadrat size of 5 × 5 m, corresponding to the minimal area of economic importance for cropping. The centers of the blocks were arranged on a regular grid with a 5 m lag. They did not fill all the possible 945 (21 × 45) grid points, but were arranged randomly at only 206 grid points. Based on the characterization of the vegetation in quadrats and the internal heterogeneity of these blocks, the soil was sampled at four points arranged regularly inside each block and then bulked. An auger with a diameter of about 3 cm was used; therefore, an approximately 200-cm^3 soil sample was available for each depth. Because gradients of salt concentration and pH can be found in small depth increments in

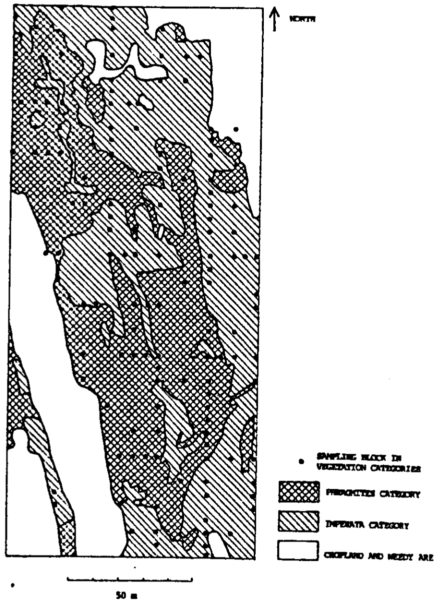

FIG. 1. Vegetation map of the study site.

soils affected by salinization, samples were taken from thin layers, at depths of 0 to 5 cm, 20 to 25 cm, and 40 to 45 cm. Using a theodolite, elevation was measured at each soil sampling point by leveling. Plant coverage was simply estimated. Penetration resistance was measured at 16 points in each block, i.e., four points were measured around each soil sampling point, using a pocket, calibrated-spring, hand-operated, direct-reading cone penetrometer. The resistance was expressed as mm of penetration of the cone, which had a diameter of about 1 in.

Analytical methods

Soil samples were used to prepare 1:5 (by weight) soil:water extracts. From these, the pH (at 0–5 cm only) was measured potentiometrically. Carbonate and bicarbonate ion concentrations were measured by titration with 0.01 *N* H_2SO_4 using phenolphthalein and methyl orange indicators, chloride ion concentration was measured by argentometry, sulfate was measured by iodometry (Lukács and Rédly 1988), and the calcium and magnesium ion concentrations were titrated with EDTA using a method essentially the same as that of Lanyon and Heald (1982). This consisted of titrating calcium with a Patton-Reeder indicator, calcium together with magnesium with an eriochrome black-T indicator, after which magnesium was calculated as the difference. Sodium ion concentration was calculated as the difference between the hypothetical cation concentration (assumed to be the same as the sum of the anions measured) and the sum of measured calcium and magnesium ion concentrations. Consequently, sodium ion concentration also included potassium and possibly some other ions.

Statistical analysis

The statistical analyses were performed on data from blocks classified as having semi-natural vegetation. The correlation between the soil and plant variables was determined by canonical correlation analysis (Digby and Kempton 1987). Discriminant analysis (Marsal 1987) was used to assess the ability of different soil variables to distinguish between vegetation categories.

RESULTS AND DISCUSSION

The correlation between plant and soil variables

A close correlation between soil and plant variables is required if vegetation categories are to be used successfully to predict soil variables. Therefore, the strength of this relationship and the variables that contribute most to it are of interest. The relationship was assessed by canonical correlation analysis, in which pairs of derived variates (X_i, Y_i) from the sets of plant (x_1–x_{13}) and soil variables (y_1–y_{26}) were calculated. The value of the i-th canonical variate (plant) for the k-th observation is received from the 13 plant variables as,

$$X_{i,k} = a_{1,i}x_{1,k} + a_{2,i}x_{2,k} + \ldots + a_{13,i}x_{13,k}$$

$$i = 1,2,3, \ldots \min(13,26).$$

Similarly the value of i-th canonical variate (soil) for the k-th observation from the 26 soil variables is calculated as,

$$Y_{i,k} = b_{1,i}y_{1,k} + b_{2,i}y_{2,k} + \ldots + b_{26,i}y_{26,k}$$

$$i = 1,2,3, \ldots \min(13,26),$$

where $a_{1,i} \ldots a_{13,i}$ and $b_{1,i} \ldots b_{26,i}$ are the coefficients for calculating i-th canonical variates from

original variables. The calculations are optimized in such a manner that the resulting canonical variates showed the highest possible correlation pairwise.

$$R(X_1;Y_1) >= R(X_2;Y_2) >= \ldots >=$$

where R(X;Y) is the correlation between X and Y.

We calculated only one pair of new variables, the so-called first canonical variates for soil properties and for plant variables, by multiplying the original variables with the coefficients, $a_1 \ldots a_{13}$ and $b_1 \ldots b_{26}$. The correlation between the first pair of canonical variates was 0.88, suggesting a strong likelihood that vegetation could be used for predicting soil variables. Since the correlations of the original variables with the canonical variates are informing on the importance of the respective variables in the relationship between the two sets of variables, these are shown in Table 1.

The cover of three plant species had large correlation coefficients with the first canonical variate created from the plant variables (Table 1): the cover percentage of *Phragmites communis* and *Aeluropus littoralis* showed a positive correlation, and the cover percentage of *Imperata cylindrica* showed a negative correlation.

