NWU-RS-LAND-003

Pattern decomposition method for

Hyper-multispectral data analysis

Motomasa DAIGO
Akiko ONO, Noboru FUJIWARA
Rika URABE

2001.10.30

Pattern decomposition method for hyper-multispectral data analysis

Motomasa Daigo
Faculty of Economics, Doshisha University, Kyoto, 602-8580, JAPAN
Akiko Ono, Noboru Fujiwara
Department of Information and Computer Sciences,
Nara Women's University, Nara 630-8506, JAPAN
Rika Urabe
Tokyo Institute of Technology University, Yokohama 226-8502, JAPAN
E-mail: mdaigo@mail.doshisha.ac.jp / Fax: +81-(0)75-251-3646

August 8, 2001

Abstract

The 'Pattern decomposition method' (PDM) is a new analysis method originally developed for Landsat TM satellite data. Applying the PDM to the radio-spectrometer data of ground objects, 121 dimensional data in the wavelength region of 350 – 2,500 nm were successfully reduced into three-dimensional data. The nearly continuous spectral reflectance of land cover objects could be decomposed by three standard spectral patterns with an accuracy of 4.17% per freedom. We introduced a concept of supplementary spectral patterns for the study of specific ground objects. As an example, availability of a supplementary spectral pattern that can rectify standard spectral pattern of vivid vegetation for spectra of withered vegetation was studied. The new vegetation index (RVIPD) for hyper-multispectra is proposed as a simple function of the pattern decomposition coefficients including the supplementary vegetation pattern. It was confirmed that RVIPD is linear to the area cover ratio and also to the vegetation quantum efficiency.

1 Introduction

Reflecting the progress of technology, some recent satellite sensors can provide hyper-multispectral data. Several analysis methods have been developed for these data. There are two well known methods in them. One is the class of principal component transformation (Gonzalez and Wintz 1977, Merembeck and Turner 1980). The principal component analysis method is mathematically pleasing, but the new coordinate system has no physical meaning. The other is spectral mixing analysis (endmembers method) (Adams et.al. 1995). In this analysis, the spectrum of each pixel is approximated as the linear sum of a spectrum for each classification category. The coefficients of that linear sum express the 'belonging ratios' of each pixel to those categories.

We have developed a new analysis method for multi-spectral satellite data called the 'Pattern decomposition method' (PDM) (Fujiwara et.al. 1996, Muramatsu et.al. 2000). The PDM is one type of spectral mixing analysis but the spectrum of each pixel is basically expressed as the linear sum of fixed three standard spectral patterns, namely the spectral patterns of three representative land objects; water, vegetation and soil. In the PDM, these three patterns have no meaning of special classification category but are standard spectral pattern components of general spectra. The spectral pattern for most ground objects can be reconstructed by the three standard patterns. The usage of standard spectral patterns makes possible the comparison of data from different time and also from different sensors with the same criteria (Hayashi et.al. 1998).

According to the application of the PDM to Landsat TM data and simulated ADEOS-II GLI data, it can be said that reflected light from most land cover objects can be approximated with a linear combination of the three standard patterns of water, vegetation and soil with good accuracy (Hayashi et.al. 1998, Furumi et.al. 1998). Especially in hyper-multispectral data analysis, data reduction without losing any information is important. In a sense, the standard patterns in the PDM are one kind of three principal axes in n-dimensional space but have physical meaning. To make it possible, the PDM adopts a oblique coordinate system. The three standard patterns are the axes of the coordinate system.

In general, spectral patterns of ground objects as a function of wavelength are decom-

posed into a smoothly varying component and a resonance component.

$$spectral\ pattern = (smooth\ component) + (resonant\ absorption\ component) \quad (1)$$

For example, spectral patterns of water, soil and concrete vary smoothly as a function of wavelength. Spectral patterns of vegetation and mineral have a smooth component and a broad resonant absorption component.

To express a resonant part well, a supplementary spectral pattern in addition to the three-spectral pattern is available. For example, a characteristic absorption pattern for minerals could be used as a supplementary pattern. In the case of vegetation, the sharp resonant absorption part of the spectrum smooths on progress from a vivid vegetation state to a withered state, and finally shows like a soil spectral pattern. A supplementary pattern which rectifies resonant absorption pattern of vivid standard vegetation is useful for the detailed analysis of vegetation change.

In this paper, the applicability of the PDM to nearly continuous spectrum data is studied and the PDM included a supplementary spectral pattern for withered vegetation is developed. For this study, we used radio-spectrometer data measured in the field and data measured with an airborne multispectral scanner.

