

Paddy Soils in Tropical Asia

Part 3. Correlation and Regression Analyses of the Soil Data

Keizaburo KAWAGUCHI* and Kazutake KYUMA**

In the preceding papers of this series a number of fertility and material characters of paddy soils in Tropical Asia were described as individual items. In the discussions significant correlations were found among many of these characters. This shows that itemwise descriptions alone are inadequate. Therefore, in the present paper correlation and regression analyses of the previously described data are presented as a preliminary step towards material classification and fertility evaluation of the sample soils.

Data and Methods

The data obtained in the routine chemical analysis, mechanical analysis, clay mineralogical analysis, and total chemical analysis conducted on plow layer samples of 410 tropical Asian paddy soils were used in this study. The origin and description of these samples and the methods of analyses were briefly given in the preceding papers.^{1,2)}

The Pearson's product moment correlation coefficients between all pairs of these variables (or characters) were computed. In this computation observed absolute values of correlation coefficient larger than 0.127 may be considered significant at $\alpha=0.01$.

Multiple regressions of cation exchange capacity (CEC) and some other variables upon their relevant characters were studied by assuming a linear regression model. The resulting multiple correlation coefficients were sufficiently high only in case of CEC to make the use of the regression equation practically meaningful.

For assessing contributions of climate, relief, and texture to accumulation of organic matter in the soil, Hayashi's theory of quantification No. 1³⁾ was adopted. The method aims at estimating a criterion variable, which is a continuous quantitative variable, such as total carbon content in the soil, from p items of qualitative variables, such as climate and relief, each of which has $k_j (j=1,2,\dots,p)$ categories. The structure of the data is shown in Table 1, in which the check mark indicates the response of a sample to that particular category under

* 川口桂三郎, Faculty of Agriculture, Kyoto University

** 久馬一剛, The Center for Southeast Asian Studies, Kyoto University

Table 1 Schematic Data Matrix used for Hayashi's Theory of Quantification No. 1

Item	1	2	P
Category	$c_{11} c_{12} \dots c_{1k_1}$	$c_{21} \dots c_{2k_2}$	$c_{p1} \dots c_{pk_p}$
Criterion				
A_1	√	√		√
A_2	⋮	√	⋮	√
A_i	⋮	⋮	√	√
A_n	⋮	√	⋮	√

each variable. Suppose that each of the categories c_{jk} is quantified as x_{jk} , and $\delta_i(jk)$ is defined as follows:

$$\begin{cases} \delta_i(jk)=1 & \text{when } i\text{th sample shows response to } k\text{th category of } j\text{th item,} \\ \delta_i(jk)=0 & \text{when it does not.} \end{cases}$$

The score to be given to the i th sample a_i , as an estimate of the measured value A_i , is computed by the addition of the relevant x_{jk} , as follows:

$$a_i = \sum_{j=1}^p \sum_{k=1}^{k_j} \delta_i(jk) \cdot x_{jk}$$

If the x_{jk} values are given to c_{jk} so as to maximize the correlation coefficient between A and a , the a_i computed by the above equation may be regarded as the best estimate of A_i from a set of p qualitative variables that are assigned to the i th sample.

All the above computations were done with the programs contained in SPSS (Statistical Package for Social Sciences⁴⁾) of the Data Processing Center of Kyoto University.

Results and Discussions

I. Correlation Analysis

The correlation coefficients between all pairs of 29 variables (or characters) are given in Table 2 in a matrix form. The variables may be grouped into 6, i.e., the ones related, respectively, to base status, mechanical composition, clay mineralogical composition, organic matter status, phosphorus status, and total chemical composition. In order to facilitate distinguishing different degrees of correlation, the table is transformed into a figure by dividing the range of correlation coefficients into 5 grades each designated by a specific pattern as shown in the legend of Fig. 1(A). In the figure appear many areas delineated by thick lines, which are numbered as in Fig. 1(B). The discussions are given hereafter by referring to the numbers.