Many soil variables showed a strong correlation with the first canonical variate created from the soil variables (Table 1). A positive correlation was shown by the concentrations of cations, chloride, sulfate, soluble salts, and elevation; a negative correlation was shown by pH, bicarbonate concentration, and penetration resistance. The same soil variables determined from different depths did not differ in the sign and magnitude of correlation. These correlations indicated the importance of the respective soil variables in the existence of different types of semi-natural vegetation on the site studied. Blocks from the PHRAGMITES category (see below) were characterized by high values for those variables that showed positive correlation with both first canonical variates (soil and plant). Blocks from the IMPERATA category exhibited high values for those variables that showed negative correlation with the same.

The vegetation categories distinguished

When setting up the vegetation categories, the governing principles were the presence of dominant and codominant species and the state of the

TABLE 1

Correlation of first canonical plant and soil variates with original variables, respectively

Variable	Correlation coefficient
Total plant cover	0.210
Phragmites communis	0.801
Scorzonera albicaulis	0.124
Aeluropus littoralis	0.584
Imperata cylindrica	−0.651
Agropyron dahuricus	0.031
Puccinellia chinampoensis	0.009
Suaeda salsa	−0.170
Messerschmidia sibirica	0.057
Melilotus sp.	−0.014
Artemisia annua	0.239
Lactuca tatarica	−0.335
Cirsium segetum	−0.172
Soil surface	
Elevation	0.728
Penetration	−0.458
0–5-cm depth	
pH	−0.371
HCO_3	−0.615
Cl	0.761
SO_4	0.862
Ca	0.738
Mg	0.709
Na	0.836
Sum	0.821
Na %	0.348
20–25-cm depth	
HCO_3	−0.199
Cl	0.678
SO_4	0.549
Ca	0.118
Mg	0.509
Na	0.770
Sum	0.659
Na %	0.569
40–45-cm depth	
HCO_3	−0.457
Cl	0.821
SO_4	0.702
Ca	0.354
Mg	0.578
Na	0.811
Sum	0.794
Na %	0.623

pH is the pH of the 1:5 soil:water extract of the 0 to 5-cm layer; HCO_3 is the bicarbonate concentration of the 1:5 soil:water extract (cmol$_c$ kg^{-1}); Cl is the chloride concentration of the 1:5 soil:water extract (cmol$_c$ kg^{-1}); SO_4 is the sulfate concentration of the 1:5 soil:water extract (cmol$_c$ kg^{-1}); Ca is the calcium concentration of the 1:5 soil:water extract (cmol$_c$ kg^{-1}); Mg is the magnesium concentration of the 1:5 soil:water extract (cmol$_c$ kg^{-1}); Na is the sodium concentration of the 1:5 soil:water extract (cmol$_c$ kg^{-1}); Sum is the salt content (%); Na% is the weight percentage of sodium ion relative to the sum of the ions.

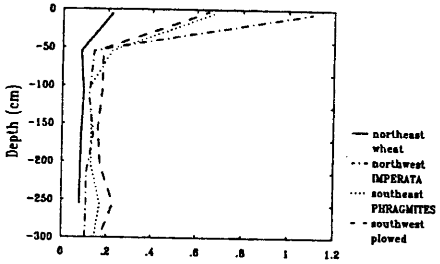

FIG. 2. Depth distribution of soil salt concentration at four points placed regularly from the center toward the corners of the site. Legend shows direction of corner and vegetation category of point.

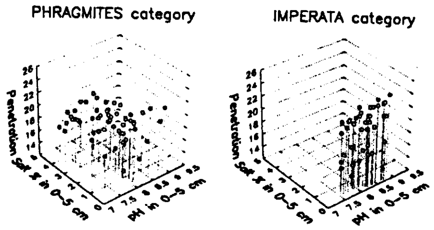

FIG. 3. Scatterplot of surface soil penetration resistance, pH, and salt concentration of the 0 to 5-cm layer in the vegetation categories.

vegetation (grazed, plowed, etc). Soil variables, such as salt effervescence, were not taken into account. The system consisted of one category of actual or recent cropland, and two categories of semi-natural vegetation. The criteria for designating an area as recent cropland were the signs of cropping and the presence of weeds such as *Messerschmidia sibirica* and *Setaria viridis.* The two categories of semi-natural vegetation were decided on the basis of the dominant species, either *Phragmites communis*, that is the PHRAGMITES category, or *Imperata cylindrica*, that is the IMPERATA category. To characterize the occurrence of certain plant species, Table 2 shows the cover of the most common ones in the two semi-natural categories. Only four species had mean cover larger than 1%. Besides the dominant species of the categories, i.e., *Phragmites communis* and *Imperata cylindrica*, no more than two plant species had a significantly different mean cover between the two categories, these were *Aeluropus littoralis* and *Lactuca tatarica*. These two species and *Puccinellia chinampoensis*, *Suaeda salsa*, *Scorzonera albicaulis*, *Tamarix chinensis*, *Limonium bicolor*, and *Heteropappus aetaicus* are considered indicators of high salt concentration in the soil. The other species in Table 2 are common weeds in the Huang-Huai-Hai Plain.