2 The pattern decomposition method

2.1 PDM with three standard spectral patterns

In the PDM, a set of reflectance (or brightness) data of n-bands for each pixel is used. As mentioned above, the spectral patterns of water, and soil are smooth varying patterns. The spectral pattern of vegetation has a smooth part and a broad resonant absorption part. The set of data for each pixel is decomposed by standard spectral patterns of water, vegetation and soil as follows:

$$A(i) \rightarrow CwPw(i) + CvPv(i) + CsPs(i)\ , \quad (2)$$

where $A(i)$ is the reflectance of band i, $Pw(i)$, $Pv(i)$ and $Ps(i)$ are standard spectral patterns of water, vegetation and soil and normalized as $\sum_i |Pk(i)| = 1$, $(k = w, v, s)$ and Cw, Cv and Cs are pattern decomposition coefficients. The three coefficients Cw, Cv and Cs are evaluated using the least squares method.

The sum of $A(i)$ for all bands is approximated as

$$\sum_i A(i) = Cw + Cv + Cs \ . \tag{3}$$

When brightness data are used as $A(i)$, the sum of the three coefficients is total brightness of the pixel. In the case of the sum of the three coefficients being normalized to unit;

$$Cw + Cv + Cs = 1 \tag{4}$$

each coefficient represents the ratio of spectral patterns of three components. In this case, the coefficients for the pixel are not affected by the shadow in the satellite image.

Using the residual of i band's reflectance,

$$R(i) = A(i) - \{CwPw(i) + CvPv(i) + CsPs(i)\} \ , \tag{5}$$

the error index of reduced-χ^2 for equation (2) is defined as follows:

$$\chi^2 = \sum_i R(i)^2/(n-3) \ , \tag{6}$$

where $n-3$ is the degree of freedom for a data set of n-bands.

2.2 PDM with a supplementary spectral pattern in addition to three standard patterns

A supplementary pattern is available for special study such as detailed analysis of vegetation change or analysis of minerals which includes a resonance absorption spectral pattern. In this section, PDM with a supplementary pattern in addition to three standard patterns is discussed for detailed analysis of the vegetation change from vivid state to dead state as an example.

$R(i)$ data set of equation (5) for dead leaves is used as a supplementary spectral pattern $Pd(i)$. The decomposition coefficients Cw, Cv, Cs and the coefficient for the supplementary pattern Cd are evaluated using the least squares method.

$$A(i) \rightarrow CwPw(i) + CvPv(i) + CsPs(i) + CdPd(i) \ , \tag{7}$$

where $Pd(i)$ is normalized as $\sum_i |Pd(i)| = 1$. Even in this case, equation (3) is right, because $Pd(i)$ is the residuals of dead leaves and $\sum_i Pd(i)$ is equal to zero.

The revised residual of i band's reflectance is as follows;

$$R(i) = A(i) - \{CwPw(i) + CvPv(i) + CsPs(i) + CdPd(i)\} \ . \tag{8}$$

The reduced-χ^2 in this case is defined as follows

$$\chi^2 = \sum_i R(i)^2/(n-4) \ . \tag{9}$$

2.3 Information Transfer

To study information transfer from original data to transformed coefficients by the PDM, the information defined by Shannon in bit unit

$$I = -\sum p(i) \log_2\{p(i)\} \tag{10}$$

was used, where $p(i)$ is the probability of event i. For satellite data, $p_k(x)$ is probability of frequency distribution of reflectance x for band k.

3 Data used in this study and three standard & supplementary spectral patterns

Reflectance was measured in the field with the radio-spectrometer of Field Spec FR (Analytical Spectral Devices Inc.) or MSR7000 (Opto Research Corp.). Both radio-spectrometers give raw spectral values every 1 nm from wavelength of 350 nm to 2,500 nm with a spectral resolution of 3 nm to 10 nm. However, many satellites' sensors have resolutions of about 10 nm. Therefore we averaged every 10 data points to simulate 10 nm resolution data in the analysis. This procedure also makes our results not to sensitive to the performance difference of two radio-spectrometers.

The samples were measured indoors using a halogen lamp or outdoors with solar light. The distances between the receptor and samples were about 50 cm. Due to the usage of solar light, the spectral region used for this analysis was restricted to where the atmospheric transmittance was higher than 80%. After measuring spectra for a sample and a standard white board (Spectralon Reflectance Target of Labsphere Inc.) one by one, a reflectance for the sample was calculated as a ratio of the raw spectral value of the sample to that of the standard white board.