1. Mutual correlations among base status characters

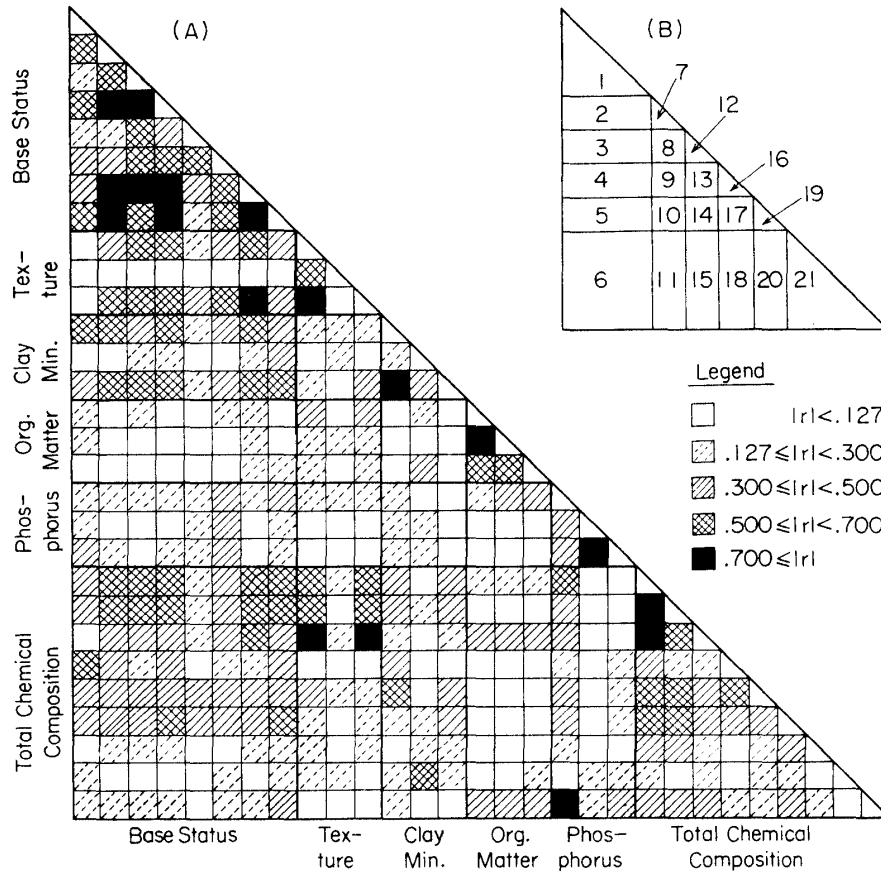


Fig. 1 Patternized Expression of the Correlation Matrix given in Table 2

The correlation coefficients in this area are generally high and positive. Ex-(Ca+Mg) is very highly correlated with Ex-Ca, CEC, Ex-Mg, and Available Silica. It is also correlated highly with pH and Ex-K, and moderately with Ex-Na. In addition to Ex-(Ca+Mg), CEC is highly correlated with, Ex-Ca, Ex-Mg, and Available Silica, and moderately with Ex-K. Correlation between Available Silica and Ex-Ca is also very high.

The lowest correlations in this area are seen between Ex-Ca and Ex-Na, pH and Ex-Mg, Available Silica and Ex-Na, and pH and Ex-Na.

2. Correlation between textural composition and base status

Except for pH, characters related to base status are moderately to highly correlated with Sand and Clay. Silt has no significant correlation with all the base status characters. Characteristically, Sand shows negative and Clay shows positive correlations with the base status characters.

Although the sign is opposite, Clay and Sand show the highest correlations with CEC, followed by the ones with Ex-(Ca+Mg), Ex-Mg, Ex-Ca, and Ex-K. Available Silica shows a moderate correlation and Ex-Na the lowest.