The distribution of the soluble salt concentration with depth in the soil profile was determined by augering at four locations approximately halfway from the center of the site to the four corners (Fig. 2). These locations also represented different vegetation categories. As a result of the preferential selection by the farmers of croplands on the lowest places, and of occasional irrigation, the topsoil samples from the cultivated locations had a small salt concentration. Salt concentration had the highest value in the surface soil layer. Below a depth of about 50 cm, the salt concentration does not vary much at 0.1 to 0.2%. The salt concentration in the surface layer of the soil profile shows great changes during the year, possibly increasing or decreasing as much as twofold, as was shown by Inanaga (1991). He reported salt concentrations in the 0 to 30-cm layer ranging from 0.3 to 0.9% throughout a 2-year observation period on a site some 20 km to the west in a less saline area not affected by periodic inundations.

The scatterplots of some variables, namely elevation and penetration resistance, as well as the pH and soluble salt concentration in the 0 to 5-cm layer, showed differences between the two vegetation categories (Fig. 3). In general, soil salt concentration and pH showed a negative correlation. With increasing elevation, the salt concentration also increased. pH and penetration resistance showed a positive correlation. Elevation and penetration resistance showed a negative correlation, as did elevation and pH in both vegetation categories. Inside the PHRAGMITES category, the salt concentration was highly variable and the penetration was usually small. Inside the IMPERATA category, salt concentration was small, and soil penetration was variable. For PHRAGMITES category the correlations of the soil variables were less than for the IMPERATA category, but the plant variables showed similar strength of correlation in the two categories. Correlation coefficients calculated between these variables are given in Table 5.

For the PHRAGMITES category, the samples had a small bicarbonate concentration and sum

TABLE 2

Means and mean ranks (MR) of the variables in the Mann-Whitney test calculated between the two vegetation categories

Variable	PHRAGMITES category (n=74)		IMPERATA category (n=54)		
	Mean	MR	Mean	MR	2-tailed P
Plant cover %					
Total	12.58	73	10.18	54	0.0041
Phragmites communis	7.30	86	2.39	35	<0.00005
Scorzonera albicaulis	0.09	65	0.04	64	0.4476
Aeluropus littoralis	1.74	77	0.08	47	<0.00005
Imperata cylindrica	1.51	42	5.80	95	<0.00005
Agropyron dahuricus	0.05	65	0.04	64	0.7543
Puccinellia chinampoensis	0.07	65	0.11	64	0.8270
Suaeda salsa	1.11	64	0.60	65	0.8923
Messerschmidia sibirica	0.01	65	0	64	0.3930
Melilotus sp.	0.19	65	0.06	64	0.8857
Artemisia annua	0.01	64	0.01	65	0.8310
Lactuca tatarica	0.28	60	0.60	70	0.0612
Cirsium segetum	0	63	0.06	67	0.0410
Soil surface					
Elevation	335.97	82	184.76	41	<0.00005
Penetration	19.42	53	21.02	80	<0.00005
0–5-cm depth					
pH	8.00	44	8.20	59	0.0087
HCO_3	0.27	37	0.38	68	<0.00005
Cl	21.02	70	2.57	29	<0.00005
SO_4	15.44	70	4.05	29	<0.00005
Ca	5.01	69	1.75	31	<0.00005
Mg	13.01	70	2.10	29	<0.00005
Na	18.96	71	2.85	28	<0.00005
Sum	2.20	71	0.43	29	<0.00005
Na %	19.31	60	15.17	40	0.0005
20–25-cm depth					
HCO_3	0.33	60	0.36	71	0.1063
Cl	3.57	83	1.80	40	<0.00005
SO_4	2.27	82	1.36	41	<0.00005
Ca	0.87	70	0.71	57	0.0412
Mg	1.89	80	1.05	43	<0.00005
Na	3.41	83	1.76	38	<0.00005
Sum	0.37	82	0.22	40	<0.00005
Na %	21.22	78	17.80	46	<0.00005
40–45-cm depth					
HCO_3	0.33	52	0.38	81	<0.00005
Cl	2.10	87	0.70	34	<0.00005
SO_4	1.58	84	0.82	38	<0.00005
Ca	0.52	75	0.41	50	0.0002
Mg	1.19	84	0.63	38	<0.00005
Na	2.31	86	0.96	34	<0.00005
Sum	0.25	86	0.12	35	<0.00005
Na %	21.03	81	17.13	42	<0.00005

Because of rain during sampling, the number of blocks evaluated for the soil chemical variables of the 0 to 5-cm layer was 54 in the PHRAGMITES category, and 47 in the IMPERATA category.

of ions weight ratio, moderate pH values, and smaller penetration. These samples were collected at higher elevations. Figure 4 shows that because the PHRAGMITES category was typical on the ridges and highest places, the IMPERATA category occupied the depressions of the site.

