CLASSIFICATION TM (THEMATIC MAPPER) DATA BY USING STATISTICAL ANALYSIS AND ARTIFICIAL NEURAL NETWORK

Nikom Suvonvorn
Department of Computer Engineering
Prince of Songkla University

1. INTRODUCTION

The satellite imagery plays more and more important role on land classification and planing. Since these images can obtain as frequency as possible, the use of appropriate method to analyze and obtain the precise data should be considered.

All the information spread around a digital image are known as group of pixel, the spectral reflectance of each pixel in specific frequency (BAND) contains a part of data. The use of multiple frequency data to identify or classify land use pattern is challenged.

Recently, most of scientists applied statistical analysis and artificial intelligence to classify land use pattern, which give a better result. However, the precision of analysis still doubtable.

In general, the data obtained from satellite imagery for land use classification are BAND #1 to BAND #5 of spectral reflectance of Landset TM. The combination of pixel data of this spectral reflectance will be able to identify land use pattern. By this manner, the interactions between each pixel data of spectral reflectance will show an important information. This information will convey to artificial neural network for training of land use pattern identification.

2. OBJECTIVES

The main objective of this study is to apply a statistical method to obtain an important information as input for 4 layers artificial neural network for land use pattern classification.

3. APPROACH BACKGROUND

To be classified, we first discuss for the remote-sensing instrument (TM sensor), fundamental of factorial experiment and fundamental of artificial neural network, respectively. After this following sections the methods used for Landsat TM classification are described.

3.1 LANDSAT TM SENSOR

Landsat TM sensor consists of 7 band including visible band, infrared band and thermal infrared band. Each band has been designed for detecting the appropriate things as described in the following.

(1) *BAND #1:* 0.45 to 0.52 μm--designed to provide increased penetration into water bodies as well as supporting analysis of land use, soil and vegetation characteristics.

(2) *BAND #2:* 0.52 to 0.60 μm--primarily designed to look at the visible green reflectance peak of vegetation lying between the two-chlorophyll absorption bands. Responses in this band are

intended to emphasize vegetation discrimination and vigor assessment.

(3) *BAND #3:* 0.63 to 0.69 μm--the most important band for vegetation discrimination. It resides in one of the chlorophyll absorption regions and emphasizes contrast between vegetation and non-vegetation features as well as contrasts within vegetation classes.

(4) *BAND #4:* 0.76 to 0.90 μm--chosen to be responsive to amounts of vegetation biomass present in a scene. This will aid in crop identification and will emphasize soil-crop and land-water contrasts.

(5) *BAND #5:* 1.55 to 1.75 μm--a band known to be important to the determination of crop type, crop water content and soil moisture conditions.

(6) *BAND #6:* 10.4 to 12.5 μm--a band important in the discrimination of rock formations.

(7) *BAND #7:* 2.08 to 2.35 μm--a thermal infrared channel known to be contributory to vegetation classification, vegetation stress analysis, soil moisture discrimination and a host of other thermally related phenomena.

The resolution of BAND #1, BAND #2, BAND #3, BAND #4, BAND #5 and BAND #7 is 30x30 m but BAND #6 is 120x120 m. The smallest unit of the TM data is pixel consisted of 7 spectral values as show in figure 1.

