Arid Land Research and Management 16:55–68, 2002

1532-4982.02 $12.00+.00

Statistical Prediction of the Presence of Salt-Affected Soils by Using Digitalized Hydrogeological Maps

T. TÓTH
L. PÁSZTOR

Research Institute for Soil Science and Agricultural Chemistry
Hungarian Academy of Sciences
Budapest, Hungary

S. KABOS

L. Eötvös University
Budapest, Hungary

L. KUTI

Geological Institute of Hungary
Budapest, Hungary

Since spatial databases created from maps in Geographical Information Systems (GIS) do not typically meet the requirements of multivariate tests such as multiple linear regression analyses, we used novel, more robust statistical techniques for predicting the occurrence of certain categories of a soil map. Based on published maps of salt-affected soils (SASs), (hydro)geological conditions (all at the scale of 1:500,000) and regionalization, we showed that the spatial occurrence of SASs in the Great Hungarian Plain can be predicted with an overall precision of 91; 96% for the SASs and 99% for the non-SASs. When we separated the non-SASs into potential SAS and non-SAS categories, the overall precision was 91%, and 87% for the SASs, 92% for the potential SASs and 94% for the non-SASs. The technique of classification trees proved to be better than the classical Multiple Linear Regression, since it does not have limitations regarding the distribution of the data set created from the maps and because it flexibly provided a possibility for incorporating nominal variables through the previous use of homogeneity analysis by means of alternating least squares (HOMALS). Inside the multidimensional space, the first and most important (regarding spatial extent) splits for the separation of non-salt-affected cases from salt-affected ones were realized in the plane of two variables derived with HOMALS from the map of "Textural classes of near-surface geological formations" and the map of "Dominant ion type of groundwater" combined with the map of "Taxonomical division of the regions of

Received 24 October 2000; accepted 27 March 2001.
This research and publication was financed by the Hungarian National Science Foundation under grants nos. T030738 and T023271. L. Pásztor acknowledges the help of Bólyai Grant. We thank Miss J. L. Wardell for polishing the style of the paper.
Address correspondence to Dr. T. Tóth, MTA TAKI, 1022 Budapest, Herman Ottó út 15, Hungary. E-mail: tibor@rissac.hu

Hungary." The plane of these variables was split into five main bands. The clay band has 70% of SASs and the mixed silty layer was divided to a bicarbonatic one with 12% SAS, and a sulfatic one with 65% SAS cover. The remaining sandy band has few SAS, but the dominance of sulfate in the groundwater indicates a higher cover of SASs and these soils constituted the two last bands. Other variables entering the classification were the groundwater level above sea level and the depth to groundwater.

Keywords classification tree, cross-tabulation, HOMAL, Hungary, saline regions, salinity risk

Salt accumulation in otherwise potential agricultural lands is a worldwide problem, affecting 340 million hectares globally. Although most research focuses on man-made salinization of irrigated lands, the areas affected by natural salinization are larger than the land affected by salinization resulting via irrigation. In several countries, these naturally salt-affected areas are important as prospective croplands or as wildlife reserves.

For the rational management of salt-affected lands, the use of maps on salt accumulation are indispensable. Due to the recent development of techniques and concepts, the mapping of SASs is facilitated at every scale of managemental decisions (Tóth et al., 1998). The use of existing maps as predictors of salt accumulation has great importance where (a) there is no detailed information on spatial distribution of salinity and (b) the main factors and mechanisms of salt accumulation have not been clarified.

For predicting "salinity risk" Tickell (1997) rated categories of groundwater salinity, vegetation, rainfall, aquifer yield, and presence of laterite from 1 to 10 regarding the salinity hazard. These maps were then intersected, the hazard ratings of the resultant delineations summed, and the hazard map was compiled, using four categories ranging from very low to high hazard of salinity. Wei and Liu (1988) and Eilers and others (1997) used similar concepts. Bradd and others (1997) used the method of "weights-of-evidence" based on the cross-table of the map of salinity versus maps of geology, soils, land use, rainfall, landforms, groundwater characteristics, and vegetation similarly as done by the previous article of Tóth and others (2001).

This article shows a comprehensive, simultaneous treatment of several predictor maps for the prediction of salt accumulation and the analyses of the importance of hydrogeological and geomorphological factors. Our objective was to test the recently developed spatial data management and statistical techniques for the prediction of different levels of salt accumulation based on previously prepared maps. Since the statistical prediction methods like the classical ANOVA and linear regression are based on the assumption that the sample elements are stochastically independent (i.e., the error term is independent from observation to observation), which is not typical in the case of soil properties using GIS techniques, we propose here an alternative statistical inference taking into account the spatial regionalization of sites.

Materials and Methods

In the Great Hungarian Plain, the elevation of typical salt-affected areas above sea level is 80–90 m, climate is temperate continental, with mean annual temperature of 10°C (– 2°C in January, 21°C in July), average annual precipitation is 527 mm (the most rainy month is June with 71 mm, the least rainy is January with 30 mm), mean annual pan evaporation is 900 mm.