In the area studied, the factors accounting for the differences between the two vegetation categories seemed to be elevation, salt concentration, and the concentration of ion(s) with alkaline hydrolysis. The only such anion was bicarbonate inasmuch as carbonate was not found in the water extract samples. Because of more limited solubility of $NaHCO_3$ (8.7 g/100 g aqueous solution at 20°C) relative to NaCl (26.4 g), Na_2SO_4 (16.1g), $MgCl_2$ (41 g), $MgSO_4$ (25.2g), and $CaCl_2$ (42.7g), the bicarbonate anion is the one that is most difficult to leach. Consequently, its concentration relative to the sum of the ions increased after leaching. When the weight percentage of bicarbonate to the sum of the ions was about 1%, the pH was no more than about 8, but when it was about 50%, the pH was about 9. The higher lying sites are less leached than the lower lying ones because of the effect of accumulated water in the lower areas. Because leaching does not greatly affect the concentration of bicarbonate ions, the ratio of bicarbonate ions to the sum of anions is greater in the lower lying sites. The resulting rise in pH is added to the effects of small ion concentration and the dominance of sodium cations. These conditions favor increased dispersion and cause the formation of a compact layer (Rengasamy and Olsson 1991), which is shown by increased penetration resistance. Within the site studied, at present only a limited number of blocks had a pH greater than the diagnostic criterion for sodic soil, pH 8.5 (Fig. 3); therefore, the IMPERATA category is more suitable for crops than the PHRAGMITES category.

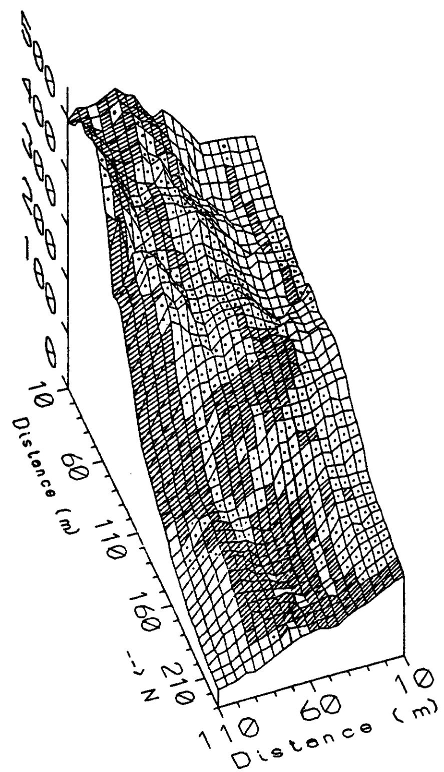

FIG. 4. Map of relief of the studied site. PHRAGMITES category is indicated by a dot (•), IMPERATA category by lines (≡), and CROPLAND category by none. Distances are in meters and elevation is in millimeters.

The separability of the vegetation categories

We compared the means of the variables in order to see which of these were different in the two vegetation categories. Because the distributions of the values of the plant cover data and some soil properties were not normal (see Fig. 5), a nonparametric method, the Mann-Whitney U test, was used to test the equality of the means of soil and plant variables in the two vegetative categories (Davis 1986). This test is based on sorting all observations in order, from smallest to largest, and giving each a rank. Then the ranks of the observations are compared with the test statistic U, of variable *a*, which was calculated as

$$U_{a,PHRAGMITES} = 74 \times 54 + ½54*55 - R_{a,IMPERATA}$$

and

$$U_{a,IMPERATA} = 74 \times 54 + ½74 \times 75 - R_{a,PHRAGMITES}$$

where R is the sum of the ranks for the given vegetation category. The smaller of the two Us is then compared with the significance levels belonging to the number of observations (74 and 54).

The means, mean ranks, and the significance levels for the differences between the studied variables are given in Table 2. It shows that only the plants used as classificators of the blocks, i.e., *Imperata cylindrica* and *Phragmites communis* together with *Aeluropus littoralis* had significantly different means. The means of most of the soil variables were different at a significance <0.00005. Though the field classification was based on plants, the soil properties showed greater difference between the vegetation categories than did the plant cover percents. This situation seemed to be promising for indicating particular ranges of soil variables by vegetation.

We wanted to obtain information on the precision with which the vegetation categories can predict the ranges of individual soil variables. The variables, whose range can be predicted precisely by vegetation categories, contribute much to the numerical separation of these categories. Discriminant analysis (DA) is the test that helps to find out how much the variables contribute to the separation. DA quantifies the degree of separability of different, established categories when the categorization is performed with a linear combination of numerical variables. It is analogous to a regression analysis (Davis 1986) in which the dependent variable is categorical. The algorithm of DA creates a series of discriminant functions in such a manner that it maximizes the ratio of the between-groups sum of squares to the within groups sum of squares of the discriminant scores. In this study, because there were two classes, the blocks were separated by one discriminant function. The correlation between the original variables and the discriminant function shows the degree of separability of the discriminating variables. If the correlation is high, the use of the categories for predicting different ranges of variables is promising. For instance, the large correlation between the discriminant function and the sulfate concentration of the 0 to 5-cm layer (0.76) indicated that the ranges of this variable in the two vegetation categories do not overlap much. On the other hand, the calcium concentration of the 20 to 25-cm layer showed the smallest correlation (0.24) with the discriminant function, and this suggests that the ranges of this variable in the two vegetation categories overlap more than, for example, the sulfate concentration in the 0 to 5-cm layer. The sign of the correlation coefficients was the same as in Table 1, and the order of magnitude was also similar.

The discriminant score determines the probably membership of the individual cases. When there are two classes, the cutting point between the two classes is calculated simply as the middle point between the means of the discriminant scores for the two classes. The cases having a discriminant score higher than this value belong to one class, and those having a lower score belong to the other. Then the precision of the classification is expressed by the classification matrix of DA. For our case it was created by comparing the original, field categorization of observations to the classification provided by the discriminant scores (Table 3). The classification matrix shows the number of properly classified, and misclassified cases for both vegetation categories. Of 54 total PHRAGMITES cases, 47 were classified correctly using the discriminant function; of the 47 IMPERATA cases, 43 were classified as belonging to that category by DA. Properly classified cases are placed on the diagonal; consequently, the precision of the classification was (47+ 43)/(54 + 47), i.e., 89%.