3. Correlation between clay mineral composition and base status

Except for Ex-Na, characters related to base status are again moderately to highly cor-

related with 7 Å and 14 Å minerals in the clay fraction. The 10 Å minerals, however, show low correlations with all the characters, the highest being the one with Available Silica.

4. Correlation between organic matter status and base status

TC, TN, and $\text{NH}_4\text{-N}$ show only low to insignificant correlations with base status characters. The pH shows negative correlation with organic matter, presumably because of the climatic influence. As will be seen later in the section III, the drier the climate, the lower the organic matter content and the higher the pH. CEC shows positive correlations probably through the effect of texture (see also Section III).

5. Correlation between phosphorus status and base status

Generally, phosphorus status shows low to insignificant positive correlations with base status characters. Only Ex-K shows moderate correlations with the three phosphorus characters. This observation is difficult to explain.

6. Correlation between total chemical composition and base status

Base status characters are highly correlated negatively with TSIO and positively with TFEO. This implies that the stronger the weathering and leaching, the more the silica and the less the iron are contained. To some extent, the latter relation between weathering and iron content contradicts with our experience that in the case of extreme weathering iron is residually accumulated in the form of pisoliths and lateritic nodules. But the correlation we have seen here deals only with paddy soils, the main occurrence of which is in recent alluvial sediments.

High correlations are also found between TALO and CEC, TCAO and pH, TMNO and Ex-(Ca+Mg), and TMNO and Available Silica. On the contrary, TKAO and TTIO show insignificant to low correlations with all the base status characters. TPHO behaves similarly to the above two, except in the case of Available Silica.

Although TMGO and TMNO are moderately correlated with base status characters, TCAO shows relatively low correlations. This may be due to the occurrence of soils containing a high amount of calcium in the sand fraction.

It is interesting to note that total silica and available silica are highly negatively correlated, and total potash and exchangeable potash are not significantly correlated.

7. Mutual correlations among textural separates

A very high negative correlation is found between Sand and Clay. Silt is moderately negatively correlated with Sand, but its correlation with Clay is insignificant.

8. Correlation between clay mineral composition and textural composition

Correlation coefficients in this area are all low. The 7 Å minerals are correlated positively with Sand and negatively with Clay. The 10 Å minerals show a positive correlation with Clay, which is the highest in this area, and a negative correlation with Sand. Silt shows a positive correlation only with 10 Å minerals.

9. Correlation between organic matter status and textural composition

Again all the correlations are low. The fact that Sand is negatively and Clay is positively

correlated with all the three characters, however, indicates a certain control exerted by texture on the accumulation of organic matter, especially of TC. This point will be further elaborated in Section III.

10. Correlation between phosphorus status and textural composition

All the correlations are low to insignificant; especially those of available phosphorus (Bray-P and HCl-P) with texture are insignificant.

11. Correlation between total chemical composition and textural composition

Sand and Clay are highly correlated with TSIO, TFEO, and TALO. TSIO shows a positive correlation with Sand and a negative correlation with Clay. The reverse is true for TFEO and TALO. The rest of the correlations are low to insignificant, except a moderate negative correlation between TMGO and Sand.

12. Mutual correlations among clay mineral species

The correlations are all negative. A very high negative correlation between 7 Å and 14 Å minerals is noteworthy.

13. Correlation between organic matter status and clay mineral composition

In this area only the 10 Å minerals show low negative correlations with TC and NH₄-N.

14. Correlation between phosphorus status and clay mineral composition

Low negative correlations are seen between 7 Å minerals and total as well as available phosphorus, and low positive correlations between 10 Å minerals and available phosphorus.

15. Correlation between total chemical composition and clay mineral composition

Many of the correlations are low to insignificant. A high positive correlation between TKAO and 10 Å minerals is noteworthy. The 10 Å minerals contain about 10% of potash in their interlayer positions. Another high, but negative, correlation is seen between 7 Å minerals and TMGO. Unlike many of the 14 Å minerals, 7 Å minerals are devoid of Mg in their octahedral layers, thus high 7 Å minerals tend to accompany a low TMGO content.