The base maps and the development of the digitalized database is described in the accompanying article of Tóth and others (2001). The variables derived from the base maps are referenced as "TFM—gw asl" derived from the "Map of groundwater level above sea level," "TVM—gw depth" derived from the "Map of depth to

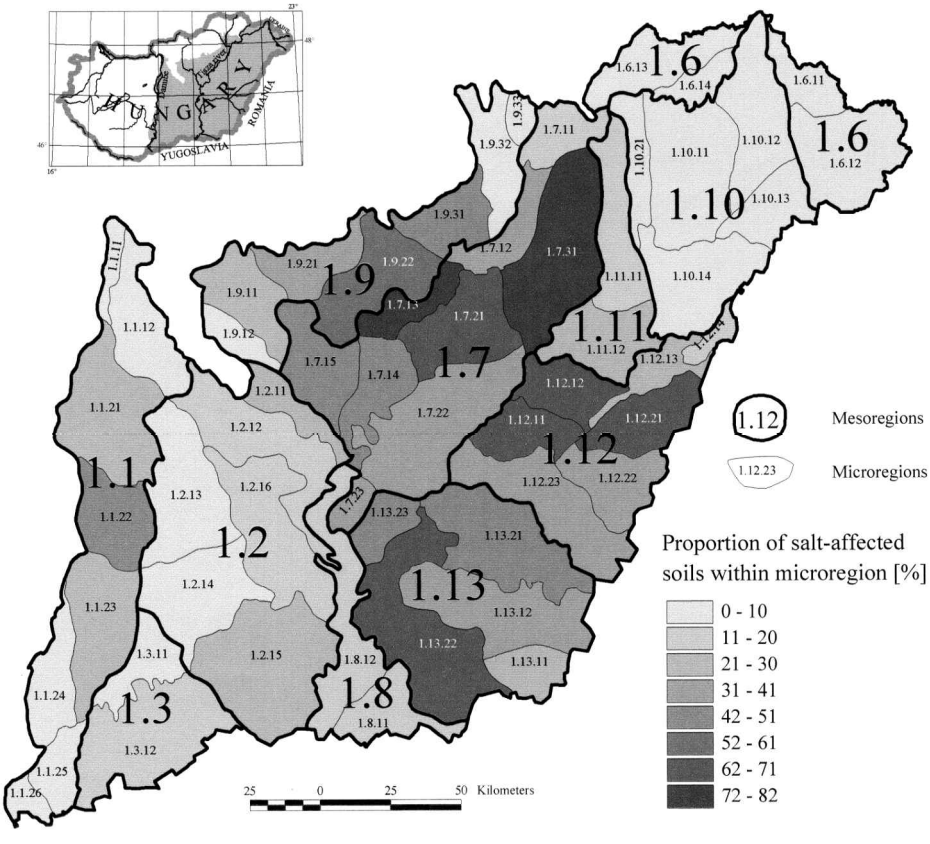

FIGURE 1 Geographical meso- and microregions in the Great Hungarian Plain.

groundwater,'' and ''OAT—gw cc'' derived from the ''Map of total soluble salt concentration of the groundwater.''

The map ''Taxonomical division of the regions of Hungary'' (Somogyi, 1991) (Figure 1) was based on recent geomorphological and regional research. It presents a hierarchical taxonomy of the regions. Inside the Great Hungarian Plain, there is one macroregion with 13 mesoregions which are divided into 70 microregions. In our analysis, the mesoregions were used in the quantification of the categories. The original database of Tóth and others (2001) was supplemented by the addition of this map. Based on the map of salt-affected soils by Szabolcs (1974), a soil was considered to be salt-affected (SAS) if it belonged to the first eight salt-affected soil categories of the original map, and it was considered non-salt-affected (non-SAS) if it belonged to either the ''Potential salt-affected soil'' or ''Non-salt-affected soil'' categories.

Statistical Techniques

Category quantification HOMALS (Homogeneity analysis by means of Alternating Least Squares) from package SPSS Categories (SPSS, 1996) converts the nominal variables into numerical ones by giving quantification scores to the categories. The mathematical background is described by Gifi (1990). Contingency tables that were analyzed (Tables 1 and 2) consist of the area (km^2) covered by the delineations by FAK (ion types) and TKT (texture) as a row variable and the mesoregion as column variable. The variables calculated by SPSS HOMALS were named TTKT1, TTKT2,

TABLE 1 Contingency Table of the Maps, "Taxonomical Division of the Regions of Hungary" and "Textural Classes of Near-Surface Geological Formations in the Great Hungarian Plain," as Expressed by the Territory of the Delineations

Textural sequences	Mesoregions												
	Code	1.1	1.2	1.3	1.6	1.7	1.8	1.9	1.10	1.11	1.12	1.13	Total
SSSSS	21	435	2,400	414	264	334	66	946	3,766	30	87	168	8,911
SSCCC	22	75	1,528	288	32	87	141	99	292	15	10		2,568
SCCCC	23	13	117	83	11	105	21		12			8	370
SISSI	24	153	1,610	317			124		140	14			2,358
SSGGG	25	342	15					235					592
SCSCC	26				12	62		8	2		47		131
SCGGG	27	154											154
IIIII	31	2	150	222		84	24		0	176			658
IISSS	32	4	128	77		564	150		1	134	80	493	1,631
ISSSS	33	123	17	104	42	146	323		42	42	46	5	890
ISIIS	34	16	417	479	1	4	45		79	53			1,095
ICCCI	36						2			8	25	44	79
CCCCC	41	60	207		1,005	4,552	515	1,759	74	831	2,379	3,184	14,565
CCSSS	42	398	40	4	910	426	99	194	84	251	536	319	3,261
CSSSS	43	1,073	122	7	227	533	57	176	25		656	672	3,549
CICII	44	73	141			88	29			56	11	37	434
CCGGG	45	870	24					79					972
CSCSS	46		40		107	128	12	294	33		358	208	1,180
Total km^2		3,789	6,958	1,994	2,610	7,114	1,608	3,790	4,551	1,611	4,235	5,138	43,398