The two vegetation categories showed similar good predictability. However, the precision of predicting the IMPERATA category was greater (43/47) and was associated with higher correlations between the soil variables (Fig. 3) in this category than in the PHRAGMITES category (47/54).

One-dimensional thresholding

Six variables were selected from Table 2 for studying the precision of predicting their ranges separately, based on vegetation categories. Surface penetration and elevation showed differing strengths of correlation with the discriminant function. The salt concentration and the pH of the 0 to 5-cm layer were selected because of their traditional role in judging the quality of saline lands, and the chloride and sodium concentration of the 40 to 45-cm layer were selected as examples of

TABLE 3

Classification matrix for the separation of vegetation categories by soil variables with discriminant analysis (DA)

DA classification	Category distinguished in the field	
	PHRAGMITES	IMPERATA
PHRAGMITES	47	4
IMPERATA	7	43
Cases	54	47

Precision of classification: 89%

variables showing a relatively strong correlation with the discriminant function.

For each of the soil variables mentioned, the full range of the values of soil variables were divided, with a cutting point, into two parts, each resulting range corresponding to one vegetation category and named accordingly. The cutting point was initially positioned at the mid point between the means of the variables in the separate vegetation categories and then shifted until the ranges had the closest match, which is shown in Fig. 5. This simple method is applicable to every distribution; it is called one-dimensional thresholding and can be considered a primitive univariate version of discriminant analysis. It shows the maximal precision percent it is possible to attain in predicting ranges of a property.

Elevation was a variable for which the full range of variables was covered in both categories, except for the highest two points. In the case of penetration (Fig. 5a) and surface salt concentration (Fig. 5b), one range was combined with the other. The penetration range of the PHRAGMITES category was fully inside the range of the IMPERATA category. The surface salt concentration range of the IMPERATA category was a narrow range of the full PHRAGMITES range. Chloride and sodium concentrations of the deepest layer had bimodal distributions (Fig. 5d), and they showed the most precise match among the vegetation categories and assigned ranges.

Table 4 shows the match between the vegetation categories and the ranges of soil variables. If the value of the given variable in a block falls into the assigned range, this block is added to the cases with a good match. Table 4 shows that among the blocks classified as PHRAGMITES in the field, 55 had elevation values higher than 260 mm, thereby matching the range assigned to this vegetation category (PHRAGMITES range). Twenty blocks had lower elevation values, i.e., these were misclassified by this cutting point, and were placed into the IMPERATA range. Among the IMPERATA blocks, there were 44 blocks with an elevation lower than 260 mm, meeting the requirement of the IMPERATA range, and 10 blocks with a higher elevation than the cutting point; these were misclassified and placed in the PHRAGMITES range. Consequently the precision of match was (55 + 44)/129, i.e., 77%.

The variables that showed a stronger significance in separating the two vegetation categories by the Mann-Whitney test and that had a stronger correlation with the discriminant function (salt

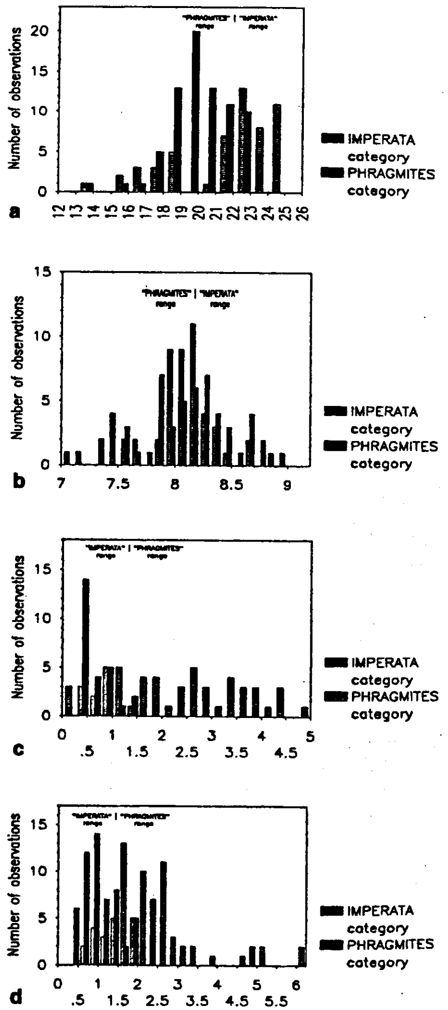

FIG. 5. Histograms of some soil variables with the indication of the cutting points between the separate ranges of the values of soil variables for the two vegetation categories

a. Surface soil penetration (mm), cutting point is 22 mm

b. pH in 1:5 water extract of the 0 to 5-cm layer, cutting point is pH 8.2

c. Salt concentration in the 0 to 5-cm layer, cutting point is 1.3%

d. Sodium ion concentration in 1:5 water extract of the 40 to 45 cm layer, cutting point is 1.5 $cmol_c\ kg^{-1}$