16. Mutual correlations among the characters related to organic matter status

A very high positive correlation is seen between TC and TN, which is actually the highest of all the correlations. As stated in a preceding paper, C/N ratio of the soil converges to a narrow range (10–12), which suggests the high correlation. NH₄-N is also highly correlated with TC and TN.

17. Correlation between phosphorus status and organic matter status

Total phosphorus content is moderately correlated with all TC, TN, and NH₄-N, while available phosphorus contents are not.

18. Correlation between total chemical composition and organic matter status

Again, TPHO is moderately correlated with organic matter. Besides, TALO is positively and TSIO is negatively correlated with organic matter, which may be explained in terms of texture. Others are insignificant, except for a low negative correlation between TKAO and NH₄-N.

19. Mutual correlations among phosphorus status characters

Available phosphorus contents determined by two different methods (Bray and HCl) are very highly correlated, while correlations between TP and available phosphorus are moderate. It appears that the more drastic the method of available phosphorus determination, the higher the correlation.

20. Correlation between total chemical composition and phosphorus status

Naturally, total phosphorus contents determined with chemical method and X-ray fluorescence method are highly correlated. But it is not as high as expected because of the discrepancy of the two data from the causes stated earlier.²⁾

Except for TKAO, all the elemental oxides are more or less correlated with TP. The highest correlation is seen with TSIO. The other elements show low to moderate positive correlations.

Available phosphorus is mostly not correlated with elemental oxides, but here again the one determined with a more drastic method (HCl-P) tends to be relatively more highly correlated.

21. Mutual correlations among total contents of elemental oxides

Very high correlations exist between TSIO and TFEO and between TSIO and TALO. TSIO shows high negative correlations also with TMGO and TMNO, and moderate negative correlations with TPHO, TCAO and TTIO. A low negative correlation is seen between TSIO and TKAO.

TFEO behaves similarly to TSIO but correlations are all positive. TALO has a high correlation with TFEO, but shows generally lower correlations with other elements as compared to TFEO. TCAO has a high correlation with TMGO, but its correlations with other elements are not high. TMGO and TMNO behaves similarly, showing high correlations with TSIO and TFEO. The highest correlation of TTIO is seen with TFEO. TKAO shows low to insignificant correlations with other elements. As a matter of course, TPHO behaves similarly to TP.

From the above discussions on the correlation matrix, some important general remarks may be made:

- a) Base status characters are highly correlated not only among themselves but also with the textural composition, clay mineralogy, and a part of the total chemical composition. In this relation, Silt, 10 Å minerals, TTIO and TKAO are exceptional, showing only low to insignificant correlations with base status characters.
- b) Among the mutually correlated characters referred to under a), Sand content in the soil, 7 Å mineral content in the clay fraction, and TSIO content in the soil are negatively correlated with most of the other characters, though their mutual correlations are positive.
- c) Characters representing organic matter are not highly correlated with any of the character groups, though their mutual correlations are high.
- d) The same could be said of the characters related to available phosphorus status.

II. Regression Analysis

As stated above, CEC and Clay are highly correlated, $r=0.791$. Regression of CEC upon Clay can be formulated as follows:

$$\text{CEC} = 0.44 \text{ Clay} + 1.72 \quad (1)$$

The equation (1) indicates that every 1% of clay contributes to CEC by 0.44 me, or 100 g clay has, in average, 44 me of CEC. The coefficient of determination for this regression is $r^2=0.63$.

To further elaborate CEC-Clay relationship, multiple regression of CEC on different clay species was studied. The content of each clay species in the soil was computed from total clay content and relative proportions of different clay species. Thus 7 Å mineral content (CL7), 10 Å mineral content (CL10), and 14 Å mineral content (CL14) of each soil were obtained. Correlation coefficients among CEC, CL7, CL10, and CL14 for 410 sample soils were computed as follows:

	CEC	CL7	CL10	CL14
CEC				
CL7	.150			
CL10	.067	.164		
CL14	.915	-.011	-.061	

The correlation matrix indicates that CEC is correlated very highly with CL14, but only lowly with CL7 and insignificantly with CL10. The mutual correlations among the three clay contents are low to insignificant. Therefore, multicollinearity is not a problem in this multiple regression analysis.

