INTERPRETATION AND DISCRIMINATION OF MARSHY WETLANDS BY SOIL FACTORS IN THE KUAN-TU NATURAL PARK, TAIWAN

SHAO-WEI LIAO[1,*] and WEN-LIAN CHANG[2]

[1]*Department of Environmental Engineering and Science, Tajen Institute of Technology, Ping-Tong, Taiwan, Republic of China;* [2]*Department and Graduate Institute of Bioenvironmental Systems Engineering, National Taiwan University, Taipei, Taiwan, Republic of China*
(*author for correspondence, e-mail: swliao@ccsun.tajen.edu.tw)

Abstract. This work investigated soil samples collected from Kuan-Tu wetlands, Taiwan. Factor analysis was performed to explain the impact of various soil factors on this marshy wetlands located in suburban Taipei. The results indicated that the latent factors were heavy metals, salinity, and soil organic matter. Canonical discriminant analysis was used to improve an existing vegetation classification scheme by identifying the physical–chemical properties of sediment in Kaun-Tu wetlands, Taiwan. Predictive discriminant analysis was used to examine the ability of the models to predict class membership for unknown soil sample. Multivariate analysis of the spatial patterns of soil quality and vegetation types showed that different properties of soil grew different types of vegetation and absorbed contaminants differently. We can feasibly conserve a suitable habitat for wetland biology by processing these unstable predictor variables. The methodology and results provide useful information concerning the Kuan-Tu wetlands and may be applicable to other wetlands with similar properties that are experiencing similar environmental issues.

Keywords: discriminant analysis, factor analysis, soil properties, vegetation classification

1. Introduction

Plants are primary producers and their abundance affects the biomass of the ecosystem. In the last few decades, the recovery of habitats from environmental contaminants has become an important subject. Recovery is controlled by hydrology, meteorology and changes in food chain. Numerous studies have illustrated the influence of soil properties on vegetation distribution. Different hydraulic conductivity could impact the conduction of water by infiltration and percolation because of swelling and shrinking in different soil texture especially in coastal wetlands, diverse habitats and biota (Pierzynski *et al.*, 1994). Aber *et al.* (1982) documented, in the prairie peninsula region in Wisconsin, a gradient from silt clay loams to sandy loams along which mean leaf height, canopy layering and species composition changed continuously. With respect to soil chemistry, clay particles influence plant activity significantly because clay colloids had a great specific surface area and thus created a higher cation exchange capacity than coarse soil (Chang and Tang, 1996). Perry (1994) mentioned the existence of ponderosa pine on old fissured lava flows

that allow deep roots to access stored water. Soil texture affects the absorption of contaminants through plant roots and is thus an important factor that governs the distribution of vegetation. Spatial or temporal measurements of physical and chemical properties generally do not reveal the underlying processes directly. Biota chooses suitable sites for living according to their biological features. Polluted water and suspension particles can degrade the quality of soil, reduce primary productivity, and furthermore, affect the living of amphibians, reptiles, insects, fishes, and birds.

Plant communities are influenced by biotic factors (e.g., animals, plants, microorganism, human beings) and abiotic factors (e.g., water, carbon dioxide, light, wind, rain, soil contaminants, soil texture, porosity, soil moisture) (Freedman, 1989). Thomas *et al*. (1999) found that soil characteristics and elevation were strongly correlated with the diversity of native and exotic plants. The spatial pattern of salinity and sodicity can be related to the spatial structure of the vegetation (Amiaud *et al*., 1998). Vegetation participated in the water purification process within a wetland by facilitating sedimentation, storing nutrients in the biomass, and creating suitable microenvironments and carbon sources for beneficial microbes that metabolized pollutants (Luckeydoo *et al*., 2002). Plant parts underground are sources of food for humans, herbivorous animals and insects, and they can influence the type and activity of microorganisms living in and around the plant roots. Schwarz *et al*. (1996) found that nitrogen, phosphorus, potassium, and organic matter on fallow farms influenced wetlands by soil erosion. Nitrogen and organic matter, in particular, affect population dynamics of plants, vertebrates, insects, and birds of wetlands. According to Chang and Tang (1996), the distribution densities of tidal crabs at Kuan-Tu wetlands showed that habitats and mangroves were destroyed by changes of soil texture. Soil moisture and soil texture are two significant factors in determining vegetation distribution. Basically, soil texture affects the distribution of vegetation under the same climatic conditions. Since spatial or temporal measurements of physical and chemical properties generally do not explain the underlying governing processes directly, knowledge of soil properties and vegetation distributions is essential in recovering habitats.

Environmental monitoring usually generates a large bulk of data, which can be difficult to analyze and interpret due to their complex inter-relationship between variables. To extract meaningful information from such databases, multivariate techniques have been used successfully to aid interpretation of complicated field measurements. Factor analysis can determine the most important factors that contribute to structure of the data (Borovec, 1996; Gupta *et al*., 1999; Soares *et al*., 1999; Jenerette *et al*., 2002). Canonical discriminant analysis is used interpret spatial distribution of bioassemblage with various environmental parameters (Cruze-Castillo *et al*., 1994; Mitchell *et al*., 1997; Momen and Zehr, 1998; Ogutu, 1999; Shin and Fong, 1999). In this work, first we analyzed soil properties in the Kuan-Tu wetlands. Second, factor scores were obtained from a factor analysis of soil properties, which were used in interpretation pollutant of this study sites. Third, homogeneous vegetation sites were chosen by field visual reconnaissance. Then canonical

discriminant analysis was used to improve an existing vegetation classification scheme by identifying those sediment physical–chemical properties that can be jointly differentiated. Fourth, we perform predictive discriminant analysis to examine the ability of the models constructed in predicting class membership for vegetation with unknown soil samplers into known classes. And then the variables, to which different vegetation regimes are most sensitive, can be determined. The objective of this paper is to present a description of multivariate analysis and demonstrate its applicability and potential in environment research. This is the first study in Taiwan based on such approach and the results could be useful in providing a methodology for the government in refining its management programmed.

2. Materials and Methods

2.1. Site description

The Kuan-Tu wetlands used to be farms that have been lost due to sinking lands, seasonal floods, invasive tides, and deteriorating irrigation water. This process led to the formation of wetlands complete with aquatic plants. Built in 1954, tidal dikes divided the wetlands into two ecological zones: the north is the Kuan-Tu Natural Park and the south is the Mangrove Conservation region. This study considered only the north wetlands and marshlands. The study site is a salt marsh in northern Taiwan, located in east of the Kuan-Tu flood plain (121°27′E, 25°07′N, Figure 1), suburban Taipei, the total site area of around 56.69 hectares is bordered by the Chi-Lung River to the south, Chung-Kang River to the north, Kuan-Tu Plain to the east, and Tan-Shui River to the west. In past 20 years the population of this area increased rapidly with the majority of the population concentrated northwest of the research area. In 1996, for the first time in Taiwan, the Taipei City government approved 500 million US dollars to establish Kuan-Tu Natural Park, primarily as bird habitat. Water resources in the Kuan-Tu wetlands included the tidal Tan-Shui River and the Chi-Lung River from the south, the domestic wastewater of Kuei-Tzu-Keng Canal and Shui-Mo-Keng Canal from the north, and the agricultural wastewater of Pa-Hsien Canal from the east. South of the tidal dikes is the Kuan-Tu Natural Reservation, which is influenced by Chi-Lung River, Tan-Shui River, and coastal tides; vegetation types are defined as mangrove ecosystems. According to surveys conducted by Zoology Institute, Academia Sinica, Taiwan, Kuan-Tu Natural Park was home to 153 species of plants of 118 genera and 110 species of birds recorded during the period between September 1998 and June 2000 (Taiwan Taipei City Government (TWTCG), 1998–2000).