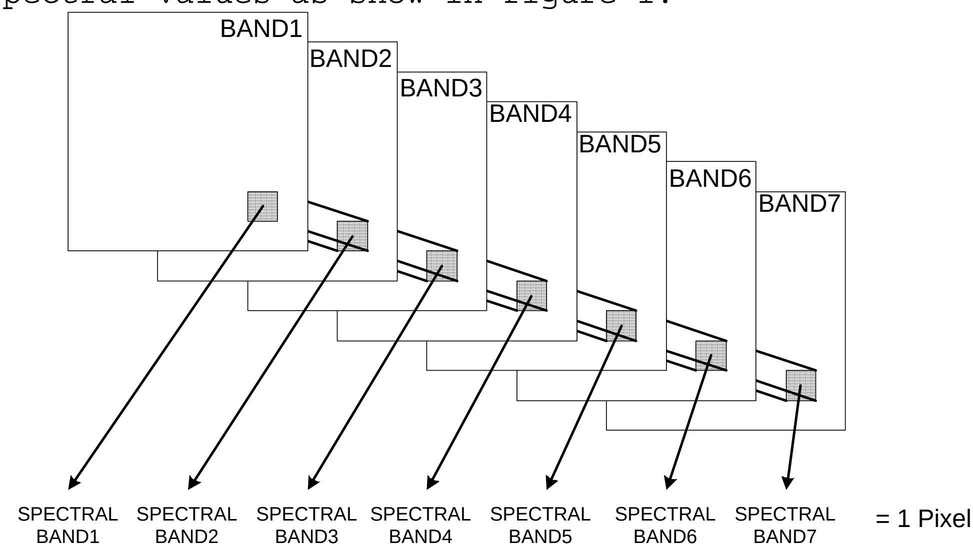

Figure 1: A pixel of Landsat TM data

3.2 FUNDAMENTAL FACTORIAL EXPERIMENT

Factorial experiment is designed to evaluate the combined effect of two or more experimental variables, permitting the evaluation of interaction effects. The interaction effect is an effect attributable to the combination of variables. At the end of experiment, the experimenter has information permitted him to make decision which have a broad range of applicability.

3.2.1 TERMINOLOGY AND NOTATION

The term *factor* will be used interchangeably with the term treatment and experimental variable. More specifically, a factor is a series of related treatment which making up a factor composes the *levels* of that factor.

The number of factors and the number of levels for each factor indicate the *dimensions* of a factorial experiment. For example, a factorial experiment in which there are two factors, one having P levels and the other having Q levels, is called PxQ factorial experiment as show in figure 2.

	b_1	b_2	$\cdots$	b_J	$\cdots$	b_Q	
a_1	μ_{11}	μ_{12}	$\cdots$	μ_{1J}	$\cdots$	μ_{1Q}	$\mu_{1.}$
a_2	μ_{21}	μ_{22}	$\cdots$	μ_{2J}	$\cdots$	μ_{2Q}	$\mu_{2.}$
.	.	.		.		.	.
a_I	μ_{I1}	μ_{I2}	$\cdots$	μ_{IJ}	$\cdots$	μ_{IQ}	$\mu_{I.}$
.	.	.		.		.	.
a_P	μ_{P1}	μ_{P2}	$\cdots$	μ_{P3}	$\cdots$	μ_{PQ}	$\mu_{P.}$
	$\mu_{.1}$	$\mu_{.2}$	$\cdots$	$\mu_{.J}$	$\cdots$	$\mu_{.Q}$	$\mu_{..}$

Figure 2: PxQ factorial experiment

The notation in figure 2

μ_{IJ} denotes the mean under treatment combination of factor A at level I and factor B at level J.

$\mu_{I.}$ denotes the mean of all treatment combination which factor A is at level I is given by

$$\mu_{I.} = \frac{\sum_J \mu_{IJ}}{Q} \qquad (1)$$

$\mu_{.J}$ denotes the mean of all treatment combination which factor B is at level J is given by

$$\mu_{.J} = \frac{\sum_I \mu_{IJ}}{P} \qquad (2)$$

$\mu_{..}$ denotes the grand mean of all treatment combination is given by

$$\mu_{..} = \frac{\sum_I \sum_J \mu_{IJ}}{PQ} = \frac{\sum_I \mu_{I.}}{P} = \frac{\sum_J \mu_{.J}}{Q}$$

3.2.2 FACTOR EFFECT

Factor effect composes of two types including main effect and interaction effect.