In all, 121 bands with 10 nm width were selected as shown in figure 1 with three standard reflectance spectra as examples. The number of samples for this analysis was 1068.

Airborne multispectral scanner (AMSS) data of Sakata, Japan were also used for this study. The AMSS was developed for the ADEOS-II/GLI project. The ground resolution of AMSS is 5m. The characteristics of the bands used in this analysis are shown in table 1. Only Rayleigh scattering are subtracted as atmospheric correction. It is the same correction method applied in the analysis of TM and MSS data (Fujiwara et.al. 1996, Muramatsu et.al. 2000).

The normalized standard spectrum patterns of water, vegetation and soil are shown in figure 2. The samples for the standard spectrum patterns are the sea at Kata port, Wakayama in Japan, ten overlapped green leaves of Quecus glauca and dry desert sand near Dunhuang in China, respectively.

The normalized supplementary pattern $Pd(i)$ in equation (7) is obtained from the residual $R(i)$ of equation (5) for dead leaves of Cinnamomum camphora gatherd at the ground of Nara Women's University on April 19th, 1997 and shown in figure 3.

4 Reproducibility of observed spectra and information transfer with PDM

4.1 Reproducibility of observed spectra with PDM and χ^2

The typical examples for pattern decomposition coefficients and residuals are shown in figure 4. These examples were arbitrarily selected. The green leaves have large Cv values and small Cs values. On the other hand, the dead leaf has small Cv value and large Cs value. Amplitudes of the original band data are almost exactly expressed by the three coefficients and the residuals are very small.

The original reflectance spectra and the reconstructed reflectance spectra of 121 bands from the three standard pattern decomposition coefficients and from the supplementary coefficient in addition to the three coefficients are shown in figure 5. These leaf samples were arbitrarily selected to display the discolour process from green to broun. The spec-

trum for yellow-brown and dead leaves reconstructed with the additional supplementary coefficient is reproduced well in comparison with the spectrum reconstructed with only three coefficients. The supplementary spectral pattern, which is the rectified resonant absorption pattern of vivid standard vegetation pattern, was useful for detailed analysis of the vegetation change from vivid state to withered state.

The frequency distribution of χ^2 defined by equation (6) for all samples of 1068 is shown in figure 6 and the average value is 0.00174 for PDM with three standard patterns. The square root of 0.00174 is 0.0417 (4.17 %) and is the fitting error per degree of freedom. The continuous spectral response patterns could be decomposed using only the three standard patterns with 4.17 % error per degree of freedom. For the PDM with the supplementary coefficient, the average χ^2 value was a little smaller than the value for the PDM with the three standard patterns.

4.2 Information transfer for the PDM with three standard patterns

We applied the principal component transformation method using a correlation matrix to about 600 samples measured on the ground. In the following analysis, we compare two kind of subspace in the 121 dimensional space. Both subspace have dimension of 3. One is constructed from the first three principal axes obtained by the principal component transformation. The other is constructed from the three standard patterns of the PDM. To avoid confusion, we give the names as "principal subspace" and "PDM subspace" respectively.

We calculated contribution rates of principal axes in the principal component transformation method. The results are shown in table 2 for the first 5 principal axes. The cumulative contribution rate for first three principal axes is 94.9 %.

The contribution rate contained in the PDM subspace was also calculated. It's value is 92.2%. This value is a little bit lower than that of the principal subspace. However each standard spectrum pattern of the PDM subspace has a meaning individually unlike the three principal axes of the principal subspace.

We have checked the positional relation of the two subspaces next. The projected lengths of three standard unit vectors into the principal subspace were 0.938, 0.994 and

0.993 for water, vegetation and soil, respectively.

Therefore we can conclude that the PDM subspace is close to the principal subspace, and both subspaces contain most information of image data of ground objects.

Using equation (10), the information included in the original reflectance data and the three PDM coefficients were evaluated. As a result, 95.5 % of the information in the original data could be transformed into three decomposition coefficients. The PDM is thus available for data reduction of hyper-multispectral data.

4.3 Reproducibility of NDVI

The traditional vegetation index NDVI uses only information of reflectance for red and near infrared wavelengths. The reproducibility of NDVI thus depends on the reproducibility of only these two bands. Figure 7 shows the relationsip between NDVI calculated from the reconstructed spectrum for PDM with the supplementary pattern and NDVI calculated from the original observed spectra for field measurement data and AMSS data except water. The reproducibility was good, especially for the AMSS data.