Soil salinity showed marked differences among the wetlands with fresh water to the east and those with alkaline water to the south. On the eastern side, the main irrigation systems of the 150 hectares Kuan-Tu rice paddy fields came
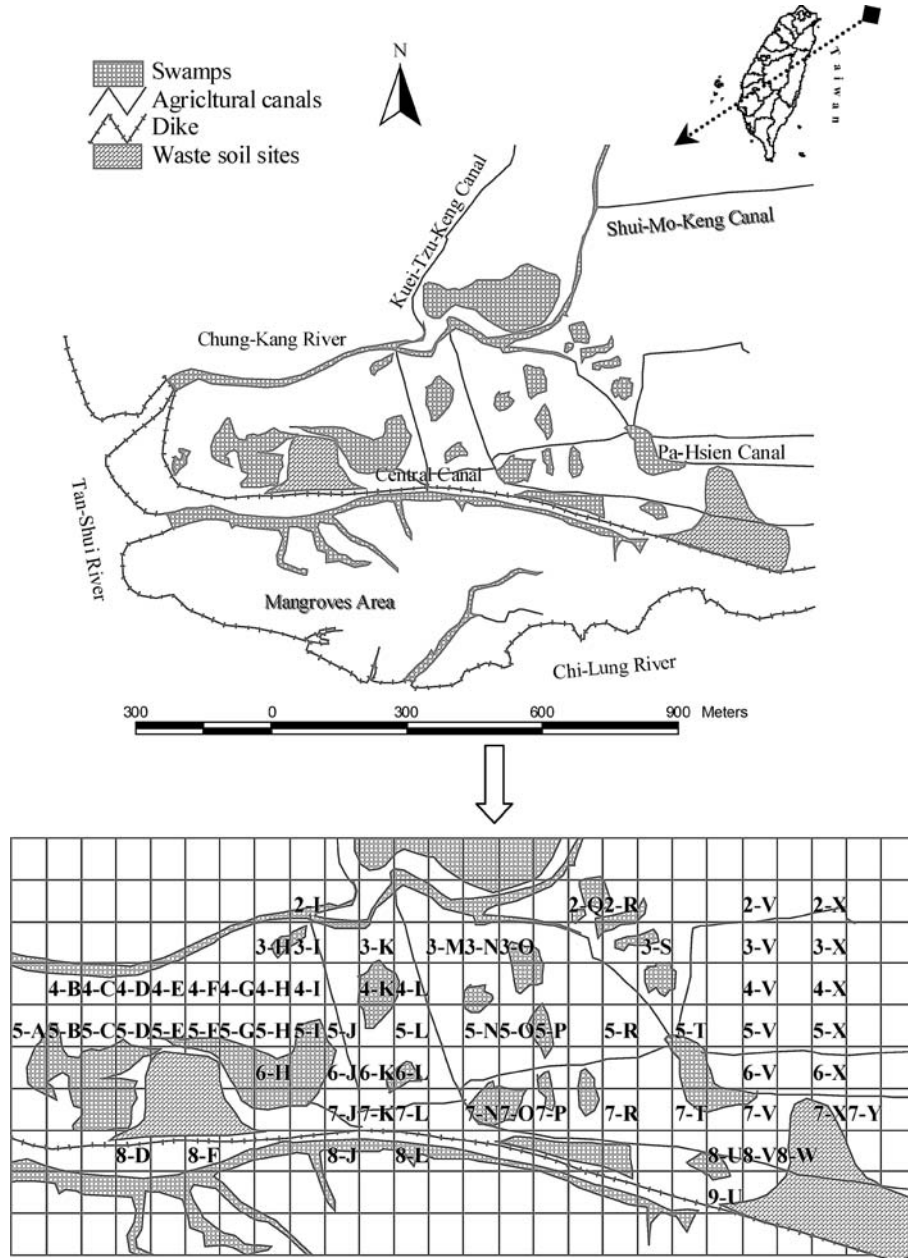
Figure 1. Soil sampling stations in Kuan-Tu wetland, Taiwan.

from the Sulfur Canal, through the Pa-Shien Canal. The primary sources of pollution in Kuan-Tu Plain, include domestic and municipal wastewater and agricultural fertilizers, that have resulted in eutrophication and degraded the water quality.

2.2. SOIL COLLECTION AND ANALYSIS

Grid sampling at Kuan-Tu wetland was undertaken at low tide in October 1998. The Kuan-Tu wetlands were divided into a 10×25 grid by rectangular grid method. Each grid has an area of 3000 m^2 and soil sample was collected compositively in each grid. Due to technical infeasibility (precarious terrain, dangerous waterways) and instrument limitation (some of the measurements fell below the minimum detectable limits), 69 sampling sites were valid out of the 250 potential sampling sites. The triplicate samples taken from each sampling site were mixed together, then, the 16 measurements were taken in the laboratory. Thus, a 69×16 data matrix was generated. Owing to soil, the surface horizon was full of intertwined plant roots, subsurface soil within 10–20 cm was samples and stored in sturdy, tightly sealed plastic bags in icebox with ice and analyzed as soon as possible. Soil samples were air-dried outdoors about 25–30 °C and ground using a wooden pestle, and then passed through a 2 mm sieve. Soil texture which is a manifestation of their component proportions of sand, silt, and clay, by Buoyancy's hydrometer method (D-422; ASTM, 1994). Samples were individually placed in separate aluminum flat plates, and dried at 105 °C to measure the moisture percentage (W). Organic matter with regard to soil was performed kilograms organic carbon by soil volume cubic meter (kg/m^3). An ash-free dry weight determination was used to approximate soil organic matter (SOM) that is calculated as the difference between the weight of the crucible containing the oven-dried sample and the weight of the crucible with the ash (Brower *et al.*, 1998). pH value and electrical conductivity (EC) of soil samples were measured at a 1:1 soil–water extract ratio by pH meter (WTW pH-320/Set-2) and EC meter (WTW LF 320/Set), respectively. Apparent specific gravity (ASG) was determined by dividing the dried solid weight by the total mass volume. Concentrations of Cu, Ni, Pb, Mn, Fe, Zn, and Cd were analyzed by ICP-AES (ICP LIBERTY II) using an ultra sonic nebulizer. The minimum detection limits were found to be 1.2, 2.3, 3.1, 1.4, 1.3, 1.1, and 1.2 μg/kg, respectively.

2.3. PLANT RECOGNIZATION

In soil sampling, also we recognized plant distribution by a priori classification. The a priori classification indicated that the habitat could be classified into four regions – short marshes, rice paddy fields, tall marshes, and drier areas. Figure 2 presents vegetation types classified by field visual reconnaissance to select the sampling sites with homogeneous plant species. From the field observation, one can detect a range of short marshes in 24 soil sampling stations, the dominant species of which were *Paspalum distichum* L., *Cynodo dactylon* Pers., *Brachiaria mutica* (Forsk.) Stapf., *Panicum repens* L., and *Alternanthera philoxeroides* (Mart.) Griseb, which is mostly located west of the Central Canal. In rice paddy field included 20 soil sampling stations, are all located in the east of the study area. In tall marshes area included 16 soil sampling stations, all of which are located east of the Central Canal
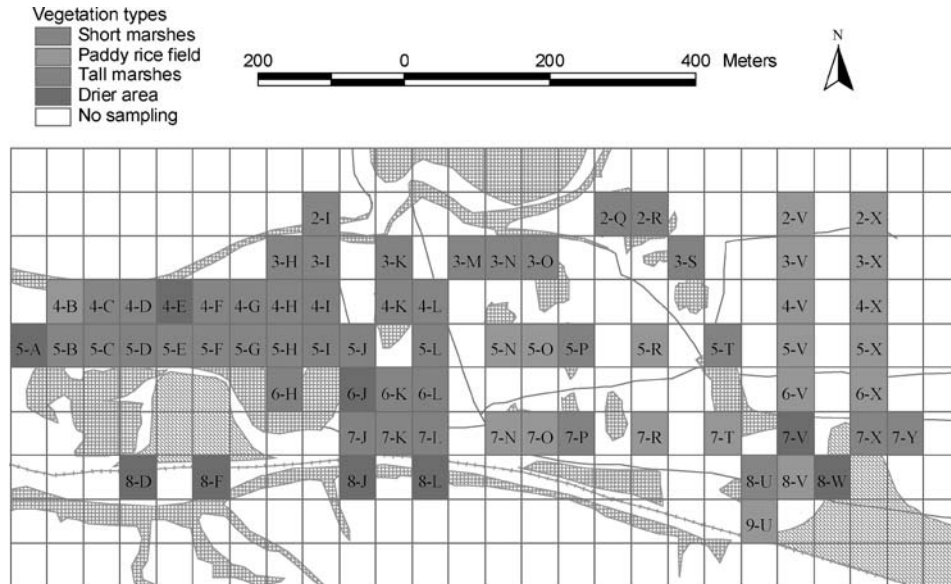
Figure 2. Vegetation types classified by field visual reconnaissance in Kuan-Tu wetland.

and west of the rice paddy fields. The dominant species were *Arundo donax* L., *Typha orientalis* Presl, and *Cyperus malaccensis* Lam. subsp. *Monophyllus* (Vahl.) *T. Koyama*. And the drier area included nine soil sampling stations, which are located in the neighboring waste soil sites. We selected the sampling sites with homogeneous plant species – 14 short marsh sites, 10 rice paddy field sites, 15 tall marsh sites, and 8 drier area sites. Forty-seven homogeneous vegetation sites were used to create the canonical discriminant functions. Twenty-two remaining sites were retained to evaluate the ability of the constructed model.