Notes: In the sequence of 2 m thick layers from 0 to 10 m depth, G is gravel, S is sand, I is silt and C is clay. Codes are as displayed by Figure 2 of Tóth et al. (2001). Mesoregions are shown in Figure 1.

quantificated variables of dominant ion types (originally TKT); and TFAK1, TFAK2, quantificated variables of textural classes of near-surface formation (originally FAK). In terminology of the data analysis, our aim was to predict the response variable (SAS) from the explanatory variables (TFAK1—quantificated texture, TTKT1—quantificated ion type, TVM—gw depth, TFM—gw asl and OAT—gw cc). Statistical analyses were performed both excluding and including quantificated variables. The case of a two-level response (SAS/non-SAS) was treated in detail, but results for a three-level response (SAS/potentially SAS/non-SAS) were also reviewed. We aggregated our data sets to 508 parts, which are homogeneous with respect to SAS, TVM, TFM and OAT. The variables TTKT1-2 and TFAK1-2 were calculated as weighted averages of the corresponding quantificated variables generated by HOMALS. The weighting was taken by area of the delineations.

Three methods of statistical inference were used: Multiple Linear Regression, Logistic Regression, and Classification Tree. The first two are standard tools in statistical analysis; the third seems to be neglected in applications for soil science, and will be treated here in detail.

A recursive partitioning of a data set according to selected threshold values of the explanatory variables, may be expressed in form of a Classification and Regression Tree (CART, Breimann et al., 1984; Breimann & Friedmann, 1985). The computer program "tree" that we have used is provided by the statistical package S-PLUS (1994). The algorithm starts with the whole set of objects and chooses explanatory (splitting) variables, optimizing the precision of the Classification Tree. The CART method accepts categorical variables as explanatory variables, but the number of levels is very limited (because an L-level categorical variable causes to form $2^{(L-1)} - 1$ subsets). The variables, "Dominant ion types of groundwater" (TKT, 29 levels) and "Textural classes of near-surface formations" (FAK, 18 levels), were beyond these limits. Due to these limitations, we required a quantification method to use TKT and FAK as explanatory variables.

Results and Discussion

The results of using multiple linear and logistic regression techniques for the prediction of salt-affected soils are shown in Table 3. In the case of using ordinal variables—TFM or "Groundwater level above sea level," TVM or "Depth to groundwater," OAT or "Total soluble salt concentration of the groundwater"—there was a weak correlation explaining only 15% of total variance. Logistic regression improved the correlation explaining 70% of the total variance, but the precision of predicting SASs was as low as 31%.

We had two more variables which could not be included in the analysis since they were nominal ("Dominant ion types of groundwater" and "Textural classes of near-surface formations"), and we used HOMALS for their quantification. The salt-affected areas of Hungary have been divided into specific regions as reported by Treitz (1924,1934), Sigmond (1927), Herke and others (1959), Arany (1956). The pattern of the regions of the Great Hungarian Plain as suggested by Somogyi (1991) showed a good agreement with these previously cited reports. We used these divisional units which express the spatial nature of salt accumulation. The map of the regions (Figure 1) shows that those microregions with high cover of SASs form an almost contiguous area (microregions 1.7.13, 1.7.21, 1.12.11, 1.12.12, 1.12.21), but are surrounded by areas that have decreasing cover of SASs with smooth transition. On the western edge of the Great Hungarian Plain by the Danube River there is a north-south oriented area (microregions 1.1.21, 1.1.22, 1.1.23) which includes most of the solonchak-solonetz areas.

The contingency tables of mesoregions (Tables 1 and 2) were used in the transformation of the above two categorical variables by the HOMALS procedure as

TABLE 2 Contingency Table of the Maps, "Taxonomical Division of the Regions of Hungary" and "Dominant Ions in the Groundwater in the Great Hungarian Plain," as Expressed by the Territory of the Delineations

Dominant			Mesoregions											
Cation	Anion	Code	1.1	1.2	1.3	1.6	1.7	1.8	1.9	1.10	1.11	1.12	1.13	Total
Na	biClsu	10					76	7			2		89	173
Na	Cl	11					18	28					12	58
Na	su	14		98			600	69	98		41	52	355	1,314
Na	subi	16											38	38
Na	bi	17	934	1,177	221	21	3,842	677	560	142	1,290	1,662	4,092	14,617
Na	bisu	19											26	26
NaMg	su	24		2			118		48					168
NaMg	bi	27					38		36	22	19	16		131
NaMg	bisu	29						23						23
NaCa	bi	37					13							13
NaCa	biCl	38										23		23
Mg	su	44	57											57
Mg	subi	46	19											19
Mg	bi	47	246	968	551		28	1	333	52		6		2,185