The following 3 regression equations were obtained stepwise:

$$\text{CEC} = 7.15 + 0.68^{**}\text{CL14} \quad (\text{R}=0.915) \quad (2)$$

$$\text{CEC} = 4.66 + 0.68^{**}\text{CL14} + 0.15^{**}\text{CL7} \quad (\text{R}=0.929) \quad (3)$$

$$\text{CEC} = 3.83 + 0.69^{**}\text{CL14} + 0.13^{**}\text{CL7} + 0.20^{**}\text{CL10} \quad (\text{R}=0.934) \quad (4)$$

**Significant at $\alpha=0.01$

The partial regression coefficients for all the three clay species are highly significant, though the increase in the coefficient of determination is not great in the 2nd and 3rd steps.

The contribution of organic matter (in terms of TC) to CEC was checked by putting TC as the 4th independent variable. The following equation was obtained:

$$\text{CEC} = 3.58 + 0.68^{**}\text{CL14} + 0.12^{**}\text{CL7} + 0.20^{**}\text{CL10} + 0.34 \text{ TC} \quad (\text{R}=0.935) \quad (5)$$

The partial regression coefficient for TC in this equation is not significant even at $\alpha=0.05$, indicating that organic matter has no statistically significant contribution to CEC. We do not know at this moment whether this result is generally applicable to tropical soils or it is only applicable to tropical paddy soils.

Each of the partial regression coefficients in the equation (4) has the following confidence limits at $\alpha=0.05$;

$$0.66 \leq b_1 \leq 0.71 \quad \text{for CL14}$$

$$0.10 \leq b_2 \leq 0.17 \quad \text{for CL7}$$

$$0.13 \leq b_3 \leq 0.27 \quad \text{for CL10}$$

The ranges appear reasonable in comparison with the generally accepted values of CEC for various clay mineral species. The 14 Å minerals consist of montmorillonite, vermiculite, and Al-interlayered vermiculite-chlorite intergrades, of which montmorillonite is dominant. The 10 Å minerals are mostly illite and clay size micas. The 7 Å minerals includes kaolinite and meta-halloysite.

Other trials of multiple regression analyses taking Available Silica, Ex-K, and NH₄-N as dependent variables were not successful, as judged from their low coefficients of determination.

III. Effect of Climate, Relief, and Soil Texture on Organic Matter Status

In the preceding papers of the series it was noted that the content of organic matter in the soil is controlled to some extent by such factors as climate, local relief or swampiness of the terrain, and soil texture. In order to ascertain this aspect of correlation in a quantitative manner we introduced Hayashi's theory of quantification No. 1.

Climatic regions were earlier established by ourselves (Kyuma⁵) for the entire region of tropical Asia. Referring to the sampling location on the map of climatic regions, a climatic class code is given to each sample. The samples from the Philippines were not included in this study because of the complexity of the climatic regions in the archipelago. Altogether 356 samples from 6 climatic regions, I through V and VII, were used in this study. They bear the climatic class code that is equal to the region number.

Local relief factor has two codes, 1 and 2; the code 1 is for ordinary non-swampy terrain and the code 2 is for permanently swampy terrain. The codes were given to the samples based on our field observations.

As a preparatory step of analysis mean TC, TN and NH₄-N contents were computed for each of the combinations between climatic class and local relief class, as seen in Table 3. The numbers of samples for climatic classes 3 and 4 are small and no samples occur in swampy terrain in these two climatic classes. Apparently fewer samples occur in swampy conditions in the alternately wet and dry climate (class 4,5, and 7) than in the permanently humid climate (class 1,2, and 3).