2.4. STATISTICAL METHODS

2.4.1. *Factor Analysis*

Factor analysis identifies a few underlying dimensions of many variables. These dimensions can be used as a new set of variables, called latent factors, which are neither observed nor expressible in terms of the observed variables. Pair wise linear correlations between the soil variables were determined. The analysis was conducted on a correlation matrix, rather than a variance–covariance matrix, to eliminate any effects of different measurement units in the determination of factor loading (Johnson and Wichern, 1992; Sharma, 1996). A varimax rotation redistributes the variance of each variable so that each variable has a high loading on only one factor with near-zero loadings on other factors to obtain easy interpretation of factors (Sharma, 1996). Factor analysis consists of a family of procedures for removing the redundancy from a set of correlated variables and represents the

variables with a small set of latent factors. And factor scores are commonly obtained by two approaches: the weighted least squares method and the regression method. The regression method (Johnson and Wichern, 1992) is used in this study to compute factor scores.

2.4.2. *Discriminant Analysis*

Canonical discriminant analysis (CDA) determines how a set of quantitative variables may differentiate among several known classes. CDA obtains linear functions of quantitative variables that maximally separate two or more groups of individuals, while minimizing variation within groups (Rencher, 1992). This method distinguishes uncorrelated canonical discriminant functions (CDFs), which are linear combinations of the original variables that best separate the means of groups of observations (Rencher, 1992). The maximum number of CDFs equals the number of independent variables or the number of class variables minus one whichever is smaller. The first CDF, denoted as CDF_1, quantifies the maximum possible variation between groups. CDF_2 reflects group differences not captured by CDF_1, while CDF_1 is not correlated with CDF_2. Similarly, CDF_3 is not correlated with CDF_1 or CDF_2, and reflects group differences not displayed by CDF_1 or CDF_2, and so on (Sharma, 1996). This implies that each CDF extracts a unique dimension of information from the dataset. This method also allows the relationships among groups to be a graphically represented by plotting the values of canonical scores for sample observations. As well as identifying outliers in the data, these plots may be used to assign a new observation to an existing group. The canonical scores for the new individual are calculated for the first two (or few) CDFs and its position is plotted. The new individual is assigned to the group whose mean is closest to its position.

Stepwise multiple discriminant analysis was conducted to correlate the vegetation groupings with the physical and chemical variables obtained from the sediment samples. These datasets (soil properties: electrical conductivity (EC), pH, % sand, % silt, % clay, apparent specific gravity (ASG), % soil moisture (W), soil organic matter (SOM), nickel (Ni), copper (Cu), lead (Pb), cadmium (Cd), manganese (Mn), zinc (Zn), iron (Fe) and osmosis pressure (OS)) were divided into groups according to the plant community pattern that resulted from classification and ordination. The class variables consisted of four vegetation classes obtained by field visual reconnaissance. The calculation was performed using the STATISTICA package from StatSoft (1996).

3. Results and Discussion

3.1. Factor analysis

Valid samples were obtained from 69 sampling stations because some sampling sites were technical infeasibility (precarious terrain, dangerous waterways) and some

sample measurements fell below the minimum detection limits of the instruments (Figure 1). A 69 × 16 data matrix was generated after the determination of 16 measurements of physical and chemical properties of the soil in a laboratory. First, data from soil sampled in October 1998 were analyzed to identify underlying factors for the observed responses. Each variable was standardized to a mean of zero and a variance of one, to prevent variables with large variances unduly influences the factor construction. Table I presents the correlation matrix for the 16 variables. The five most significant factors were extracted according to the "eigenvalue-greater-than-one" rule, while Table II shows the factor-loading pattern, the eigenvalues that represent the factors, and the proportion of total variance of the sample that is explained by each factor. The result of factor analysis of the five most significant latent factors indicated that they explained 76.36% of the total variance of the sample (Table II). Hence, the five latent factors can be used to explain mainly the background physical and chemical processes of the soil, without loss of any significant characteristics.

Factor loadings are simple correlations between the original soil variables and the factors. The first latent factor accounted for 26.18% of the total variation in the soil variables. It meant sand, Cu, and Zn represents the underlying dimensions that account for the correlation among the variables. So we termed the first latent factor as the "heavy metal factor" because soil particles can attract cation charges, and both Cu and Zn are heavy metals. Table II showed that Cu and Zn had high positive factor loadings with the first latent factor, which meant it had high positive correlation with the first factor. Factor scores were calculated to determine the level of pollution, in Figure 3, the darker and lighter shades represented the higher and lower concentration level of contaminants, respectively. As seen in Figure 3, Cu and Zn are most abundant in the south and west areas especially at the tidal dike nearby the Central Canal – at sampling stations 7-J, 7-K, 7-L, 7-P and 8-L, which coincided with the laboratory data. Based on our field measurements, gate of the tidal dike is 2-m wide and the water level is 1.45 m. The in-flow of tidal water resulted soil contamination due to heavy metal precipitation. Three main pollutant sources were: (1) domestic and industrial wastewater from Shui-Mo-Keng Canal, (2) auto-body-shop wastewater from the east, and (3) the domestic wastewater from the north, the Pei-To district. The population of this area has increased rapidly in the past 20 years and the majority of the population is concentrated in the northwest of the research area. Furthermore, the results were affected by the altitude, tides, and accumulation of metal. The dominant soil texture was clay or clay loam in this area.

The second latent factor explained 18.46% of the total variation in soil variables and showed high positive factor loadings for Cd and Fe (Table II), the second latent factor was also termed as the "heavy metal factor". Concerning the level of pollution, as seen in Figure 4, the areas with the highest concentration levels of Cd and Fe were found in the east, which coincided with the laboratory data, especially in the rice paddy fields at the end of Pa-Hsien Canal. Soil

TABLE I

Pairwise correlation coefficients between physical and chemical properties of Kuan-Tu wetland's soil

	EC	pH	Sand	Silt	Clay	ASG	W	SOM	Ni	Cu	Pb	Cd	Mn	Zn	Fe	OP
EC	1.00															
pH	−0.14	1.00														
Sand	**−0.27**	**0.45**	1.00													
Silt	**−0.27**	**−0.27**	**−0.54**	1.00												
Clay	**0.53**	**−0.32**	**−0.75**	−0.15	1.00											
ASG	**−0.41**	0.07	0.10	0.04	−0.15	1.00										
W	**0.40**	**−0.29**	**−0.33**	0.07	**0.33**	**−0.67**	1.00									
SOM	−0.10	**−0.32**	**−0.25**	−0.10	**0.37**	0.03	**0.31**	1.00								
Ni	0.09	0.02	0.06	0.06	−0.11	**−0.24**	**0.30**	−0.01	1.00							
Cu	0.02	0.09	**0.45**	−0.15	**−0.41**	−0.10	0.15	−0.19	**0.59**	1.00						
Pb	−0.15	0.20	**0.28**	0.02	**−0.34**	0.05	0.10	−0.04	**0.29**	**0.36**	1.00					
Cd	0.06	−0.11	0.03	0.22	−0.21	−0.22	**0.38**	0.09	**0.34**	**0.29**	**0.36**	1.00				
Mn	−0.13	**0.41**	**0.48**	**−0.34**	**−0.29**	−0.02	−0.01	0.07	0.11	0.14	**0.58**	0.19	1.00			
Zn	−0.10	**0.29**	**0.72**	**−0.48**	**−0.48**	0.13	**−0.28**	−0.20	−0.01	**0.61**	**0.33**	−0.06	**0.35**	1.00		
Fe	−0.05	**−0.40**	−0.23	**0.45**	−0.08	−0.19	**0.35**	0.09	0.20	0.14	0.06	**0.56**	**−0.25**	−0.16	1.00	
OP	**−1.00**	0.14	**0.27**	**0.27**	**−0.53**	**0.41**	**−0.40**	0.10	−0.09	−0.02	0.15	−0.06	−0.06	0.10	0.05	1.00
S.D.	4743	0.76	9.28	6.16	7.91	0.26	12.18	23.48	16.48	11.93	6.02	6.02	34.91	125.6	1420	173

Note: The abbreviations used from Tables I–IV were EC, electrical conductivity; ASG, apparent specific gravity; W, soil moisture; SOM, soil organic matter; OP, osmosis pressure; S.D., standard derivation (bold values are significant at $p < 0.05$).