5 Relationship between χ^2 and band number

Next, we studied the relationship between the band number and χ^2 for the PDM with three standard patterns. For this study, several data sets with different numbers of bands were selected from the data of 1069 samples observed on the ground with 121 bands. We picked up one band from every 2 bands, every 4 bands, every 8 bands and so on. For example, 1 band was picked up from every 2 bands and two data sets were made from the data with 121 bands. 75 data sets were prepared in total. We also made simulated TM and ADEOS-II/GLI data sets which had 6 and 18 bands, respectively.

Figure 8 shows the correlation between the χ^2_{n-3} and the band number n. The χ^2_{n-3} decreased according to the increase in band number n and converged to a constant value of 0.00174 for n larger than 15. The square root of 0.00174 is 0.0417 (4.17 %) and is the fitting error per degree of freedom. Therefore, in the PDM, the spectral reflectance for band numbers greater than 15 can be fitted within the error of 4.17 % at all measured points using only the three standard patterns. It was concluded as a general rule that

land cover objects could be decomposed by the three standard spectral patterns.

6 New vegetation index and It's availability

6.1 New vegetation index for hyper-multispectral data

A new vegetation index VIPD based on the PDM with three standard patterns has been proposed by A. Hayashi (Hayashi et.al. 1998) and S. FURUMI (Furumi et.al. 1998). The VIPD is a linear function of three pattern decomposition coefficients. Revised VIPD (RVIPD) based on the PDM with the supplementary pattern in addition to the three standard patterns is given by the following simple formula;

$$RVIPD = (Cv - Cd)/(Cw + Cv + Cs) \ . \tag{11}$$

Owing to minimized shadow effects, this formula is normalized with total reflectance or total brightness. The NDVI is a function of reflectance of two wavelengths, namely red and near infrared. On the other hand, the VIPD and RVIPD are functions of all observed wavelengths and sensitive to vegetation state even if the pixel is mixed.

In figure 9, the relationship between NDVI and RVIPD is shown. The NDVI is saturated above 0.8 as a function of RVIPD. This means that the one state in NDVI corresponds to a different state in RVIPD.

6.2 Relationship between RVIPD and Area Cover Ratio

The same data of reflectance and area cover ratio measured in the laboratory using green leaves and soil by Hayashi et.al. (Hayashi et.al. 1998) were used. Here the area cover ratio means purely geometrical one. The details of the measurement is described in the reference.

The relationships between two vegetation indices and area cover ratio are shown in figure 10. The horizontal axis shows the area cover ratio and the vertical axis shows the vegetation index obtained from measured reflectance. The dotted lines are quadratic curves fitted by the least squares method. The first and third terms of RVIPD are negligibly small. Meanwhile, these terms of NDVI are not so small. Linear relation between RVIPD and area cover ratio was confirmed.

6.3 Relationship between RVIPD and Quantum Efficiency

The same data of reflectance and chlorophyll for many kinds and many statuses of leaves used by Furumi et.al. (Furumi et.al. 1998) were used to obtain the relationships between two vegetation indices and chlorophyll content. The chlorophyll content of leaves 3 – 5g in weight was extracted using acetone and $MgCO_3$, and measured its density by an absorptiometer (the Arnon method (Yoshimura et.al. 1994)). The details of the measurement are described by Yoshimura et.al. (Yoshimura et.al. 1994) and Furumi et.al. (Furumi et.al. 1998). The results are shown in figure 11. Using the relationship between chlorophyll and quantum efficiency (figure 12) (Gabrielsen 1948), relationships between vegetation indices and quantum efficiency were obtained as shown in figure 13. It was confirmed that the RVIPD and NDVI are both linear to quantum efficiency.

7 Summary

We studied the applicability of the Pattern Decomposition Method (PDM) to a set of nearly continuous spectral reflectance data in the wavelength range from 350 nm to 2,500 nm. 121 bands were selected where the atmospheric absorption was less than 20%. For all 1068 samples measured in the field, spectral response patterns could be decomposed using only three standard spectral patterns with about a 4.17 % error per degree of freedom. It was understood as a general rule that land cover objects can be decomposed by three standard spectral patterns, namely water, vegetation and soil.

The information included in the original reflectance data and the three PDM coefficients were evaluated and it was confirmed that 95.5 % of the information in the original data can be transformed into three decomposition coefficients. The PDM is available for data reduction of hyper-multispectral data.

The relationship between the band number and χ^2 for the PDM with three standard patterns showed that χ^2 is decreased with increasing band number and converged to a constant value of 0.00174 above a band number of 15. The value of 0.00174 corresponds to 0.0417 (4.17 %), the fitting error per degree of freedom.