MgNa	su	54					3		46				49
MgCa	subi	66	27										27
MgCa	bi	67	598	95	171								864
Ca	Cl	71					11						11
Ca	Clsu	72				9							9
Ca	su	74		5		234	60	92	36	10		30	467
Ca	subi	76				13							13
Ca	bi	77	1,533	4,272	944	2,216	1,917	626	2,310	4,313	260	2,248	433	21,071
Ca	bisu	79		13									13
CaNa	biClsu	80					13	1					14
CaNa	bi	87	27	19	18	80	52						196
CaMg	bi	97	142	240	55		17	1	87			25	568
CaMg	Nasu	104				13				12		57	21	104
CaMg	Nabi	107	208	69	35	23	308	52	235		1	115	73	1,117
CaMg	NabiCl	108						31						31
Total km^2			3,789	6,958	1,994	2,610	7,114	1,608	3,790	4,551	1,611	4,235	5,138	43,398

Cations and chloride are abbreviated by their respective chemical symbols, sulfate = su, bicarbonate = bi. The order of the ions express their order of dominance in the groundwater. Codes are as displayed by Figure 3 of Tóth et al. (2001). Mesoregions are shown in Figure 1.

described in Materials and Methods. In the first row (original code on the map is 21) and column of Table 1, the number "435" shows the extent (km^2) of the area having texture SSSSS (sand for each 2 m depth-sequences down to 10 m depth) located inside the mesoregion No. 1.1. The number "31" in the last line of Table 2, for which the original code on the map is 108, shows the extent (km^2) of the area where the dominant ions are $Ca^{2+}/Mg^{2+}/Na^{+}$ and HCO_3^{-}/Cl^{-}. All these fall into the mesoregion No. 1.8.

Two new quantificated texture variables were received by HOMALS from the categories of the map: "Textural classes of near-surface formations." TFAK1 is the clay-silt-mixed gravel-sand variable, which can be interpreted as a series of formations of slow fluvial activity, fast fluvial activity, and eolian activity. TFAK2 expresses the contrast of a mixed gravelly sequence and a more homogeneous sequence series.

Similarly, two quantificated new ion type variables were received from the categories of the map: "Dominant ion types of groundwater." TTKT1 (quantificated ion type) can be considered as expressing decreasing solubility of dissolved salts from the salts of sodium through magnesium salts to calcium salts. TTKT2 can be considered as the sulfate-bicarbonate variable.

The first of these numerical variables, TTKT1—quantificated ion type and TFAK1—quantificated texture (Table 3), were included in the regression analysis.

TABLE 3 Regression Model Summary Statistics and Estimated Parameters

	Without nominal variables	With quantificated nominal variables
A. Multiple linear regression		
R^2	0.149	0.296
Residual MS	0.183	0.151
Model F test	253	363
Probability of F	0.0000	0.0000
	Parameter estimates	
Intercept	– 0.399 (0.042)	0.550 (0.040)
b1:TFM—gw asl	– 0.049 (0.002)	0.044 (0.004)
b2:TVM—gw depth	– 0.019 (0.008)	– 0.078 (0.007)
b3:OAT—gw cc	0.083 (0.112)	– 0.119 (0.012)
b4:TTKT1—ion types	Not included	– 0.317 (0.015)
b5:TFAK1—texture	Not included	– 0.187 (0.020)
B. Logistic regression		
Overall precision %	69	79
Precision for salt-affected soils %	31	63
Model Chi square	791	1,548
Significance of test	0.0000	0.0000
	Parameter estimates	
Intercept	0.334 (0.256)	1.830 (0.339)
b1:TFM—gw asl	– 0.347 (0.018)	0.333 (0.036)
b2:TVM—gw depth	– 0.239 (0.050)	– 0.964 (0.071)
b3:OAT—gw cc	0.409 (0.063)	– 1.007 (0.097)
b4:TTKT1—ion types	Not included	– 2.022 (0.127)
b5:TFAK1—texture	Not included	– 1.682 (0.167)

gw = groundwater; asl = above sea level; cc = concentration.

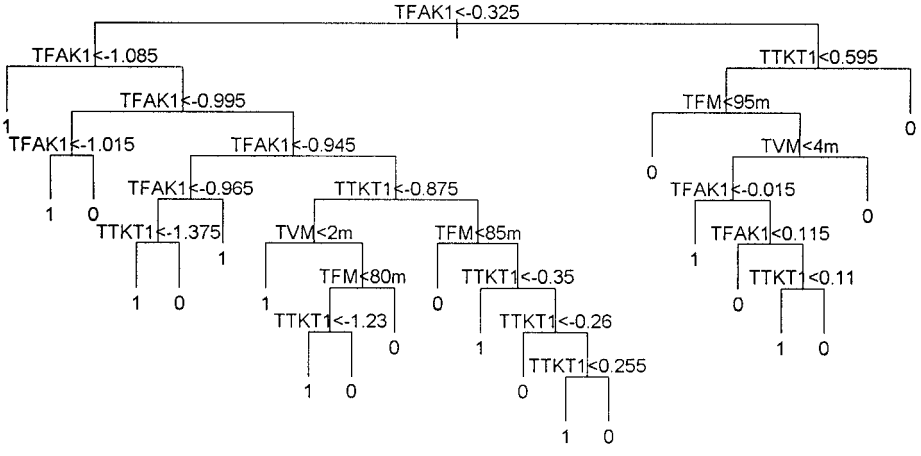

FIGURE 2 Classification tree for the separation of salt-affected soils (1) from non-salt-affected soils (0). Misclassification rate is 0.036.