TABLE II

Rotated factor loadings of a model, with five factors, of Kuan-Tu wetland's soil

Soil properties	Factor loadings				
	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5
EC	0.01	−0.02	**−0.94**	−0.12	−0.04
pH	0.15	−0.26	0.05	0.63	−0.45
Sand	**0.77**	−0.06	0.25	0.38	−0.21
Silt	−0.63	0.44	0.36	−0.27	−0.27
Clay	−0.42	−0.27	−0.58	−0.23	0.46
ASG	0.11	−0.38	0.61	−0.11	0.04
W	−0.20	0.56	−0.54	0.05	0.34
SOM	−0.12	0.04	0.08	0.03	**0.93**
Ni	0.20	0.62	−0.18	0.10	−0.06
Cu	**0.74**	0.52	−0.05	−0.01	−0.15
Pb	0.18	0.43	0.15	0.63	0.00
Cd	−0.03	**0.78**	−0.02	0.17	0.06
Mn	0.22	0.05	0.06	**0.88**	0.12
Zn	**0.88**	−0.06	0.12	0.16	−0.10
Fe	−0.14	**0.73**	0.12	−0.36	0.07
OP	−0.01	0.02	**0.94**	0.12	0.04
Eigenvalues	4.19	2.95	2.52	1.44	1.11
Total variance (%)	26.18	18.46	15.76	9.02	6.93
Cumulative (%)	26.18	44.64	60.40	69.43	76.36

texture in this area was clay loam or silty clay loam. This result was an evidence of industrial pollution because natural soil had little concentration of Cd. A potential source of pollution was from the auto-body-shop wastewater from the east through Pa-Hsien Canal. Heavy metal accumulated in the intersection of the canal, because the water level was not enough to wash out pollutants during flooding.

The third factor explained 15.76% of the total variation in the soil variables and showed high positive factor loadings for osmotic pressure, and a high negative factor loading for electrical conductivity (Table II). Electrical conductivity will cause salinity stress to the freshwater plants. So we termed the third latent factor as the "salinity factor". The salinity factor score varied significantly (with factor loading higher than 0.7). The lighter shade in sampling stations 5-J, 6-J, 7-J and 7-K, 7-L in Figure 5 – neighboring the Central Canal – showed that high electrical conductivity and low osmotic pressure was caused by the tides. The function of evapotranspiration causes the accumulation of salinity and increasing electrical conductivity in the west and neighboring Central Canal. The main soil texture in the lightly shaded areas was clay or clay loam.
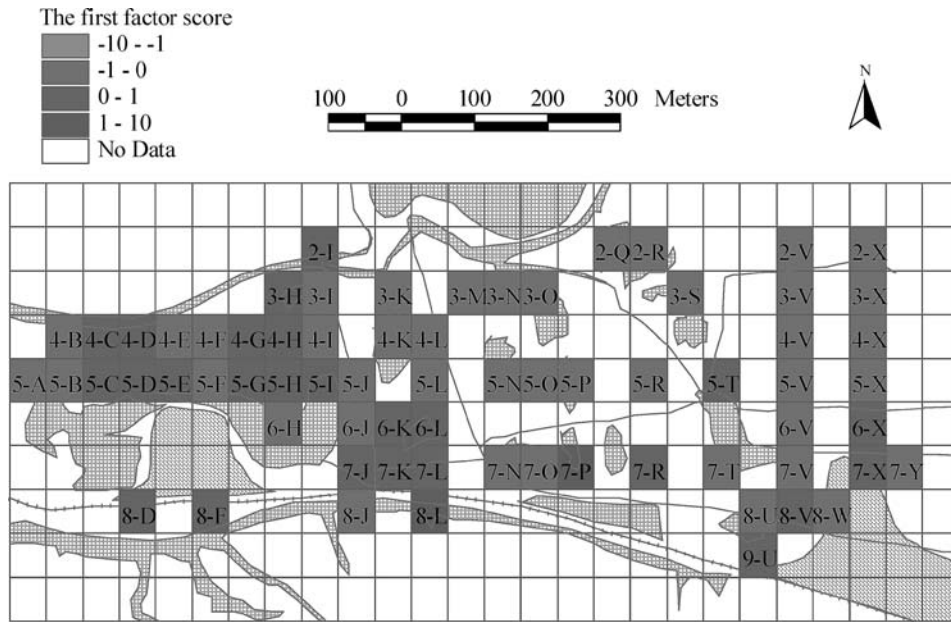
Figure 3. Distribution diagram for the first factor (sand, Cu, and Zn) score.
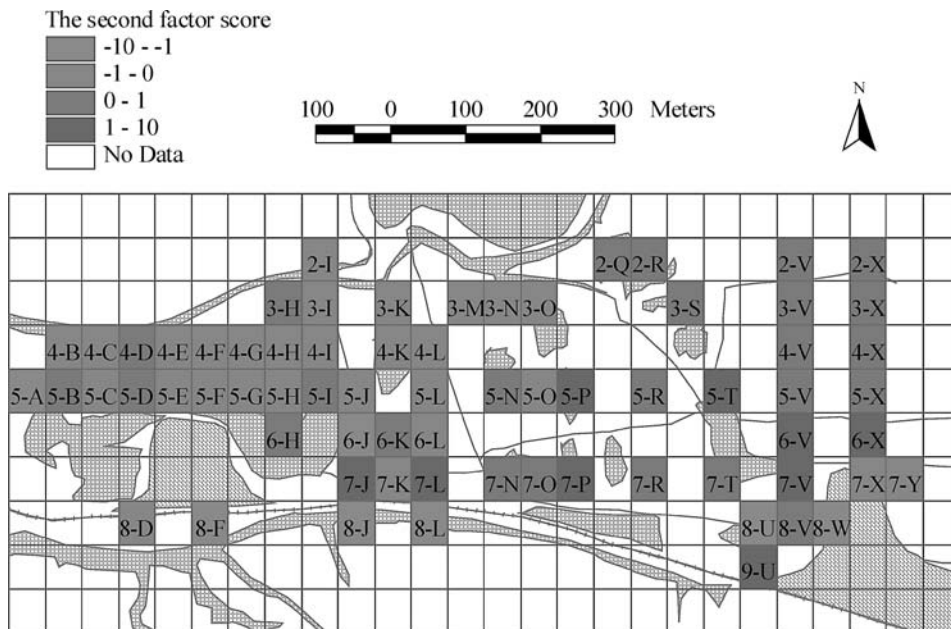

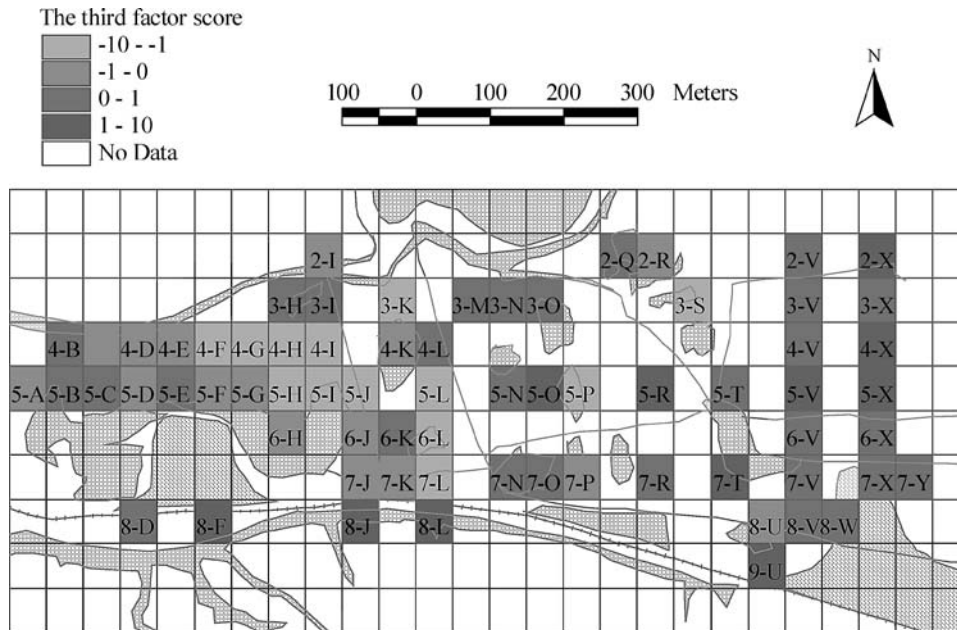*Figure 4*. Distribution diagram for the second factor (Cd and Fe) score.