For detailed analysis of the vegetation change from vivid state to withered state, a supplementary spectral patterns in addition to the three standard patterns are available.

The new vegetation index RVIPD is proposed as a simple function of the four pattern decomposition coefficients which are linear to the original data of reflectance or brightness. It was confirmed that the RVIPD is linear to the vegetation cover ration and quantum efficiencies.

The PDM is applicable to any optical sensors with several bands in the same framework. We have opened the three standard patterns and the supplementary pattern to the public on the World Wide Web.

Acknowledgements

This work was supported under the ADEOS-II/GLI project by the National Space Development Agency of Japan (NASDA) and the Academic Frontier Promotion project by Ministry of Education, Science, Sports and Culture, Japan.

References

GONZALEZ, R.C., and WINTZ, P., 1977, Digital Image Processing (Mass.: Addison-Wesley)

MEREMBECK, B.F., and TURNER, B.J., 1980, Directed canonical analysis and the Performance of Classifiers under its associated linear transformation. *IEEE Trans. Geoscience and Remote Sensing*, **GE-18**, 190–196.

ADAMS, J.B., SABOL, D.E., KAPOS, V., FILHO, R.A., ROBERTS, D.A., SMITH, M.O., and GILLESPIE, A.R., 1995, Classification of multispectral images based on fractions of endmembers: application to land-cover change in the Brazilian Amazon. *Remote Sensing of Environment*, **52**, 137–154.

FUJIWARA, N., MURAMATSU, K., AWA, S., HAZUMI, T., and OCHIAI, F., 1996, Pattern expand method for satellite data analysis. (in Japanese) *Journal of The Remote Sensing Society of Japan*, **Vol. 17, No.3**, 17–34.

MURAMATSU, K., FURUMI, S., FUJIWARA, N., HAYASHI, A.,, DAIGO, M., and OCHIAI, F., 2000, Pattern decomposition method in the albedo space for Landsat TM and MSS data analysis. *International Journal of Remote Sensing*, **Vol. 21, No.1**, 99–119.

HAYASHI, A., MURAMATSH, K., FURUMI, S., SHIONO, Y., FUJIWARA, N. and DAIGO M., 1998, An algorithm and a new vegetation index for ADEOS-II/GLI data analysis. *Journal of The Remote Sensing Society of Japan*, **Vol.18, No.2**, 28–50.

FURUMI, S., HAYASHI, A., MURAMATUS, K., and FUJIWARA, N., 1998, Relation between vegetation vigor and a new vegetation index based on pattern decomposition method. *Journal of The Remote Sensing Society of Japan*, **Vol.18, No.3**, 28–50.

YOSHIMURA, H., ISHIDA, M., KOBASHI, S., and OHTE, K., 1994, Effects of plant pigments concerning seasonal optical properties of tree leaves in the visible region. (in Japanese) *Journal of the Japanese Society of Revegetation Technology*, **Vol. 20, No.2**, 99–110.

E. K. GABRIELSEN, E.K., 1948, Effects of different chlorophyll concentrations on photosynthesis in foliage leaves. *Physiol. Plant.*, **Vol. 1**, 5–13.

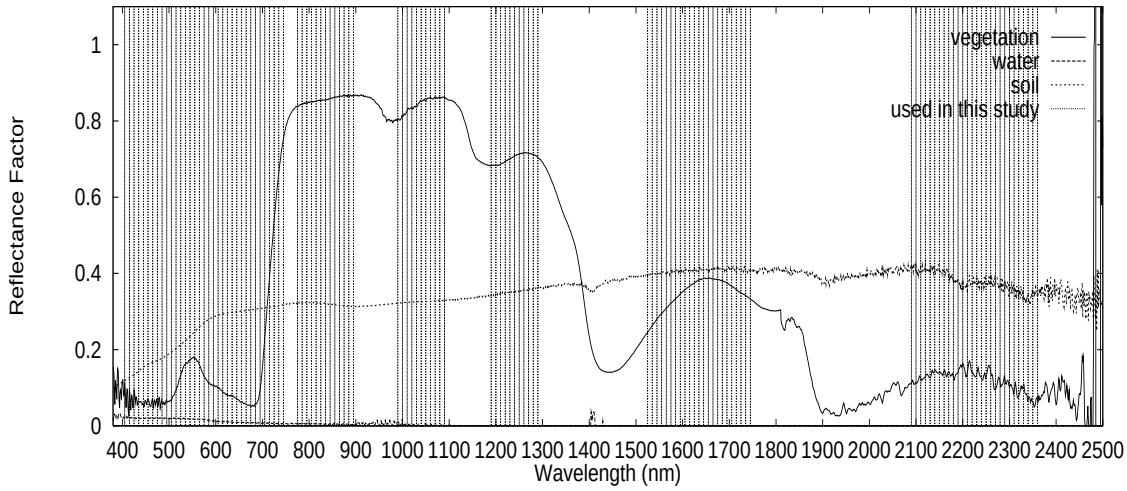


Figure 1. Selected 121 band indicated by vertical lines and three reflectance spectra of typical vegetation, water and soil.