From Table 3 it is clear that the level of organic matter accumulation is higher under the wetter climate and that the relief affects the organic matter reserve conspicuously.

Effect of soil texture is shown in Table 4. In the table textural classes are coded as follows: 1-HC, 2-LiC, 3-SiC, 4-SC, 5-CL, 6-SiCL, 7-SCL, 8-L & SiL, 9-SL, and 10-LS & S. The TC, TN and NH₄-N contents are generally higher as soil texture becomes finer. In an Indonesian Ando soil having clay loam texture (code 5) an extraordinarily high amount of organic matter is contained, i.e., 5.6% TC, 0.59% TN, and 36.2 mg/100 g of NH₄-N.

Table 3 Organic Matter Status as Broken Down by Climate and Relief

Climate	Relief	No. of Samples	TC, %	TN, %	NH ₄ -N, mg/100g
1	1	39	1.29	0.10	11.7
	2	11	5.50	0.47	26.0
	Mean	50	2.22	0.18	14.9
2	1	28	1.42	0.13	11.3
	2	14	3.45	0.28	15.4
	Mean	42	2.10	0.18	12.6
3	1	7	2.21	0.22	20.6
	2	0	—	—	—
	Mean	7	2.21	0.22	20.6
4	1	10	1.23	0.12	3.7
	2	0	—	—	—
	Mean	10	1.23	0.12	3.7
5	1	151	1.05	0.09	4.3
	2	6	2.97	0.24	6.4
	Mean	157	1.12	0.10	4.4
7	1	83	0.86	0.09	3.8
	2	7	2.34	0.22	11.8
	Mean	90	0.98	0.10	4.4

Table 4 Organic Matter Status as Broken Down by Texture

Texture	No. of Samples	TC, %	TN, %	NH ₄ -N, mg/100g
1	137	1.93	0.16	8.90
2	63	1.24	0.11	8.23
3	20	1.40	0.14	6.77
4	14	1.15	0.10	4.94
5	35	1.32 (1.20)	0.13 (0.11)	7.01 (6.15)
6	13	0.99	0.11	5.62
7	18	0.89	0.08	4.46
8	13	0.70	0.07	3.62
9	28	0.44	0.05	3.07
10	15	0.58	0.06	4.60

The mean values of the class are considerably lowered when this soil is excluded, as shown in parentheses.

As the first step, all of these three factors, i.e., climate, relief, and texture, were taken into consideration in explaining the organic matter status. The overall mean values of the criterion variables are as follows for 356 samples (excluding Philippine soils):

	mean	standard deviation
TC (%)	1.38	1.34
TN (%)	0.124	0.112
NH ₄ -N (mg/100 g)	6.15	7.03

The seven climatic divisions were reduced to four by combining those of geographical proximity. This way the number of samples per category was increased. Likewise, texture is recorded into 4,1-HC, 2-LiC, SiC, and SC, 3-CL, SiCL, and SCL, and 4-L, SiL, SL, LS, and S (see Table 5). The model, in case of TC, is as follows:

$$\text{TC (\%)} = \text{climate (4 categ.)} + \text{relief (2 categ.)} + \text{texture (4 categ.)}$$

Numeric values assigned to each category of the three items (or variables) are as shown in Table 5. Multiple correlation coefficients (R) and partial correlation coefficients for each variable in the above model are given in Table 6.

The R values are not sufficiently high to justify the use of the equations obtained for prediction. The partial correlation coefficients indicate that the relief factor is most relevant to the TC and TN contents, while climate contributes most to the amount of mineralizable nitrogen. The effect of climate on TC and TN seems to be relatively minor.