concentration of the 0–5-cm layer; chloride and sodium concentration of the 40–45-cm layer) had matches greater than 80%, and those with a weaker significance (in the Mann-Whitney test) and weaker correlations with the discriminant function (elevation, penetration resistance and pH of the 0–5-cm layer) had matches of about 70 to 75%. The differences in the precision for elevation (77%), penetration resistance (75%), and pH (67%) can be deduced from Table 1, where the correlation coefficients of these variables with the first canonical variate of soil variables were 0.73, −0.46, −0.37. The differences between the precision of matches for the separate vegetation categories accorded with the correlations of the soil variables with the first canonical variate of soil variables (Table 1). The precision of match was greater in the IMPERATA category than in the PHRAGMITES category when the predicted variable showed a positive correlation in Table 1. In Table 4, the precision of match for elevation in the PHRAGMITES category was 55/(20 + 55) = 73%; it was 44/(44 + 10) = 81% in the IMPERATA category. These precisions were, respectively, 63% and 98% for surface salt concentration, 80% and 91% for chloride concentration of the deepest layer, and 81% and 87% for sodium concentration of the deepest layer. These variables showed a positive correlation with the canonical variate of soil variables as well as with the discriminant function, and their large value is characteristic for the PHRAGMITES category. On the other hand, the range of values of these variables can be predicted more precisely for the IMPERATA category. It suggests that for the plants of the IMPERATA category, the ecological conditions described by these variables were manifested as a threshold, whereas the plants of PHRAGMITES category have a wider tolerance for these variables. Penetration and surface pH had a negative correlation with the first canonical variate of soil variables (Table 1), and these variables had larger value in the IMPERATA category than in the PHRAGMITES category. The precision of match for penetration was 87% in the PHRAGMITES and 59% in the IMPERATA category (Table 4), and for surface pH 78% in the PHRAGMITES and 55% in the IMPERATA category (Table 4). These latter variables were separated more precisely for the category where their values were usually smaller. It suggests that for the characteristic plants of the PHRAGMITES category (Table 2), the pH and penetration values exerted a limiting condition,

TABLE 4

Coincidence between the classification, based on vegetation categories, and discrete ranges of soil variables

Assigned range	Plant category: PHRAGMITES	Plant category: IMPERATA
Elevation, cutting point = 260 mm		
PHRAGMITES range	55	10
IMPERATA range	20	44
Precision of match: 77%		
Surface soil penetration, cutting point = 22 mm		
PHRAGMITES range	65	22
IMPERATA range	10	32
Precision of match: 75%		
pH in 1:5 water extract of a 0 to 5-cm layer, cutting point = pH 8.2		
PHRAGMITES range	42	21
IMPERATA range	12	26
Precision of match: 67%		
Salt concentration in a 0 to 5-cm layer, cutting point = 1.3%		
PHRAGMITES range	36	1
IMPERATA range	18	46
Precision of match: 81%		
Chloride ion concentration in 1:5 water extract of a 40 to 45-cm layer, cutting point = 1.4 $cmol_c\ kg^{-1}$		
PHRAGMITES range	60	5
IMPERATA range	15	49
Precision of match: 84%		
Sodium ion concentration in1:5 water extract of a 40 to 45-cm layer, cutting point = 1.5 $cmol_c\ kg^{-1}$		
PHRAGMITES range	61	7
IMPERATA range	14	47
Precision of match: 84%		

As a result of rainfall during sampling, the pH and salt concentration of the 0 to 5-cm layer were evaluated on a reduced data set.

but the plants of IMPERATA category had wider tolerance for these conditions.

Discrimination between the vegetation types can be used to select potential areas of land for cropping inasmuch as only the less saline sites, i.e., those associated with IMPERATA category, are likely to be suitable. The error of estimating in which of the two ranges of surface salt concentration (either less than 1.3% or more) the observed blocks fall, using the two vegetation categories, is 19%. The chance of encountering a

TABLE 5

Correlation matrix of the occurrence of misclassifiction by the variables (n=128)

Variables	Penetration	pH 0–5	Sum 0–5	Cl 40–45	Na 40–45
Elevation	0.24*	0.13	0.45**	0.17	0.15
	*(−0.36**)*	*(−0.33**)*	*(0.78**)*	*(0.59**)*	*(0.58**)*
Penetration		0.21	0.04	0.20	0.09
		*(0.34**)*	*(−0.36**)*	*(−0.42**)*	*(−0.34**)*
pH 0–5			0.01	0.14	0.09
			*(−0.42**)*	*(−0.46**)*	*(−0.47**)*
Sum 0–5				0.51**	0.44**
				*(0.85**)*	*(0.82**)*
Cl 40–45					0.62**
					*(0.92**)*

Correlation coefficients between the values of the variables are shown in italics in brackets.

Correlations with the values of pH and salt concentration in the 0 to 5-cm layer were calculated with a reduced data set.

*Shows significance <.01, and ** shows significance <.001.

block with a salt concentration greater than the limit using the IMPERATA category is very small, less than 1% (Table 4), but the error of underestimating the salt concentration from PHRAGMITES category is considerable (27%). The prediction is biased toward encountering a lower than expected salt concentration. This could be favorable for the selection of possible pieces of cropland if the cutting point coincided with the critical limits of salt concentration for agricultural crops. The cutting points do not match those limits because the prediction of salt concentration was evaluated in the 0 to 5-cm soil layer, not in the plow layer (0–20 cm), as in mapping soil salt concentration here.