In the case of multiple linear regression, there was an improvement in the coefficient of determination (R^2), from 15% to 30% explanation of total variance. With the inclusion of the texture and dominant ion variables in the logistic regression, 80% of the total variance could be explained. Additionally, for the SASs the precision was improved to 63%. Further inclusion of TKT2 and TFAK2 (results are not shown) did not improve on regression analyses.

The lack of a good correlation employing these estimation methods may be explained by the fact that the relationship between the explanatory variables and the salt accumulation process is neither linear nor additive. Therefore, we turned to the technique of the classification tree.

In Figure 2, we present our most important results determined by the CART methodology. In the subsequent splitting of the cases into increasingly homogeneous "leaves" of the "classification tree," all of the numerical variables discussed were entered, that is TTKT1, TFAK1, TFM, OAT and TVM. The major splits can be followed in the plane of TTKT1—quantificated ion type and TFAK1—quantificated texture on Figure 3. In this plane, conceptually, the probability of salt accumulation is in inverse relationship with the values of TFAK1 and TTKT1 as described by Tóth and others (2001). The dominance of fine particles (clay and silt) in soils and subsurface materials, restricts the movement of solutes; however, the thickness of the saline layers as affected by capillary rise is substantial. Consequently, salts may move up the soil profile from groundwater, and are difficult to remove by leaching. Sand and gravel do not have a degree of high capillary rise from groundwater and permit the rapid leaching of salts from the surface layers. When Na is the dominant cation in groundwater, it immediately affects the soil properties, especially when the soil is clayey due to the adsorption of Na^+ ions onto the exchange sites. Mg^{2+} behaves somewhat similar to Na^+, but Ca^{2+} typically does not cause salt accumulation as observed in the Great Hungarian Plain. The general differences in dominance of cations are modified by the dominance of anions. The presence of SO_4^{2-} is typically more closely related to salt accumulation, rather than the presence of HCO_3^-. From the surrounding mountains groundwater flows towards the deepest areas of the Carpathian Basin where it gets shallow, some of it evaporates, and its salt concentration increases. The differences in the solubility of the dissolved salts are paralleled with a change in composition. At elevated groundwater salt concentrations, SO_4^{2-} becomes the more dominant anion.

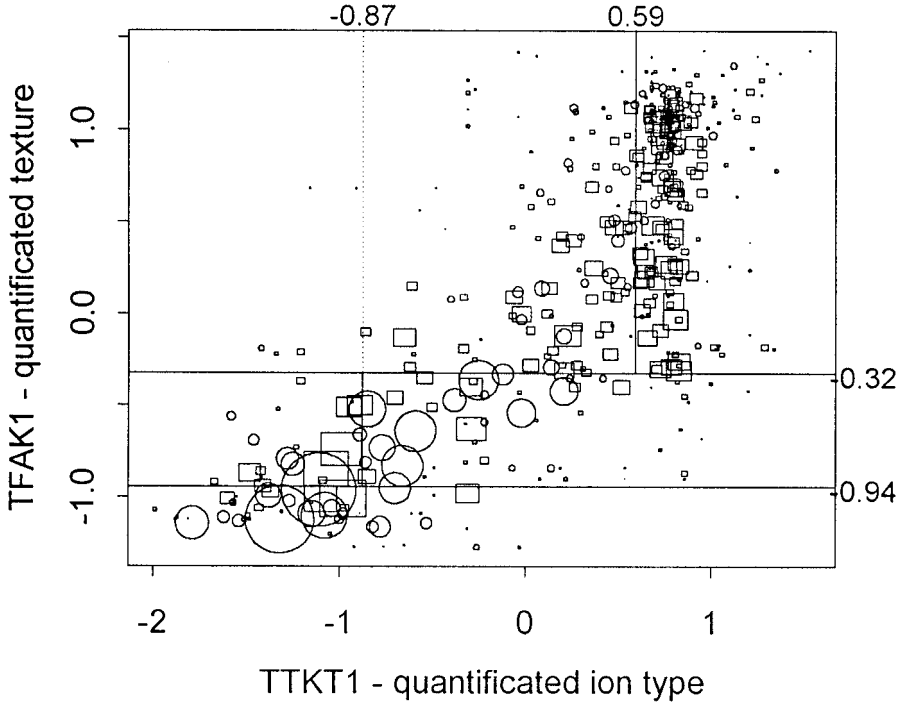

FIGURE 3 Major splits made by the classification tree in the plane TTKT1 (quantificated ion type)–TFAK1 (quantificated texture) with the indication of salt-affected (O) and non-salt-affected soils (□). The size of symbols indicates the relative coverage of the soils.