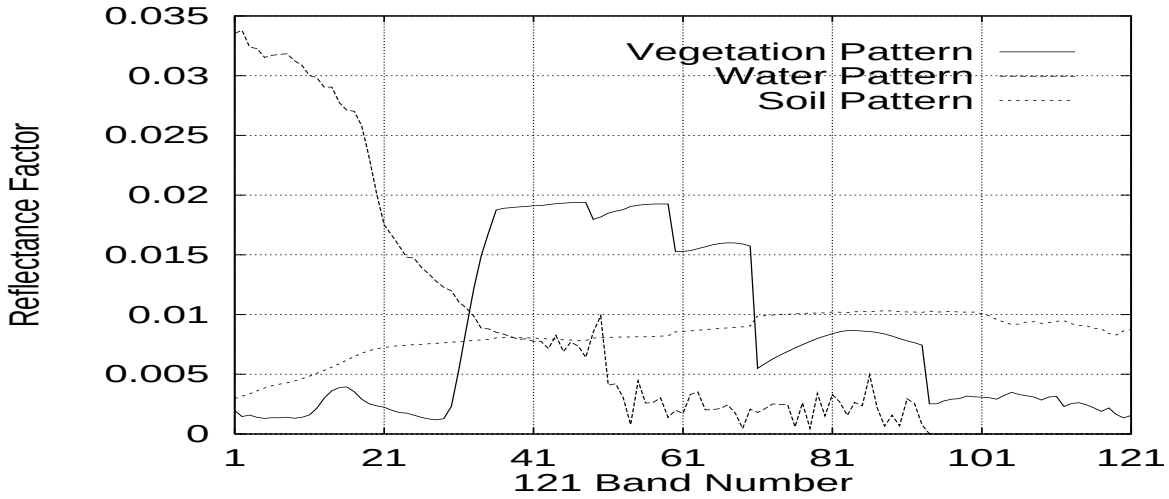


Figure 2. Three normalized standard patterns of vegetation, water and soil.

Table 1. AMSS spectral bands and the band number defined in this study.

Record Sequence of AMSS	Central Wavelength (nm)	Band Width (nm)	Band Number Defined in This Study
r1	405.3	5.5	1
r2	412.6	9.9	2
r5	442.3	9	3
r7	463.2	8.8	4
r10	489.6	8.9	5
r12	522.8	9.2	6
r15	545.4	9	7
r17	561.3	7.8	8
r20	627.3	10	9
r28	668.1	9.3	10
r23	675.7	8.9	11
r29	683.4	10.3	12
r27	750	9.4	13
r31	784.2	13.4	14 (average from r31 to r35)
r32	800.9	12.6	
r33	829.4	13.6	
r34	848.1	20.6	
r35	866	15.5	
r38	1050.8	16.1	15
r39	1239.8	20.5	16
r40	1631.8	217.5	17
r41	2233	198.3	18