The main effect is an effect inside each factor, defined in terms of parameters. The estimates of these parameters will be obtainable for corresponding statistics. The main effects of factors A and B are given by

$$\alpha_I = \mu_{I.} - \mu_{..} \tag{3}$$

$$B_J = \mu_{.J} - \mu_{..} \tag{4}$$

Defined α_I is the main effect of A at level I

β_J is the main effect of B at level J

The variance of the main effects due to factor A is given by

$$\sigma^2 = \frac{\sum_I (\mu_{I.} - \mu_{..})^2}{P-1} = \frac{\sum_I \alpha_I^2}{P-1} \tag{5}$$

The interaction effects of factor A level I and factor B J is a measure of the extent to which the criterion mean for treatment combination AB_{IJ} can not be predicted from the sum of the corresponding of main effects, a measure of the non-additivity of main effects. The interaction effects of factor A and B are given by

$$\alpha\beta_{IJ} = \mu_{IJ} - (\alpha_I + \beta_J + \mu_{..}) \tag{6}$$

Defined, $\alpha\beta_{IJ}$ is an interaction effect of factor A at level I and factor B at level J. And it also follows that $\sum_I \alpha\beta_{IJ} = \sum_J \alpha\beta_{IJ} = 0$

The variance of the interaction effects due to factor A and factor B is given by

$$\sigma_{\alpha\beta}^2 = \frac{\sum_I \sum_J (\alpha\beta)_{IJ}^2}{(P-1)(Q-1)} \tag{7}$$

And the variance of experimental error due to the combination of factor A and factor B is given by

$$\sigma_\varepsilon^2 = \frac{\sum_K (X_{IJK} - \mu_{IJ})^2}{N-1} \tag{8}$$

Defined, X_{IJK} is the observing value at K Th. of the combination of factor A at level I and factor B at level J.

N are the numbers of observing value.

3.3 FUNDAMENTALS OF ARTIFICIAL NEURAL NETWORK (ANN)

Artificial neural network consist of a large number of neurons which is a simple processing units, also referred to as neuroades of human nervous system. Each neuron is connected to the others. It is capable of receiving the input, processing input that multiplying input by a corresponding weight value and transmitting the output to the other. The weight inputs are summed to determine the activation or decision level (F) of neuron. The weights represent as the connection strengths between neurons.

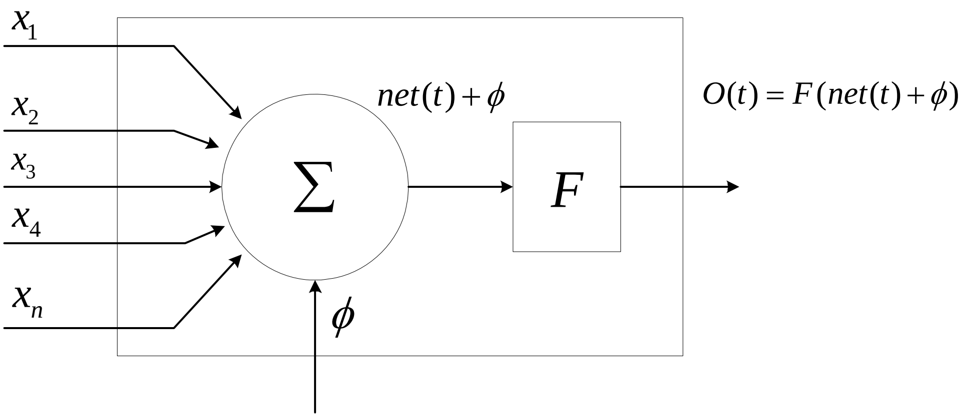

Figure 3: Artificial neural representation

In figure 3 let $x=(x_1,x_2,....,x_n)$ as the n inputs of artificial neural.

$$net(t) = \sum_{i=1}^{n} x_i w_i + \phi_i \quad (9)$$

Defined $net(t)$ denotes the sum of weight input.

ϕ denotes a bias (constant)

w denotes the weight (connection strengths)

The output of neuron is $F(net(t))$ when F is activation or decision function. The study use log-sigmoid function is given by

$$F(x)=\frac{1}{1+e^{-x}} \quad (10)$$

Therefore, the output of neuron ($O(t)$) is given by

$$O(t)=\frac{1}{1+e^{-net(t)}} \quad (11)$$

Equation (3), the output will be Non-linear. Because of this it can be solved the Non-linear problem.

Considering for artificial neural network design

(1) the number of layer

(2) The number of neurons.