Splits are discussed as shown in Figure 3, which also considers the extent of separated leaves (the final cases resulted at the end of the branches). The first split was made at the value – 0.325 of TFAK1 (quantificated texture), which separated the geological formations dominated by clayey and silty surface layers, where the occurrence of salt accumulation is much greater than in the remaining cases. On the left branch of the tree, the split at TFAK1 – 1.085 and mostly at TFAK1 – 0.945 separated the most extreme clayey soils associated with salt accumulation. The groundwater under these soils is dominated by Na^+; the resulting band is identified as "Na-clay band." Following minor (regarding spatial extent) splits in the silty band, one important split occurs at TTKT1 (– 0.875), which separates the groundwater types dominated by Na^+ and HCO_3^- from the groundwater dominated by $Na^+/Mg^{2+}/Ca^{2+}$ and SO_4^{2-}. Sulfate is linked to a larger probability of salt accumulation ("$Na^+/Mg^{2+}/Ca^{2+}$-SO_4^{2-}-silt band") and the part of the band < -0.875 accounts for little of the salt-affected soils. In this "Na^+-HCO_3^--silt band" the dominance of silt corresponds to the loessic sediments which contain 20–25% $CaCO_3$ and upon weathering, easily produce Na^+. In comparison, the next split on the left shows that the salt-affected areas have shallower groundwater level (TVM—gw depth < 2 m) and usually lie lower (TFM—gw elevation asl < 80 m) as shown by Table 2 of Tóth and others (2001) than the non-SAS. On the right branch of the tree the split at TTKT1 0.595 has separated two areas with few SASs. However, there is a "Ca^{2+}/Mg^{2+}-SO_4^{2-}-sand band," where the dominance of SO_4^{2-} indicates a somewhat higher probability of salt accumulation, whereas the "Ca^{2+}/Mg^{2+}-HCO_3^--sand band" is not salt-affected. The overall precision of the classification tree is acceptable: 96%, and 91% for SAS, and 99% for non-SAS. The splits were similarly

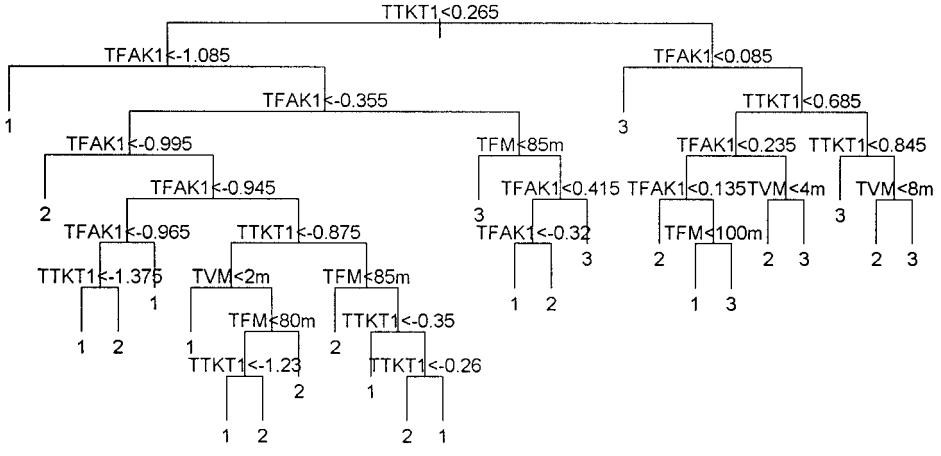

FIGURE 4 Classification tree for the separation of salt-affected soils (1) from potentially salt-affected soils (2) and non-salt-affected soils (3). Misclassification rate is 0.088.

evaluated on the basis of the microregions and neighboring salt-affected regions occupied neighboring position in the plane of TTKT1—quantificated ion type and TFAK1—quantitated texture (figure not shown).

Figure 4 shows the classification tree of three categories, distinguishing the potential salt-affected soils. The same variables were included as in Figure 2. There was an overall precision of 91%, 87% for SAS, 92% for the potential SAS, and 94% for non-SAS. The most important splits occurred in the plane of TTKT1 (quantitated ion type)–TFAK1 (quantitated texture). Most splits parallel those of Figure 2, with one exception, the first split at TTKT1 (quantitated ion type) = 0.265. On the left of this split, there are the areas where Na^+ is the dominant cation in the groundwater with a greater probability of salt accumulation, and on the right, the Ca^{2+}/Mg^{2+}-dominated groundwaters with sandy sequences. On the left side of the tree, the "Na^+-clay band" is separated, then the next split separates the upper left corner of the TTKT1–TFAK1 plane identified as potentially salt-affected soils. Continuing on the left, we arrive to the same branches that we saw on the two-category classification tree of Figure 2. On the right side of the tree, first the lower left corner of the TTKT1–TFAK1 plane is separated at TFAK1 (quantitated texture) = 0, followed with a split at TTKT1 = 0.685 separating non-SAS on the right side which is due to the dominance of Ca^{2+} and Mg^{2+} in the groundwater. Delineations containing more clay (TFAK1 < 0.135) are potentially salt-affected soils. The remaining branches, including the deeper groundwater level under the soil surface or the higher groundwater level above sea level, indicate a smaller probability of salt accumulation.