Figure 5. Distribution diagram for the third factor (osmotic pressure) score.

The fourth factor explained 9.02% of the total variation in the soil variables, and had a high positive factor loading for Mn (Table II). Similarly, the fourth factor was termed "heavy metal factor" because Mn is a heavy metal. The higher fourth factor scores were distributed over two waste soil sites and scattered in western and northern parts (Figure 6). Further investigation is required to determine the relationships between pollutant sources and neighborhood waste soil sites. Furthermore, the accumulation of Mn may be caused by irrigation water quality and industrial wastewater.

The fifth factor explained 6.93% of the total variation in the soil variables and showed a high positive factor loading for SOM (Table II). The fifth factor was termed the "soil organic matter factor" because SOM was determined by the ratio of the difference between the oven-dry weight and the ash-free dry weight, to the oven-dry weight. The darker shade represented higher primary production, and the ones, distributed in the western sampling sites indicated an area of short stem grasses (Figure 7). The soil texture in this area was clay. Soils with high clay content and small pore size impede drainage more than sandy or loamy soils. In addition, high clay content soil can resist the decomposition of litter and other organic matter in the floodplain (Mitsch and Gosslink, 1993). Seasonal changes lead short marshes to evolve quickly in this area. Plant residues in wetlands, standing water or poorly drainage will cause anaerobic conditions. Accordingly, rotted and decayed plants yielded more organic matters in soils.
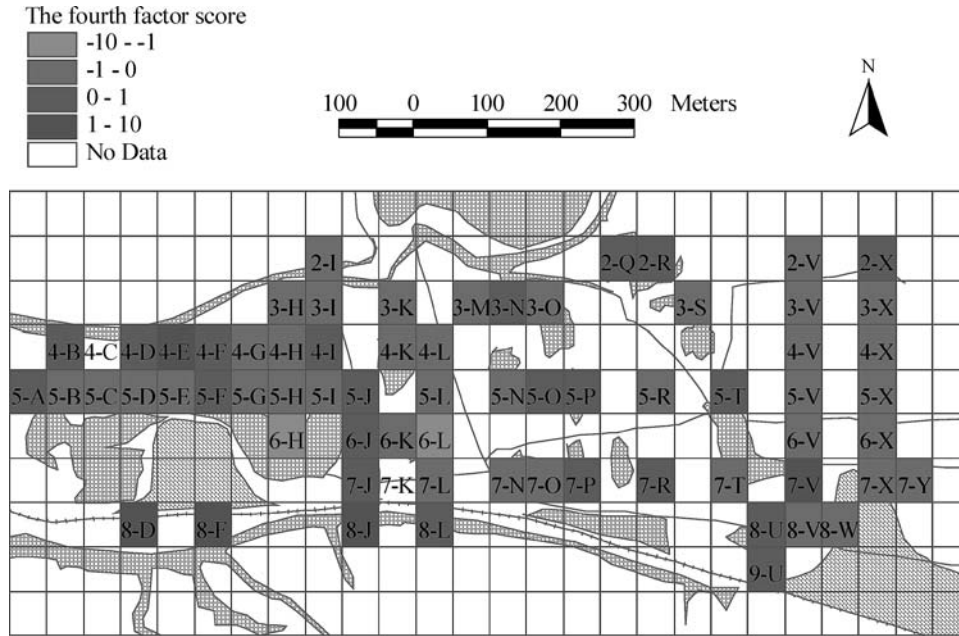
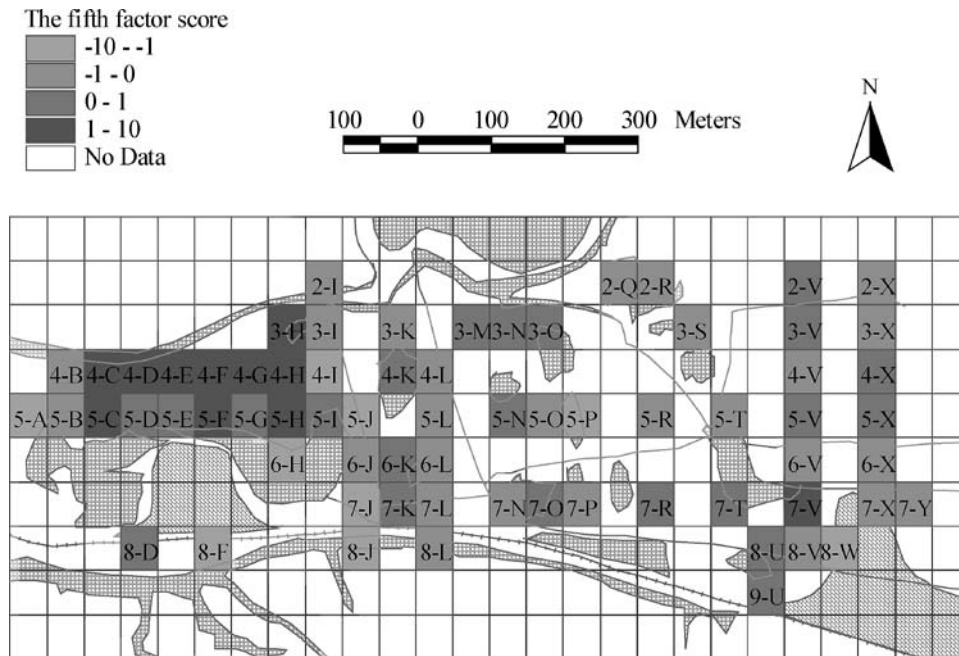*Figure 6.* Distribution diagram for the fourth factor (Mn) score.

Figure 7. Distribution diagram for the fifth factor (soil organic matter) score.

3.2. DISCRIMINANT ANALYSIS

By field visual reconnaissance – 14 short marsh sites, 10 rice paddy field sites, 15 tall marsh sites, and 8 drier area sites were chosen. Forty-seven homogeneous vegetation sites with 15 variables were used to create the canonical discriminant functions (CDFs). Twenty-two remaining sites were retained for classification using the constructed CDFs. The osmosis pressure term was deleted to expedite the discriminate analysis and avoid the problem of co-linearity because electrical conductivity was perfectly negatively correlated with osmosis pressure (Table I). Because the number of class membership (four vegetation types) was smaller than the number of independent variables (15). Consequently, four vegetation classes and three CDFs were determined. Two main statistics explain the properties and structure of the canonical discriminant functions – (1) the total canonical structure coefficients that indicates correlation coefficients between individual variables and canonical scores (similar to variable loadings in factor analysis) and (2) total standardized canonical coefficients are multipliers of standardized independent variables and produce standardized canonical scores. The validity and the use of these two indices have been discussed but controversies still remain (Rencher, 1992; Huberty, 1994; Matthew *et al*., 1994; Cruze-Castillo *et al*., 1994; Momen and Zehr, 1998). Total standardized canonical coefficients reflect the joint effects of independent variables of a given CDF, and thus, are more informative than the total canonical structure coefficients (Rencher, 1992). Whereas, total standardized canonical coefficients can be misleading when independent variables are interrelated (Cruze-Castillo *et al*., 1994). In this work, the total canonical structure coefficients was used to interpret the CDFs because of significant correlations among some of the independent variables (Table I), and the total standardized canonical coefficients was used to construct the CDFs.