Table 5 Numerical Values Assigned to Each Category of Three Qualitative Variables to Estimate Organic Matter Status

Variable	Category	TC	TN	NH ₄ -N
Climate	1+3	0.453	0.033	7.481
	2	0.316	0.025	4.164
	4+5	-0.044	-0.008	-2.151
	7	-0.352	-0.018	-2.690
Relief	1	-0.251	-0.020	-0.746
	2	2.099	0.171	6.240
Texture	1	0.328	0.024	0.453
	2	-0.089	-0.006	0.401
	3	0.081	0.012	0.613
	4	-0.745	-0.062	-2.524

Table 6 Partial and Multiple Correlation Coefficient between Organic Matter Status and Climate, Relief, and Texture

Partial Corr.	TC	TN	NH ₄ -N
Climate	0.274	0.215	0.611
Relief	0.596	0.568	0.395
Texture	0.356	0.333	0.219
Multiple Corr.	0.706	0.674	0.725

Table 7 Partial and Multiple Correlation Coefficients between Organic Matter Status and Relief and Texture after Stratifying the Samples by Climate

Permanently Humid Climate: Number of samples=99			
Partial Corr.	TC	TN	NH ₄ -N
Relief	0.648	0.625	0.410
Texture	0.393	0.354	0.367
Multiple Corr.	0.690	0.661	0.509
Alternately Wet and Dry Climate: Number of Samples=257			
Partial Corr.	TC	TN	NH ₄ -N
Relief	0.499	0.478	0.321
Texture	0.522	0.479	0.238
Multiple Corr.	0.674	0.637	0.407

Soil texture is of moderate importance for determining the TC and TN levels, but of minor importance for NH₄-N.

As it was anticipated from the field observations that texture is more important in determining the organic matter level in the drier climate, similar analysis was done after stratifying the samples into two, one being the samples occurring in permanently humid climate (codes 1, 2, and 3) and the other being those occurring in wet and dry climate (codes 4, 5, and 7). The results are shown in Table 7 in terms of partial and multiple correlation coefficients.

The multiple correlation coefficients are low for NH₄-N regardless of climatic zones. Relief factor is even more relevant to TC and TN in the permanently humid climate, while texture plays as important a role as relief for TC and TN in the wet and dry climate.

Summary

Correlation coefficients between all pairs of 29 variables that had been described previously in relation to fertility and material characteristics were computed for the same 410 tropical Asian paddy soil samples.

Chemical characters related to base status are highly correlated not only with each other but also with such character groups as textural composition, clay mineralogy, and total chemical composition.

Characters related to organic matter status are not highly correlated with any of the other character groups, though their mutual correlations are high. The same holds for characters related to phosphorus status.

Multiple regression of CEC on the contents of different clay mineral species and organic matter was studied. The 14 Å minerals alone explained more than 80% of the variance of CEC. The 7 Å and 10 Å minerals also contributed significantly, but the contribution of organic matter to CEC turned out to be insignificant.

Hayashi's theory of quantification (No. 1) was adopted to assess the contributions of such qualitative variables as climate, relief, and texture to organic matter status of the soil. Climate was proved to be most relevant to the reserve of mineralizable nitrogen, while local relief contributed most to TC and TN. Effect of texture on accumulation of organic matter was clearer under a drier climatic condition.

References

- 1) Kawaguchi, K. and Kyuma, K. 1974. "Paddy Soils in Tropical Asia. Part 1, Description of Fertility Characteristics," *Tonan Ajia Kenkyu*, Vol. 12, No. 1, pp. 3-24.
- 2) Kawaguchi, K. and Kyuma, K. 1974. "Paddy Soils in Tropical Asia. Part 2, Description of Material Characteristics," *Tonan Ajia Kenkyu*, Vol. 12, No. 2, pp. 177-192.
- 3) Hayashi, C. *et al.* 1970. *Joho-shori to Tokei-suri* (Information processing and Statistical Mathematics), Sangyo-tosho, Tokyo.
- 4) Miyake, I. 1973. *Shakaikagaku no tameno Tokei-pakkeiji* (Statistical Package for Social Sciences), Toyoshinpo-sha, Tokyo.
- 5) Kyuma, K. 1972. "Numerical Classification of the Climate of South and Southeast Asia," *Tonan Ajia Kenkyu*, Vol. 9, No. 4, pp. 502-521.