Conclusion

Each of the agrogeological variables studied was earlier reported to be influential on the occurrence of SASs in the Great Hungarian Plain. The importance of the depth to groundwater (Mados, 1943; Darab, 1967) and the salt concentration of groundwater ('Sigmond, 1927; Scherf, 1935; Endrédy, 1941; Erdélyi, 1979) were the factors most discussed, and the effect of the groundwater level above sea level (Kreybig & Endrédy, 1935; Endrédy, 1941) was also addressed by a few researchers. The importance of the distinctive particle fractions and stratification was emphasized by

'Sigmond (1927) and recently by Eilers et al. (1997) and Bradd et al. (1997). In Hungary, the relative importance of sodium in the formation of SASs was emphasized by Darab (1967). Várallyay (1968) related different chemical types of groundwater to different types of soils in the Danube Valley. Specifically, the dominance of sodium was closely related to the occurrence of solonchaks, compared to the dominance of calcium under alluvial soils. Benz et al. (1961), Bazilevich (1970), Arndt and Richardson (1989) have reported relationship between the chemical composition of groundwater and soil salinity. There are numerous theories that explain the differences in the elevation, groundwater depth, concentration, and chemical composition inside basins. Among these are the theory of groundwater flow by Tóth (1984) and the theories proposed for the Danube Valley by Kovács (1960), Várallyay (1968) and Kuti (1989).

The division of large regions into smaller regions is common practice in geography. It is one of the first things done when confronting any specific question. In Hungary, the regions that are distinguished appear to represent salt accumulation. During the transformation of FAK (textural classes of geological formations) and TKT (dominant ion types of groundwater) variables from nominal into numerical data, the algorithm of HOMALS gave similar scale points to the codes of the maps which frequently occurred together.

The precision of predicting the occurrence of SAS systematically increased with the inclusion of the quantitated nominal variables. Logistic regression provided precision that was better, but still insufficient for management. Classification trees proved to be suitable for the prediction of the occurrence of SAS. Compared to the precision of 69% provided by the logistic regression, a 73% was achieved by the classification tree without the new quantitated variables. With the incorporation of TKT1 (quantitated ion type) and TFAK1 (quantitated texture), there was an increase to 96% overall precision when predicting SAS or non-SAS. The significance of the map, "Textural classes of near-surface formations" may be reasoned by the hydrological processes responsible for the formation of salt-affected soils. Additionally, there is a close relationship between the origin and geomorphic location of these formations and the textural classes. Surface sand is found in areas of windblown sand, silt is common in loessic areas, and clay is dominant in recent floodplains. Similarly, surface sand is related to the highest topographic areas inside

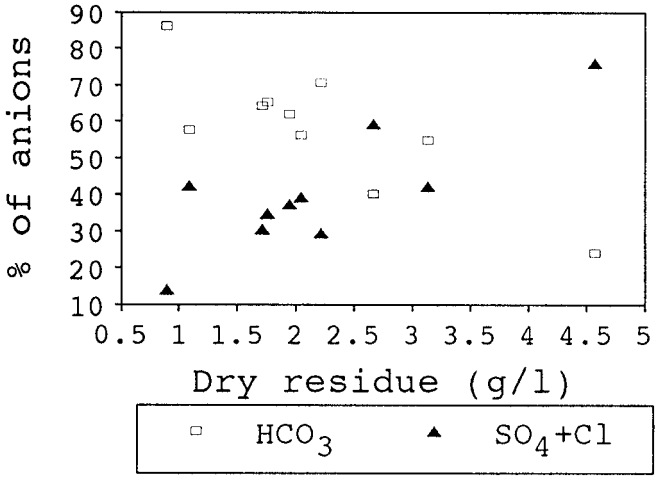

FIGURE 5 Change of dominance of anions in dependence to the concentration of groundwater in the Hortobágy Region, based on data of Rónai (1958).

the Great Hungarian Plain, silt to intermediate loessic plateaus, and clay to the lowest lying recent floodplains. Since the floodplains have most of the salt-affected soils, these can be relatively well predicted by the textural class of the geologic formation.

The large significance of the map "Dominant ion types of groundwater" may be reasoned by the solubility of Na versus Ca and Mg salts. In Table 2, the salt-affected mesoregions show dominance of Na^+ and HCO_3^-. Additionally a great importance is assigned to the dominant anion of groundwater. Employing the data of Rónai (1961), Figure 5 shows that there is an inverse relationship between HCO_3^- and SO_4^{2-} and Cl^- concentration, and at a total salt concentration of 2.5 g L^{-1} SO_4^{2-} and Cl^- become dominant. The original ranges of map "Total soluble salt concentration of the groundwater" (Figure 4 of Tóth et al., 2001) show that this important range of salt concentration parallels the reported range of 1–5 g L^{-1}. For the studies on SAS formation, in the Great Hungarian Plain more detailed information is required. The use of HOMALS and the classification tree improved the utilization of information on anion concentration. An alternative to the direct use of detailed groundwater salt concentrations the anion concentration may provide similar information, especially the dominance of SO_4^{2-} at a groundwater concentration greater than 2.5 g L^{-1}.

The approach shown here for the prediction of the occurrence of soils is readily applicable in those circumstances where there are nominal variables available for the categorization of the influencing factors and where the classical statistical tests cannot be applied.

References

Arany, S. 1956. *Salt-affected soil and its reclamation.* Akadémiai Kiadó, Budapest.

Arndt, J. L., and J. L. Richardson. 1989. Geochemistry of hydric soil salinity in a recharge-throughflow-discharge prairie-pothole wetland system. *Soil Science Society of America Journal* 53:848–855.

Bazilevich, N. I. 1970. *The geochemistry of soda soils.* USDA, NSF and Israel Program for Scientific Translations, Jerusalem.

Benz, L. C., R. H. Mickelson, F. M. Sandoval, and C. W. Carlson. 1961. Groundwater investigations in a saline area of the Red River Valley, North Dakota. *Journal of Geophysical Research* 66:2435–2443.