4. METHODOLOGY

4.1 USING NEURAL NETWORK TO CLASSIFY TM DATA

4.1.1 NEURAL NETWORK DESIGN

By preliminary study, it shows that the interactions between BAND #1 to BAND #5 are given the results in Table 1. Since the combination of 2-3 bands effect on land use pattern identification, 4 layers neural network is considered. This neural network contains 4 layers, Input Layer, 2 Hidden Layers and Output Layer.

(1) Input Layer: The input layer are BAND #1 - BAND #5 of interested area. This information will bring to hidden layer #1.

(2) Hidden layer #1: This layer considers the interaction between input layers that will result for 4 different categories. The result obtained from this layer will then pass through the hidden layer #2.

Table 1: *The number of neurons in Hidden Layer #1*

Input Data (Interactions)	Categories			
	Plant	Water	Soil	Building
1. BAND #1 & #1		✓		✓
2. BAND #1 & #2		✓		
3. BAND #1 & #3				✓
4. BAND #1 & #4		✓		
5. BAND #1 & #5				✓
6. BAND #2 & #2	✓	✓		
7. BAND #2 & #3	✓			
8. BAND #2 & #4	✓	✓		
9. BAND #2 & #5				
10. BAND #3 & #3	✓		✓	✓
11. BAND #3 & #4	✓		✓	
12. BAND #3 & #5			✓	✓
13. BAND #4 & #4	✓	✓	✓	
14. BAND #4 & #5			✓	
15. BAND #5 & #5			✓	✓
16. BAND #1 & #2 & #3				
17. BAND #1 & #2 & #4				
18. BAND #1 & #2 & #5		✓		
19. BAND #1 & #3 & #4				
20. BAND #1 & #3 & #5				✓
21. BAND #1 & #4 & #5				
22. BAND #2 & #3 & #4	✓			
23. BAND #2 & #3 & #5				
24. BAND #2 & #4 & #5				
25. BAND #3 & #4 & #5			✓	
Number of neurons	7	7	7	7

Note: Black rows show the interactions which not used.

(3) Hidden Layer #2: The result of hidden layer #1 which contain interaction between land use pattern

will then be determined. The statistical analysis will be applied to obtain the precise output.

(4) Output Layer: This layer is the result of hidden layer #2 and some necessary information and threshold value.

4.1.2 TRAINING DATA

Back-propagation learning algorithm, which will be described later, is used to trains neural network system; to prepare the neural network for the TM data classification (supervised classification). The data for training the neural network system is training data

Training data is groups of data, which are used to calculate change in weight values of neural network system. The adjustments of weight values make the neural network system can be classified TM data to the desire categories.