The study of the spatial distribution of misclassified blocks can inform whether misclassification can be linked to specific spots, to the borders between categories, or is scattered haphazardly. The location of the blocks and the success or failure of classification is shown in Fig. 6. A large, misclassified PHRAGMITES patch was evident in the pH and salt concentration of the 0 to 5-cm layer at around East 30, North 190. Most misclassified patches were located close to the borders between categories (Fig. 6), which suggests that the manifestation of the change in soil properties and vegetation is different.

To test the joint or independent (occurrence of) misclassification for the values of soil variables, a correlation matrix of misclassification was calculated (Table 5) using a misclassification score. The latter was created by assigning 1 to misclassified blocks and 0 to properly classified blocks. Consequently, the correlation coefficient of misclassification indicates the relative frequency of the coincidence of misclassification of two variables in the same block. A comparison of the correlation coefficients of the values of the variables is given in the same table because the difference between the two correlation coefficients suggests whether it is coincidental that both are correlated and both have misclassified blocks in the same locations. When both correlation coefficients of one variable (calculated from the values and also from the misclassification scores) are large, the misclassification can be considered to be regular, as in the case of surface

TABLE 6

Misclassified and precisely classified blocks through the use of the vegetation categories

Variables	At the borderline				Inside the category			
	P*	P	I*	I	P*	P	I*	I
Penetration	5	23	15	2	5	42	7	30
Na 40–45	6	14	4	9	8	47	3	38

P* P, I*, and I are blocks falling into the PHRAGMITES (P) or IMPERATA (I) category with values falling (no asterisk) or not falling (*) into the assigned range.

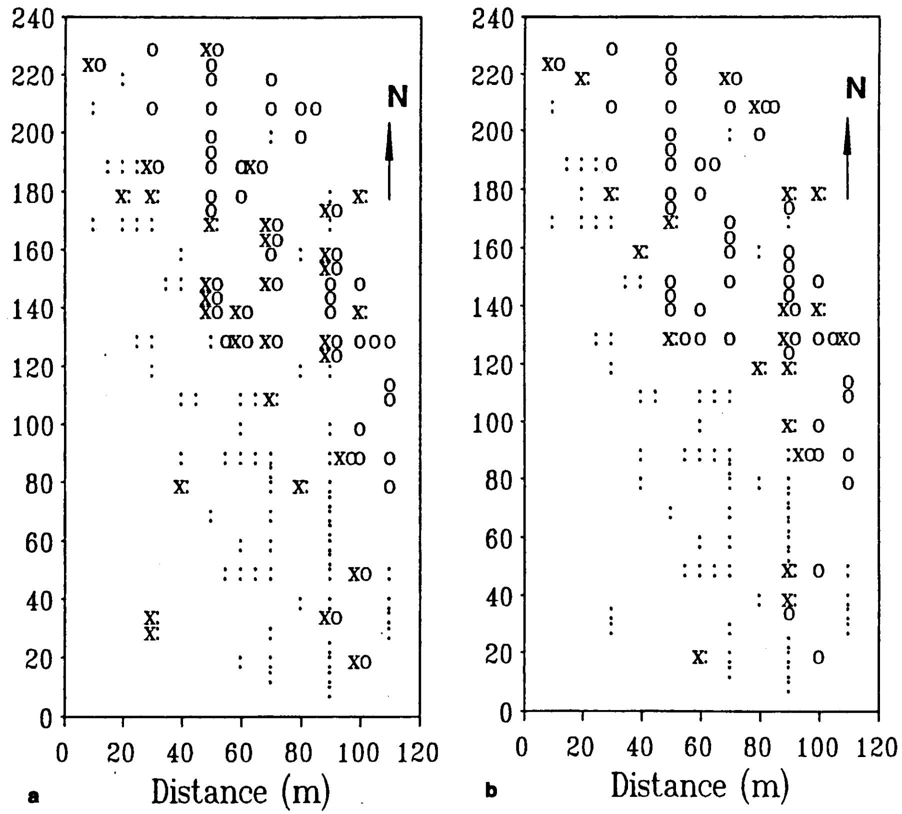

FIG. 6. Location of properly classified (":" is PHRAGMITES and "o" is IMPERATA category) and misclassified (indicated by preceding "x") blocks using the classification of Fig. 5 and Table 4.
a. Surface soil penetration (mm)
b. Sodium ion concentration in 1:5 water extract of the 40–45 cm layer

salt concentration, chloride, and sodium concentration of the deepest layer, and it indicates that there is a spatial correlation of misclassification between the pairs of variables. When the variable shows a large correlation with the values of other variables, but not with the misclassification scores, its misclassification can be considered irregular, such as was shown by surface pH. This variable was predicted least precisely, and its pattern of misclassification did not show strong correlation with other variables.

For the study of the spatial location of misclassified blocks, two variables were selected whose values were predicted with differing precision by the vegetation. Penetration resistance represented the second lowest and sodium concentration of the 40 to 45-cm layer the largest precision of prediction. The error of misclassifying the blocks by plant categories was greater at the borderline to the two vegetation categories, as was shown by Table 6. This error was expressed by calculating the ratio of the number of misclassified blocks to the number of precisely classified ones at two zones: at the borderline around the vegetation categories and within the categories. The ratio of misclassified blocks to precisely classified ones is much higher at the borders than within the categories. For penetration resistance, the ratio of misclassified blocks at the borderline of the IMPERATA category is extremely high and suggests that there is a transitional zone for this variable between the two vegetation categories. In this zone the vegetation did not reflect the changes in penetration resistance. No such transitional zone was observed in the case of sodium concentration of 40 to 45-cm layer.