Bradd, J. M., W. A. Milne-Home, and G. Gates. 1997. Overview of factors leading to dryland salinity and its potential hazard in New South Wales, Australia. *Hydrogeology Journal* 5:51–67.

Breiman, L., and J. H. Friedman. 1985. Estimation optimal transformations for multiple regression and correlation (with discussion). *Journal of the American Statistical Association* 80:580–619.

Breiman, L., J. H. Friedman, R. A. Olshen, and C. J. Stone. 1984. *Classification and regression trees.* Wadsworth and Brooks/Cole.

Darab, K. 1967. Remarks on the study of Dr. H. Franz "Data on the Quaternary stratigraphy and the formation of salt-affected soils in Hortobágy and its periphery." *Agrokémia és Talajtan* 16:459–468 (in Hungarian).

Eilers, R. G., W. D. Eilers, and M. M. Fitzgerald. 1997. A salinity risk for soils of the Canadian prairies. *Hydrogeology Journal* 5:68–79.

Endrédy, E. 1941. On the question of the formation of salt-affected soils. *Öntözésügyi Közlemények* 3(1):207–217 (in Hungarian).

Erdélyi, M. 1979. The hydrodynamics of the Hungarian Basin. *VITUKI Proceedings* 18:1–82.

Gifi, A. 1990. *Nonlinear multivariate analysis.* Wiley, New York.

Herke, S., I. Mihályfalvy, I. Prettenhoffer, E. Tury, and E. Vezekényi. 1959. *Agriculture on our salt-affected soils.* Mezõgazdasági Kiadó, Budapest (In Hungarian).

Kovács, G. 1960. The relationship of salinization and groundwater. *Hidrológiai Közlöny* 1960:131–139 (in Hungarian).

Kreybig, L., and A. Endrédy. 1935. Über die Abhängigkeit des Vorkommens von Alkaliböden im Oberen Tisza-Gebiete Ungarns von der absoluten Höhenlage, pp. 357–360, in *Transactions of the Third International Congress of Soil Science*. Oxford, UK.

Kuti, L. 1989. Young unconsolidated sediments and their chemical interaction with the groundwater stored in them, pp. 441–454, in *Annual report of the Hungarian Geological Institute of 1987*. Budapest (in Hungarian).

Mados, L. 1943. Soil salinization and water. *Hidrológiai Közlöny* 23:3–21 (in Hungarian).

Rónai, A. 1961. *Map of the groundwaters of the Great Hungarian Plain*. Hungarian Geological Institute, Budapest (in Hungarian).

Scherf, E. 1935. Geologische und morphologische Verhältnisse des Pleistozäns und Holozäns der grossen ungarischen Tiefebene und ihre Beziehungen zur Bodenbildung, insbesondere der Alkalibodenentstehung, pp. 1–37, in Annual reports of the Hungarian Geological Institute of the years 1925–28. Budapest.

'Sigmond, A. 1927. *Hungarian alkali soils and methods of their reclamation*. California Agricultural Experiment Station, University of California, Berkeley, California.

Somogyi, S., ed. 1991. *Cadaster of the microregions of Hungary*. Hungarian Academy of Sciences, Budapest (in Hungarian).

S-PLUS, Version 3.3. 1994. StatSci, a division of MathSoft, Inc. Seattle, Washington.

SPSS Categories.1996. SPSS Inc., Chicago, Illinois.

Szabolcs, I. 1974. *Salt-affected soils in Europe*. Martinus Nijhoff, The Hague, The Netherlands.

Tickell, S. J. 1997. Mapping dryland-salinity hazard, Northern Territory, Australia. *Hydrogeology Journal* 5:109–117.

Tóth, J. 1984. The role of regional gravity flow in the chemical and thermal evolution of groundwater, pp. 3–39, in *Proceedings of the First Canadian/American Conference on Hydrogeology*. Banff, Canada.

Tóth, T., M. Kertész and L. Pásztor. 1998. Mapping of salt-affected soils. CD ROM, *Proceedings of the 16th International. Congress of Soil Science. Symposium 29*. Montpellier, France.

Tóth, T., L. Pásztor, S. Kabos, and L. Kuti. 2001. Use of digitalized hydrogeological maps for evaluation of salt-affected soils of large areas. *Arid Land Research and Management* 15:000–000.

Treitz, P. 1924. *The natural history of saline and alkali soils. First part*. Stadium, Budapest (in Hungarian).

Treitz, P. 1934. Saline and alkali soils of Rump Hungary, pp. 176–206, in E. Sajó and Á. Trummer, eds., *The Hungarian salt-affected soils*. Patria, Budapest (in Hungarian).

Várallyay, G. 1968. Salt accumulation processes in the Hungarian Danube Valley, pp. 371–380, in *Transactions of the 9th International Congress of Soil Science*, vol. 1.

Venables, W. N., and B. D. Ripley. 1996. *Modern applied statistics with S-Plus*, corrected third printing. Springer, New York.

Wei, Y. Q., and S. Y. Liu. 1988. The gradation and regionalization methods on prediction of potential salinization in Huang-Huai-Hai Plain of China, pp. 348–353, in *Proceedings of International Symposium on Solonetz Soils*. Osijek, Yugoslavia.