Assume that vegetation classes were affected by these main sediment parameters. A forward stepwise method was used to determine which variables could be included in the model (StatSoft, 1996). Accordingly, an *F*-test was applied to select the most discriminating variables. The process was terminated, when differences ceased to be significant. Table III shows significance among these discriminate variables. The order of inclusion in the model, according to the *F*-test, was – EC, Ni, SOM, pH, Mn, Cu, Cd, Fe, Silt, Zn, and ASG. Only 11 out of 15 analyzed parameters were proven important in discriminating the vegetation groups. Four parameters, which were not shown to be important in differentiating, were clay, sand, Pb, and W. The former EC, Ni, SOM, and pH, was the order of discriminatory ability and these variables were most significant to these model. EC affects a plant's ability to conduct water, and identify a durable halophyte. The presence of Ni was mostly produced by human activities, and content in plant tissue varied according to vegetation bioaccumulation. SOM related to the amounts of vegetation litter. And pH indicates the soil's acidity or alkalinity. From test of Wilk's Lambada that sand, clay, W, and Pb, showed insignificance in statistics. Poor discrimination of vegetation by coarse

TABLE III

Mean values and differences of the discriminate variables for the four vegetation classes

Discriminate variable	Wilk's Lambda	F-value	p-level	Mean (standard derivation) values of discriminate variables			
				Short marshes	Rice paddy fields	Tall marshes	Drier area
EC (μs/cm)	0.039	26.15	0.000	8161.43 (3167.72)	1297.90 (618.62)	12243.0 (4485.00)	3284.63 (2134.99)
pH	0.016	4.53	0.009	5.53 (0.59)	5.71 (0.27)	6.21 (0.59)	7.37 (0.71)
Fe (mg/kg)	0.014	2.55	0.073	3522.29 (907.26)	5047.70 (982.88)	4224.53 (1262.57)	1685.50 (732.97)
SOM (kg/m^3)	0.018	6.06	0.002	109.80 (26.25)	93.07 (11.79)	68.38 (12.47)	70.68 (32.29)
Mn (mg/kg)	0.016	3.74	0.020	54.35 (23.65)	38.42 (10.74)	48.95 (20.14)	112.13 (67.42)
Ni (mg/kg)	0.019	7.05	0.001	14.05 (11.18)	6.92 (0.95)	23.33 (31.22)	6.30 (4.49)
Cd (mg/kg)	0.015	3.06	0.042	0.28 (0.12)	0.42 (0.14)	0.55 (0.35)	0.20 (0.34)
ASG (g/cm^3)	0.013	1.76	0.174	0.86 (0.29)	0.82 (0.11)	0.63 (0.27)	0.91 (0.31)
Silt (%)	0.014	2.04	0.127	35.86 (5.82)	45.60 (1.90)	40.07 (4.82)	32.50 (7.93)
Cu (mg/kg)	0.015	3.22	0.035	18.46 (3.52)	22.17 (5.36)	29.83 (15.51)	22.52 (22.65)
Zn (mg/kg)	0.014	2.01	0.132	42.83 (15.27)	35.93 (12.76)	59.67 (22.89)	186.27 (361.26)
Sand (%)	0.011*	0.78	0.516	19.21 (4.23)	18.80 (3.16)	21.27 (4.82)	36.50 (19.38)
Clay (%)	0.011*	0.81	0.500	45.00 (5.56)	35.60 (3.20)	38.73 (7.22)	31.00 (12.39)
W (%, v/v)	0.011*	0.58	0.631	65.44 (10.11)	66.35 (7.14)	70.27 (9.85)	49.83 (21.28)
Pb (mg/kg)	0.011*	0.62	0.605	12.21 (3.67)	13.84 (2.60)	15.05 (6.32)	18.34 (13.72)

Note: (*) Indicating that discriminate variables had insignificance (did not choose when Wilk's Lambada <0.012, or p-values 0.15), and the value under the parentheses was standard derivation.

TABLE IV

Total standardized canonical coefficients (TSCC) and total canonical structure coefficients (TCSC) between canonical discriminant functions (CDF_1, CDF_2, CDF_3) and discriminant variables

Discriminant variable	CDF_1		CDF_2		CDF_3	
	TSCC	TCSC	TSCC	TCSC	TSCC	TCSC
EC	**−1.25**	**−0.58**	0.25	0.15	0.24	0.16
pH	0.11	0.14	0.39	**0.57**	−0.58	−0.25
Fe	−0.17	−0.17	**−0.49**	**−0.48**	−0.18	−0.35
SOM	0.30	0.10	**−0.69**	−0.31	0.34	**0.44**
Mn	0.50	0.16	**0.62**	**0.36**	0.13	0.10
Ni	**−1.35**	−0.17	0.09	0.02	**1.21**	−0.01
Cd	**−0.74**	−0.17	0.06	−0.07	−0.17	−0.26
ASG	−0.28	0.16	0.01	0.00	0.47	0.16
Silt	−0.08	−0.04	−0.32	−0.33	−0.53	**−0.42**
Cu	**1.18**	−0.10	0.004	0.06	**−1.50**	−0.18
Zn	**−0.98**	0.07	−0.19	0.18	**0.84**	−0.01
Chi-square test	171.54		100.75		41.78	
Canonical correlation	0.92		0.89		0.81	
Eigenvalue	5.29		3.63		1.96	
Cumulative variance explain	0.49		0.82		1.00	

and fine soil texture however influenced the capacity to hold water significantly, such that soil moisture also poorly discriminated among vegetation.

Canonical correlation coefficients (the square root ratio of the between-group sum of squares to the total sum of the squares for a given CDF) exceeded 0.8 for the first three CDFs (Table IV). Eigenvalues (the ratio of between-group sum of squares to within-group sum of squares for a given CDF) also exceeded 1.0 for the first three CDFs. These three CDFs together explained for 100% (49, 33, and 18%, respectively) cumulative variance for 47 sampling sites. The value of every discriminant variable was standardized to find the relationship between the discriminant variables and functions. The standardized CDFs were obtained as follows.

$$CDF_1 = -1.25EC + 0.11pH - 0.17Fe + 0.30SOM + 0.50Mn - 1.35Ni$$
$$- 0.74Cd - 0.28ASG - 0.08silt + 1.18Cu - 0.98Zn$$

$$CDF_2 = 0.25EC + 0.39pH - 0.49Fe - 0.69SOM + 0.62Mn + 0.09Ni$$
$$+ 0.06Cd + 0.01ASG - 0.32silt + 0.004Cu - 0.19Zn$$

$$CDF_3 = 0.24EC - 0.58pH - 0.18Fe + 0.34SOM + 0.13Mn - 1.21Ni$$
$$- 0.17Cd + 0.47ASG - 0.53silt - 1.50Cu + 0.84Zn$$

TABLE V

Summary of scores for vegetation classes of the first three canonical discriminant functions (CDFs)

Vegetation classes	Sample number	Mean and standard deviation		
		CDF_1	CDF_2	CDF_3
Short marshes	14	$-0.20\,(\pm0.96)$	$-0.86\,(\pm0.85)$	$1.95\,(\pm1.03)$
Rice paddy field	10	$2.14\,(\pm0.63)$	$-2.35\,(\pm0.60)$	$-1.40\,(\pm0.87)$
Tall marshes	15	$-2.77\,(\pm1.19)$	$0.64\,(\pm1.18)$	$-0.87\,(\pm1.21)$
Drier area	8	$2.87\,(\pm1.05)$	$3.24\,(\pm1.24)$	$-0.04\,(\pm0.50)$

Vegetation classes were discriminated by three CDFs and Table V summarizes mean of discriminant scores for the four vegetation classes. CDF_1 defined by EC, Ni, Cd, Cu, and Zn, that significantly discriminated between tall marshes and drier area because it had the highest difference of average values $2.87(\pm1.05) + 2.77(\pm1.19) = 5.64(\pm1.59)$, meaning that CDF_1 discriminated between these two vegetation classes the most. CDF_1 had the highest canonical correlation of 0.92. With respect to canonical structure, EC was a factor that most affected CDF_1 so we termed CDF_1 "the salinity factor" (Table IV). In tall marshes, the mean values of EC, Ni, Cd, and Cu were significantly higher than the corresponding values in the drier area and higher than any other vegetation class (Table III). Tall marshes exhibited the highest EC because the tide brought in salinity and accumulated in porous soil. Higher contaminants of Ni, Cd, and Cu came from industrial pollution. Heavy metals do not easily dissolve in water, but adsorb and accumulate in soil colloids and plant tissues, so their content in tall marshes site is higher than almost any other class of vegetation. The drier area was full of architectural waste soil and the highest content of Zn was come from these unknown waste soils.