CONCLUSION

In the abandoned land of the Huang-Huai-Hai Plain, there is a close correlation between the composition and extent of semi-natural vegetation cover, salt concentration, pH, penetration resistance, and elevation of the sites. This vegetation of the abandoned saline site was a good indicator of differences in the values of soil variables. The sites in the PHRAGMITES vegetation category are at higher elevation, the soil has a larger salt concentration, the pH is closer to neutral, and the penetration resistance is smaller than those sites in the IMPERATA category of vegetation. The match between vegetation categories and nonoverlapping classes of soil property values had a precision of 67 to 84%, depending on the soil variable selected.

The precision of predicting soil salt concentration by vegetation is good, and it is straightforward and economical in the area studied inasmuch as the two vegetation categories are easy to distinguish. The variables studied are thought to be some of the most important in determining the suitability of the abandoned saline lands for cropping. To extend the prediction of the soil variables for other vegetation categories and areas based on existing correlations established between soil and vegetation will require further study.

ACKNOWLEDGMENTS

We are much indebted to Dr. M. Kertész, Dr. L. Pásztor, and Prof. H. Chen for continuous discussion on the paper; to Dr. S. Kabos for comments on statistics; to Dr. D. Nielsen for criticism; and to two anonymous reviewers for helpful suggestions. Thanks are due to the staff of the Nanpi and Wangsu experimental stations of the Shijiazhuang Institute of Agricultural Modernization, who, under the leadership of Prof. K. X. Tian, kindly devoted their time to the laboratory analyses. Prof. J. M. Tian has contributed much to the profile description and field work, and the work would have been impossible without the information from Prof. H. Zhu on the local vegetation.

REFERENCES

Ballenegger, R. 1929. Les méthodes de la cartographie des sols alcalins (salins). *In:* Transactions of the Alkali-Subcommission of the International Society of Soil Science, Vol. A. Budapest, Hungary, pp. 9–12.
Bodrogközy, Gy. 1965. Ecology of the halophilic vegetation of the Pannonicum. II. Correlation between alkali ("szik") plant communities and genetic soil classification in the Northern Hortobágy. Acta. Bot. Hung. 11:1–51.
Chapman, V. J. 1960. Salt marshes and salt deserts of the world. Leonard Hill Ltd., London.
Davis, J. C. 1986. Statistics and data analysis in geology. Wiley, New York.
Digby, P. G. N., and R. A. Kempton. 1987. Multivariate analysis of ecological communities. Chapman and Hall, London.
Girard, M., and L. A. Isarwa. 1990. Remote sensing of arid and semiarid regions. Desertification Control Bull. 18:13–18.
Inanaga, S. 1991. Improvement of sloping beds of row for crop establishment in saline low land in Cangzhou Prefecture, China. Proceedings of International Symposium Impacts of Salinization and Acidification on Terrestrial Ecosystem and its Rehabilitation. September 26–28. Fuchu, Tokyo, pp. 143–148.
Kearney, T. H., L. J. Briggs, H. L. Shantz, J. W. McLane, and R. L. Piemeisel. 1914. Indicator significance of vegetation in Tooele Valley, Utah. J. Agric. Res. 1:365–417.
Lanyon, L. E. and W. R. Heald. 1982. Magnesium, calcium, strontium and barium. *In* Methods of soil analysis, Part 2, Agronomy 9. A. L. Page (ed.). ASA-SSSA, Madison, WI.
Lukács, A., and M. Rédly. 1988. Quantitative determination of sulfate ions in aqueous soil extracts by iodometry. *In* Buzás, I. (ed). Methods of soil analysis 2. Physico-chemical and chemical methods of soil analysis. I. Buzás (ed.). Mezögazdasági Kiadó, Budapest pp. 185–186 (in Hungarian).
Magyar, P. 1928. Beitroäge zu den Pflanzensoziologischen und geobotanischen Verhältnissen der Hortobágy-Steppe. Erdészeti Kísérletek. 30:210–226.
Marsal, D. 1987. Statistics for geoscientists. Pergamon Press, Oxford, U.K.
Rengasamy, P., and K. A. Olsson. 1991. Sodicity and soil structure. Aust. J. Soil Res. 29:935–952.
Shantz, H. L. 1911. Natural vegetation as an indicator of the capabilities of land for crop production in the Great Plains area. U.S. Dept. Agric., Bureau of Plant Industry, Bull. No. 201, Washington, DC.
Shantz, H. L., and R. L. Piemeisel. 1924. Indicator significance of the natural vegetation of the southwestern desert region. J. Agric. Res. 28:721–802.
'Sigmond E. 1927. Hungarian alkali soils and methods of their reclamation. University of California Press, Berkeley, CA.
Tóth, T., and K. Rajkai. 1994. Correlations between soil and vegetation in a solonetzic grassland. Soil Sci. 157:253–262.
Tóth, T., S. Matsumoto, R. Mao, and Y. Yin. 1994. Plant cover as predictor variable of salinity and alkalinity in abandoned saline soils of the Huang-Huai-Hai Plain, China. Agrokémia és Talajtan 43:175–195.
Zonneveld, I. S. 1976. Survey methods for performance, monitoring and prognosis for natural vegetation and economic crops with special reference to salt-affected soils. *In* Prognosis of salinity and alkalinity. FAO Soils Bull. No. 31, pp. 195–211.