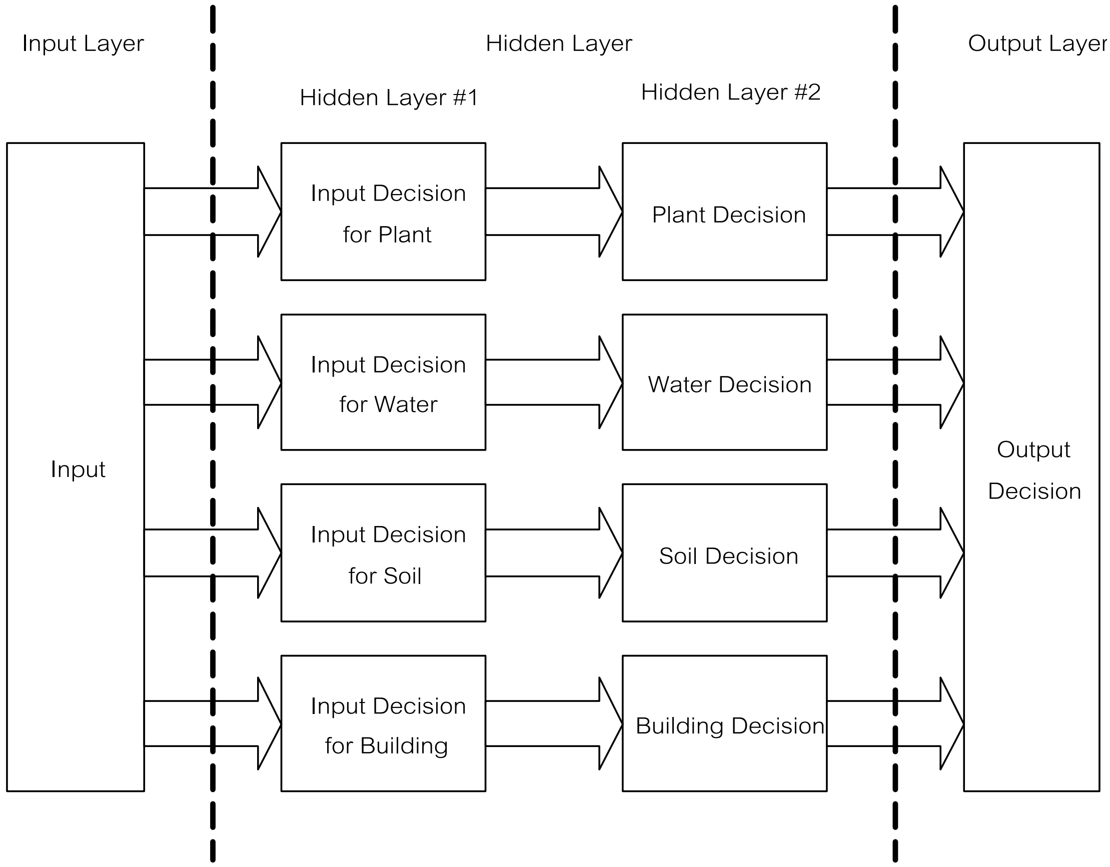

Figure 4: Block diagram of neural network design

To define training data, training data is selected from group of things that is interested, such as type of plant, water, soil and building. For each thing that is chosen, the interactions illustrated in table 6 must be obtained. The prototypes of training data including input vector and output vector illustrated in equation 12.

$$\left\{ \begin{bmatrix} Band11 \\ Band12 \\ : \\ Band345 \end{bmatrix}, \begin{bmatrix} d1 \\ d2 \\ d3 \\ d4 \end{bmatrix} \right\} \qquad (12)$$

From equation 12, show the defining of input vector and output vector of training data. Note that, $Band_{ijk}$ is the input vector (the interactions between bandwidths), d_i is output vector (desired or target output)

4.1.3 BACK-PROPAGATION LEARNING

Back-propagation learning algorithm involves the presentation of pairs of input and output vector (training data). The actual output for a given vector is compared with the desired or target output. If there is no difference, no weights are changed; otherwise, the weights are adjusted to reduce the difference.

The neuron network for a study shows in picture 12, the neurons in Input Layer ($L1$) serve only as distribution points; they perform no input summations. For each neuron in Hidden Layer #1 ($L2$), Hidden Layer #2 ($L3$) and Output Layer ($L4$) produce $net(t)$ and $O(t)$ that given by equation (9) and (11).

The back-propagation learning algorithm for TM data classification can be described in the following step:

First step: Initialize the weights

The weights between layer $L1$ and $L2$, layer $L2$ and $L3$ and Layer $L3$ and $L4$ are initialized to small random values

$$\lfloor w_{1,1} \quad w_{1,2} \quad w_{1,3} \quad . \quad . \quad . \quad w_{ij} \rfloor \qquad (13)$$

Second step: Process to obtain output

Present a continuous-valued input vectors $x = [x_1 \quad x_2 \quad x_3 \quad \ldots \quad x_{18}]^T$ and obtains output $y = [y_1 \quad y_2 \quad y_3 \quad y_4]$ at Layer $L4$. In order to obtain output vector y, calculation is done layer by layer start from layer $L2$. The $net(t)$ value of each neuron in layer $L2$ is calculated as the weighty sum of its inputs. The $net(t)$ input is then passed through the activation function F (the study use log-sigmoid function) to produce $O(t)$ for each neuron in layer $L2$. The outputs of neurons in layer $L2$ serve as inputs to neurons in layer $L3$ and the outputs of neurons in layer $L3$ serve as inputs to neurons in layer $L4$. The process is repeated to obtain the output vector y at layer $L4$ (network output).