CDF_2 significantly discriminated between rice paddy fields and drier areas because it had a highest difference of average values $3.24(\pm1.24) + 2.35(\pm0.60) = 5.59(\pm1.38)$, and a scatterplot of the scores of the first two canonical discriminant functions (CDF_1 and CDF_2) indicated a separation of vegetation types (Figure 8), and had less difference of average values than CDF_1 (Table V). CDF_2 had the higher canonical correlation of 0.89, meaning that CDF_2 most effectively discriminated between these two vegetation classes. Variables Fe, SOM, and Mn defined CDF_2, such that these variables most importantly discriminated among these types of vegetation. With respect to canonical structure, Fe, Mn, and pH were the factors that most affected CDF_2. Fe is one of the most important components in soil genesis, and much Mn is produced by human activity so CDF_2 was termed, "the heavy metal factor" (Table IV). In rice paddy fields, the mean values of Fe and SOM contents were considerably higher than the content in drier area (Table III). Soluble organic carbonate (humic acid and fulvic acid) combined with Fe in the fountainhead and formed a compound that ran off to the lake or river basin (Vaughan and Ord, 1994).
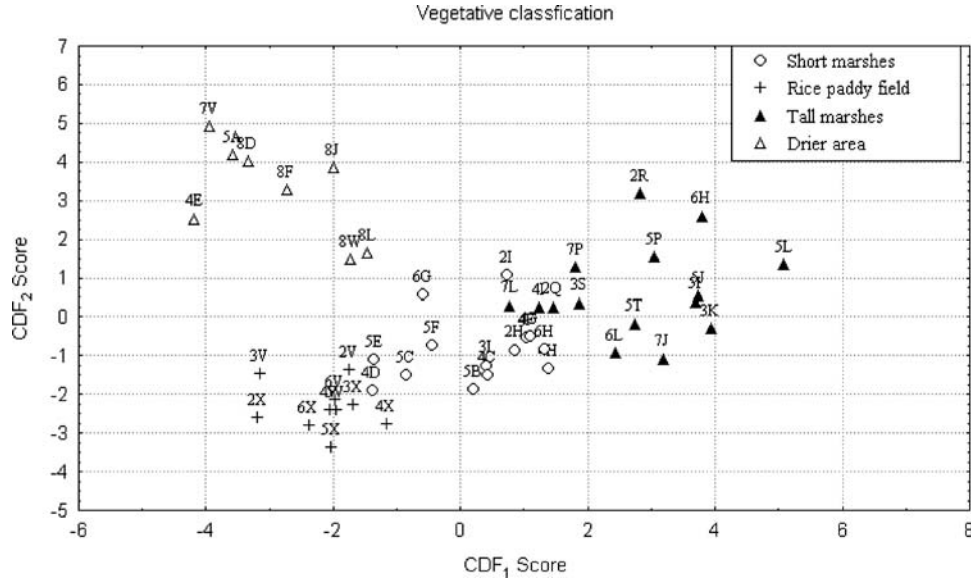
Figure 8. Scatterplot of the first two canonical discriminant functions (CDF$_1$ and CDF$_2$).

Transport, followed by accumulation in the wetlands, explains why the content of Fe exceeded that of any other element. SOM content was higher in this area mainly because of littering during harvesting. In rice paddy fields, the content of Mn was lower than in the drier area. The content was highest here because of an unknown waste soil.

CDF$_3$ significantly discriminated between short marshes and rice paddy fields because it had the highest mean difference 1.95($\pm$1.03) + 1.40($\pm$0.87) = 3.35($\pm$1.35) (Table V). A scatterplot of the scores of CDF$_1$ and CDF$_3$ indicated a separation of vegetation types (Figure 9). CDF$_3$ had canonical correlation coefficient of 0.81, meaning that CDF$_3$ most effectively discriminated between these two vegetation classes. CDF$_3$ was related to the variables Ni, Cu, and Zn. The canonical structure shows that SOM and % silt were the factors that most affected CDF$_3$. Organic matter and soil texture affected the adsorption of heavy metal in soil, so CDF$_3$ was termed, "the organic matter factor" (Table IV). In short marshes, the mean contents of Ni and Zn were significantly higher than the corresponding content in rice paddy fields (Table III). The heavy metal came from the canal and was accompanied by the settling of wastewater in Kuan-Tu wetlands. The Ni and Zn contents in areas of short marshes exceeded those in rice paddy fields, because of good bioaccumulation in the former areas. The content of Cu was lowest in short marshes and in rice paddy fields. The capacity of rice paddy fields to accumulate Cu possibly exceeded that of short marshes.

Error-count-estimate (ECS) was used to evaluate the predictive capability of the discriminant model obtained using CDA. The resubstitution method was used to
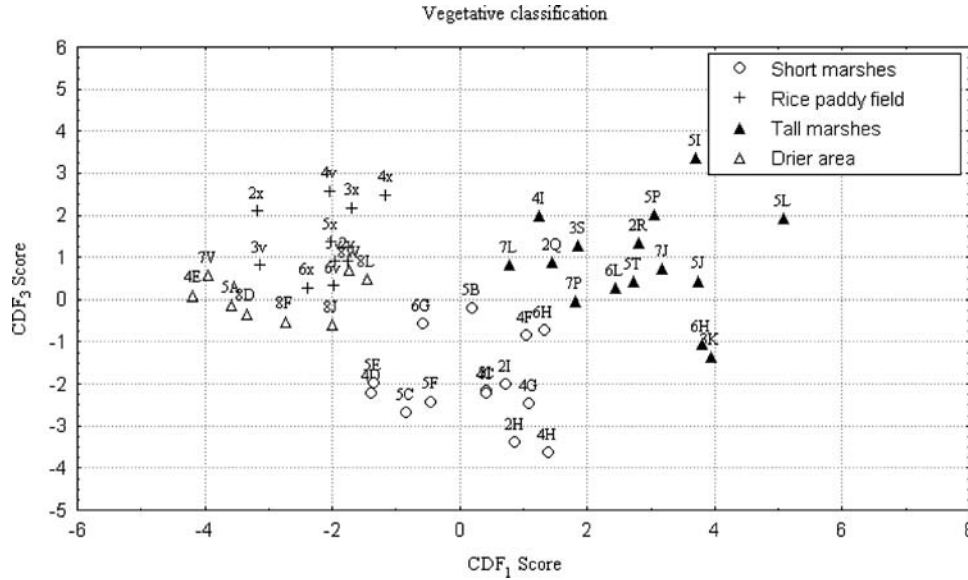
Figure 9. Scatterplot of the first two canonical discriminant functions (CDF_1 and CDF_3).

calculate ECS values and is also called internal classification analysis (Huberty, 1994). CDFs were established by 47 known vegetation soil samples and classified into four vegetation classes. Resubstitution was then used to evaluate the predictive capability of the discriminant model, with a total of 69 soil samples, 64 of which were classified correctly into pre-determined vegetation classes (results of cluster analysis coincided with results from field visual reconnaissance of the real environment), resulting in a total ECS of 8.69%. ECS in short marshes was 20.83%. Three out of a total of 24 samples were classified incorrectly into rice paddy fields and two were classified incorrectly into a drier area. Secondly, ECS in tall marshes was 6.25%. One sample out of 16 was classified incorrectly into short marshes (Table VI). The best predictor capability of the short marsh area was the CDF_3

TABLE VI
Predictive capability of the discriminant model according to error-count-estimate (ECS)

Vegetation classes	Sample number in different vegetation classes	Different vegetation classes classified by discriminant model				ECS (%)
		S	R	T	G	
Short marshes (S)	24	19	3	0	2	20.83
Rice paddy fields (R)	20	0	20	0	0	0
Tall marshes (T)	16	1	0	15	0	6.25
Drier area (D)	9	0	0	0	9	0
Total	69	20	23	15	11	8.69

model, and organic matter factor was the most affected factor. The higher existence of biomass content in rice paddy fields and in part of the drier area was the reason why higher ECS in the area of short marshes.