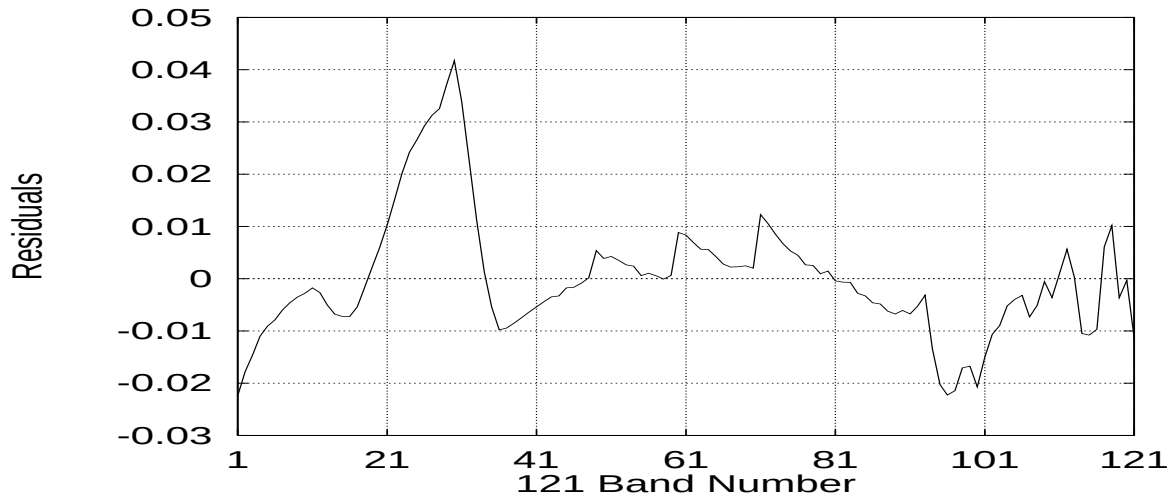


Figure 3. The supplementary pattern defined from the dead leaf of Cinnamomum camphora for 121 bands.

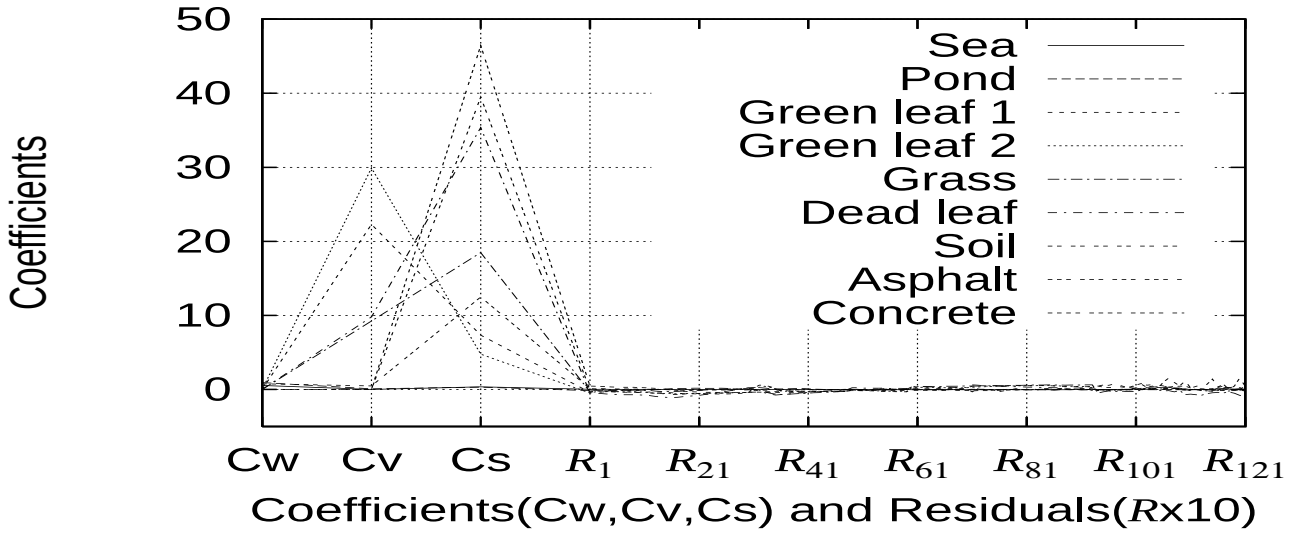
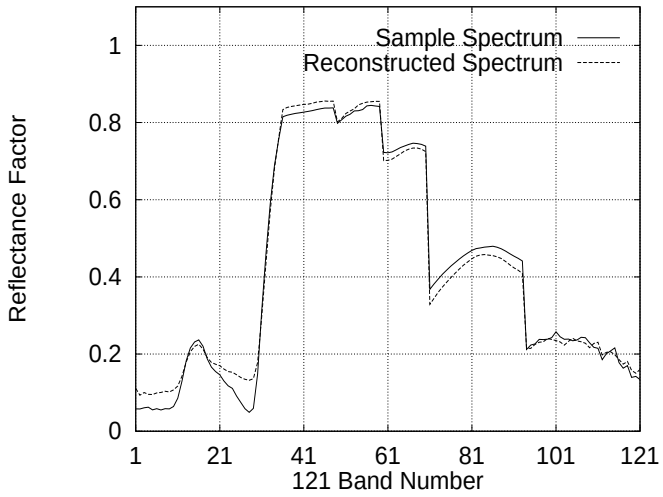
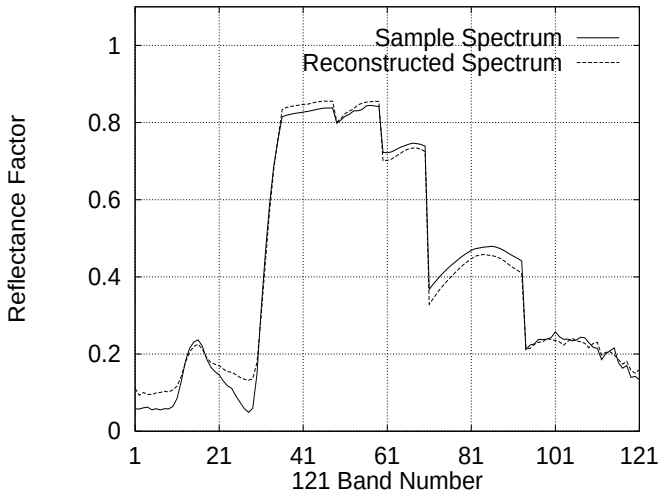