Third step: Calculate change in weights

In order to calculate change in weights, the output vector y is compared with the desired vector or the target vector d_i and the error between the two vectors is obtained. The error is then propagated backward to obtain the change in weight Δw_{ij} that is used to update the weights. The Δw_{ij} which is the weights between layers $L3$ and $L4$ is given by

$$\Delta w_{ij} = \alpha O_j \delta_i \quad (14)$$

Defined α is training rate at layer $L4$ (typically 0.01 to 1.0)

O_j is the output of neuron j at layer $L3$

δ_i is given by

$$\begin{aligned}\delta_i &= [\frac{\partial F(net_i)}{\partial net_i}](d_i - O_i) \\ &= [\frac{e^{-net}}{(1+e^{-net})^2}](d_i - O_i) \\ &= [\left(\frac{1}{1+e^{-net}}\right)\left(\frac{e^{-net}}{1+e^{-net}}\right)](d_i - O_i) \\ &= [O_i(1-O_i)](d_i - O_i) \\ &= O_i(1-O_i)(d_i - O_i)\end{aligned} \quad (15)$$

And O_i is output of neuron i at Layer $L4$ (network output)

d_i is the target output or the desired output of neuron i at Layer $L4$

Layer $L2$ and $L3$ have no target vector, so equation 15 can not be used. The back-propagation algorithm trains the hidden layers by propagating the output error back through layer by layer in order to adjust weights at each layer. The calculate change in weights between layer $L1$-$L2$ and $L2$-$L3$ are given by

$$\Delta w_{ij} = \beta O_j \delta_{Hi} \quad (16)$$

Defined β is training rate at layer $L3$ (typically 0.01 to1.0) if the calculate change between $L2$-$L3$

O_j is the output of neuron j at layer $L2$

δ_{Hi} is given by

$$\delta_{HI} = O_i(1-O_i)\sum_k \delta_k w_{ik} \quad (17)$$

Defined O_i is the output of neuron i at Layer $L3$

Forth step: Update the weights

To update the weights is given by

$$w_{ij}(n+1) = w_{ij}(n) + \Delta w_{ij} \quad (18)$$

Defined $w_{ij}(n+1)$ are the weights after adjustment.

$w_{ij}(n)$ are the weights before adjustment.

Fifth step: Obtain the error for neurons at layer $L4$

To obtain the error ε is given by

$$\varepsilon = \sum_i (O_i - d_i)^2 \quad (19)$$

If the error ε is greater than some minimum $\varepsilon_{\min}$, then repeat steps 2 through 4, otherwise terminate the training process.

4.2 BASIC CONCEPT OF TM DATA CLASSIFICATION

Method of TM data classification, the factorial experiment analysis is use to be obtained the interaction between bandwidths (BAND #1-BAND #5). This interaction is very important to indicate that which points of data is plant, water, soil and building. For example, the bandwidths which can be indicated to plant is BAND #2 (Green), BAND #3 (Red) and BAND #4 (Near infrared). So, the interactions are:

(1) Interaction inside BAND #2
(2) Interaction inside BAND #3
(3) Interaction inside BAND #4
(4) Interaction between BAND #2 and BAND #3
(5) Interaction between BAND #2 and BAND #4
(6) Interaction between BAND #3 and BAND #4
(7) Interaction between BAND #2 and BAND #3 and BAND #4

The interactions make it easy to be isolated water from plant; plant -- if the wavelength is nearly to BAND #4, the spectral reflectance is more reflect, Water--the spectral reflectance is less than.

This study can be categorized things which are interested into four groups consist of vegetation, water, soil and building (show in table 2). In order to be classified TM data, the factorial experiment analysis (statistical analysis) which obtains the interactions is to be done at first. Later, the interactions must be served as input to neural network system. Lastly, use neural network (back-propagation learning algorithm) to be classified the TM data.