Different habitats attract different species of birds to feed, breed, and rest. According to a report by TWTCG (1998–2000), the number of winter migrant birds and river birds at Kuan-Tu Natural Park declined every year between 1998 and 2000. Drier habitats enabled the number of terrestrial bird species to increase slowly. The majority of family $Anatidae$ briefly gathered in Kuan-Tu Natural Park during February 2000 when farmers did spring tillage they turned over soil and drew off water from rice paddy fields. Many families $Anatidae$ were attracted and fed in a few days. However, they flew away a few days after the water level in the rice paddy fields declined. CDF_2 could discriminate between rice paddy fields and drier area. However, contents of biomass, Fe, and Mn were the most important discriminant variables. Biomass, including organisms, rice paddies, and aquatic insects aggregated under root of rice paddies were the primary food sources for family $Anatidae$. Mn can impact a bird's life by affecting plant succession and by bioaccumulation of toxins. Therefore controlling plant succession and maintaining the tillage of rice paddies to ensure a stable supply of food and mitigation of contaminants are the most important aspects of managing Kuan-Tu wetland for family $Anatidae$ in the winter. In the drier area, a source of Mn can injure birds via the food chain. A good habitat for family $Anatidae$ in the winter can be constructed by controlling these discriminant variables. Understanding the discriminatory structure can help in managing wetlands.

4. Conclusion

Multivariate data analysis was performed via factor analysis and discriminnat analysis to characterize the physical and chemical properties of soil and it can jointly differentiate among vegetation classes in Kuan-Tu wetlands, Taiwan. The physical and chemical properties of the soil were strongly related to the distribution of vegetation. Analysis of the samples shows that physical and chemical properties of the soil at the sites can be explained by three latent factors – the heavy metal factor, the salinity factor, and the organic matter factor. Furthermore figure of factor score shows the main pollutant sources – domestic wastewater, industrial wastewater, and agricultural wastewater and also indicates level of contaminants clearly.

CDA summarized important differences between predetermined groups by reducing the dimensionality of the datasets, while recognizing the complicated relationships among many characteristics. Such insight cannot be gained from univariate statistical techniques. Sediment from tall marshes site accumulated more heavy metal, except manganese, than sediment from short marshes site. Knowledge of CDA clarifies the variables to which different vegetation types are most sensitive. Controlling these discriminant variables enables the plant succession to be

manipulated and supported the construction of a biodiversity habitat. The methodologies and results should help the government to refine the current monitoring programme by selecting determinants of physical and chemical analyses of marine sediments, which may be applicable to other wetlands with comparable properties or experiencing similar environmental problems.

Acknowledgments

References

Aber, J. D., Pastor, J. and Melillo, J. M.: 1982, 'Changes in forest canopy structure along a site quality gradient in southern Wisconsin', *Am. Midland Naturalist* **108**, 256–265.

Amiaud, B., Bouzillé, J. B., Tournade, F. and Bonis, A.: 1998, 'Spatial patterns of soil salinities in old western France', *Wetlands* **18**, 482–494.

ASTM: 1994, *Particle Size Analysis of Soils*, D-422, Annual Book of ASTM Standards, Vol. 4.08, American Society for Testing and Materials, Philadelphia.

Borovec, Z.: 1996, 'Evaluation of the concentrations of trace elements in stream sediments by factor and cluster analysis and the sequential extraction procedure', *Sci. Total Environ.* **177**, 237–250.

Brower, J. E., Zar, J. H. and Von Ende, C. N.: 1998, *Field and Laboratory Methods for General Ecology*, 4th ed., McGraw-Hill, New York.

Chang, W. L. and Tang, J. L.: 1996, 'Application of multivariate statistical method to analyze irrigation water pollution in suburb Taipei', *J. Taiwan Water Conserv.* **44**, 54–66 (in Chinese).

Cruze-Castillo, J. G., Ganeshanandam, S., Mackay, B. R., Lawes, G. S. and Woolley, D. J.: 1994, 'Application of canonical discriminant analysis in horticultural research', *HortScience* **29**, 1115–1119.

Freedman, B.: 1989, *Environmental Ecology – The Impacts of Pollution and Other Stresses on Ecosystem Structure and Function*, Academic Press, New York.

Gupta, M. K., Singh, V., Rajwanshi, P., Agarwal, M., Rai, K., Srivastava, S., Shrivastav, R. and Dass, S.: 1999, 'Groundwater quality assessment of Tehsil Kheragarh, Agra (India) with special reference to fluoride', *Environ. Monit. Assess.* **59**, 275–285.

Huberty, C. J.: 1994, *Applied Discriminant Analysis*, John Wiley and Sons, New York.

Jenerette, G. D., Lee, J., D. Waller, W. and Carlson, R. E.: 2002, 'Multivariate analysis of the ecoregion delineation for aquatic systems', *Environ. Manage.* **29**, 67–75.

Johnson, R. A. and Wichern, D. W.: 1992, *Applied Multivariate Statistical Analysis*, 3rd ed., Prentice-Hall, New Jersey, Englewood Cliffs.

Luckeydoo, L. M., Fausey, N. R., Brown, L. C. and Davis, C. B.: 2002, 'Early development of vascular vegetation of constructed wetlands in northwest Ohio receiving agricultural waters', *Agric. Ecosyst. Environ.* **88**, 89–94.

Matthew, C., Lawoko, C. R. O., Korte, C. J. and Smith, D.: 1994, 'Application of canonical discriminant analysis, principal component analysis, and canonical correlation analysis as tools for evaluation differences in pasture botanical composition', *New Zeal. J. Agric. Res.* **37**, 509–520.

Mitsch, W. J. and Gosselink, J. G.: 1993, *Wetlands*, ed., Van Nostrand Reinhold, New York.

Mitchell, R. J., Marrs, R. H., Le Le Duc, M. G. and Auld, M. H. D.: 1997, 'A study of succession on lowland heaths in Dorset, southern England: Changes in vegetation and soil chemical properties', *J. Appl. Ecol.* **34**, 1426–1444.

Momen, B. and Zehr, J. P.: 1998, 'Watershed classification by discriminant analyses of lakewater-chemistry and terrestrial characteristics', *Ecol. Appl.* **8**, 497–507.

Ogutu, Z. A.: 1999, 'An investigation of the influence of human disturbance on selected soil nutrients in Narok District, Kenya', *Environ. Monit. Assess.* **58**, 39–60.

Perry, D.: 1994, *Forest Ecosystems*, Johns Hopkins University Press, Baltimore, MD, 649 pp.

Pierzynski, G. M., Sims, J. T. and Vance, G. G.: 1994, *Soils and Environmental Quality*, CRC Press, Boca Raton, Florida.

Rencher, A. C.: 1992, 'Interpretation of canonical discriminant functions, canonical varieties, and principal components', *Am. Stat.* **46**, 217–225.

Schwarz, W. L., Malanson, G. P. and Weirich, F. H.: 1996, 'Effect of landscape position on the sediment chemistry of abandoned channel wetland', *Landscape Ecol.* **11**, 27–38.

Sharma, S.: 1996, *Applied Multivariate Techniques*, John Wiley and Sons, New York.

Shin, P. K. S. and Fong, K. Y. S.: 1999, 'Multiple discriminant analysis of marine sediment data', *Mar. Pollut. Bull.* **39**, 285–294.

Soares, H. M. V. M., Boaventura, R. A. R., Machado, A. A. S. C., Esteves, D. A. and Silva, J. C. G.: 1999, 'Sediments as monitors of heavy metal contamination in the Ave river basin (Portugal): Multivariate analysis of data', *Environ. Pollut.* **105**, 311–323.

StatSoft: 1996, *STATISTICA for Windows*, StatSoft, Tulsa, Oklahoma, U.S.A.

Thomas, J. S., Lisa, D. S. and Brian, V. H.: 1999, 'How grazing and soil quality affect native and exotic plant diversity in Rocky Mountain grasslands', *Ecol. Appl.* **9**, 45–64.

TWTCG: 1998–2000, *Environmental Monitoring and Research at Kuan-Tu Natural Park and Kuan-Tu Natural Reservation Park*, Taipei City Government Print (in Chinese), Taiwan.