Table 2: *Categories of interesting things*

First level	Second level
Pant, vegetation	Forest Farming
Water	River Lake Irrigation
Soil	-

Building	Industry High-way

4.3 FINDING INTERACTION BETWEEN DIFFERENCE BANDWIDTH

The interaction between difference spectral reflectance will be identified by multi valuable analysis. The spectral reflectance wavelength used is ranged between 0.45-1.75 μm. The analytical process can be described as follows:

First step: Data selection

At this moment, the difference type of land use such as plant, water and man-made structure are selected. Each interested data is described in order to use as an information base for classification.

Second step: Statistical experiment design

The information obtained from first step will be used as a threshold value in this experiment. The combination of each spectral band will be used to find the degree of interaction. Since the spectral reflectance is determined for BAND #2, BAND #3 and BAND #4, The formula using for this analysis is shown in table.

Table 3: *Formula for analysis*

Three factors	
Interactions	Formula
Insides A	$\alpha_I = \mu_{I..} - \mu_{...}$
Insides B	$B_J = \mu_{J..} - \mu_{...}$
Insides C	$\gamma_K = \mu_{K..} - \mu_{...}$
Between A and B	$\alpha\beta_{IJ.} = \mu_{IJ.} - (\alpha_I + \beta_J + \mu_{...})$
Between A and C	$\alpha\gamma_{I.K} = \mu_{I.K} - (\alpha_I + \gamma_K + \mu_{...})$
Between B and C	$\beta\gamma_{.JK} = \mu_{.JK} - (\beta_J + \gamma_K + \mu_{...})$
Between A, B and C	$\alpha\beta\gamma_{IJK} = \mu_{IJK} - (\alpha_I + \beta_J + \gamma_K + \alpha\beta_{IJ} + \alpha\gamma_{IK} + \beta\gamma_{JK} + \mu_{...})$

5. REFERENCES

[1] จิราพร ชมพิกุล. <u>การวางแผนการทดลอง</u>. สงขลา : ภาควิชาคณิตศาสตร์ คณะวิทยาศาสตร์ มหาวิทยาลัยสงขลานครินทร์. 2532.

[2] Bischof, H., Schneider, W. and Pinz, A.J. 1992. "Multispectral Classification of Landsat-Images Using Neural Networks". <u>IEEE Trans. Geosci. Remote Sensing</u>. 30 (May 1992) : , 482-490.

[3] Campbell, James B. 1987. <u>Introduction to Remote Sensing</u>. New York : Guilford Press.

[4] Hagan, Martin T., Demuth, Howard B. and Mark Beale. 1996. <u>Neural Natwork Design</u>. Boston : PWS Publishing Company.

[5] Kulkarni, Arun D. 1994. <u>Artificial Neural Networks for Image Understanding</u>. New York : Van Nostrand Reinhold.

[6] Lillesand, Thomas M. and Kiefer, Ralph W. 1994. <u>Remote Sensing and Image Interpretation</u>. New York : John Wiley & Sons.

[7] Paola, Justin D. and Schowengerdt, Robert A. 1995. "A Detailed Comparison of Back-propagation Neural Network and Maximum-Likelihood". <u>IEEE Trans. Geosci. Remote Sensing</u> 33 (July 1995) : 981-995.

[8] Wasserman, Philip D. 1993. <u>Advanced Methods in Neural Computing</u>. New York : Van Nostrand Reinhold.

[9] Winer, B.J. 1962. <u>Statistical Principles in Experimental Design</u>. New York : Mcgraw-hill Book Company.

[10] Heermann, Philip D. and Khazenie, Nahid. 1992. "Classification of Multispectral Remote Sensing Data Using a Back-Propagation Neural Network". <u>IEEE Trans. Geosci. Remote Sensing</u>. 30 (January 1992) : 81-88.

[11] Salu, Yehuda. and Tilton, James. 1992. "Classification of Multispectral Image Data by the Binary Diamond Neural Network and Nonparametric, Pixel-byPixel Methods". <u>IEEE Trans. Geosci. Remote Sensing</u>. 31 (May 1993) : 606-616.

[12] Paola, Justin D. and Schowengerdt, Robert A. 1995. "A Detailed Comparison of Back-propagation Neural Network and Maximum-Likelihood Classifiers for Urban Land Use Classification". <u>IEEE Trans. Geosci. Remote Sensing</u>. 33 (July 1995) : 981-996.