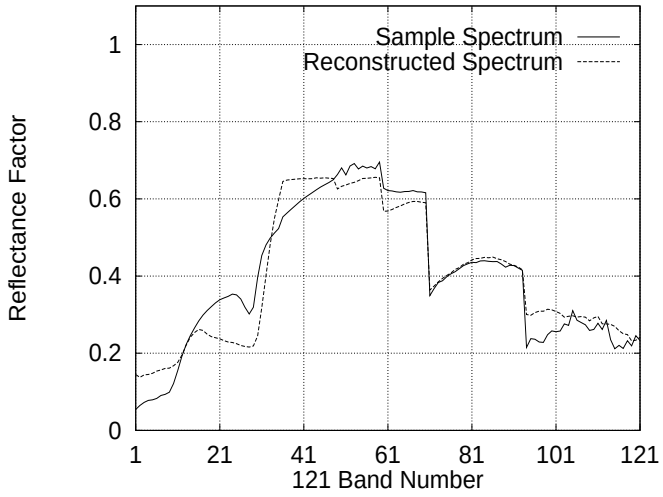
Figure 4. Three pattern decomposition coefficients and residuals for typical samples.

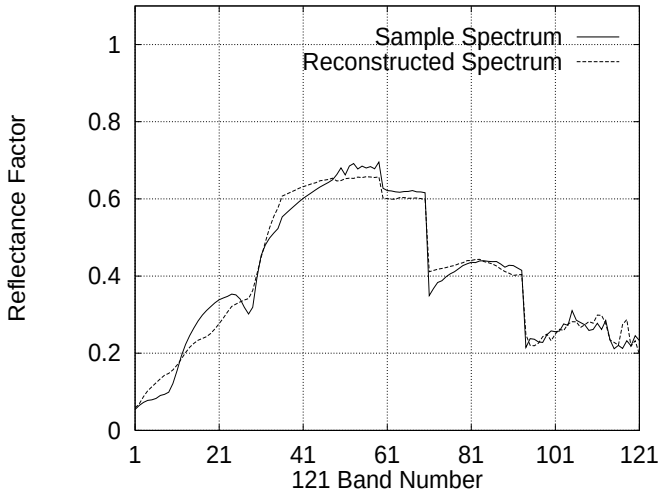
(a) Green Leaf



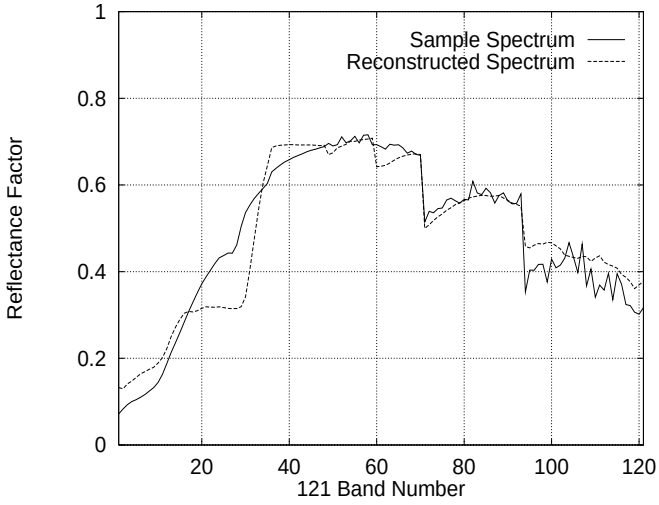


(b) Yellow-brown Leaf





(c) Dead Leaf



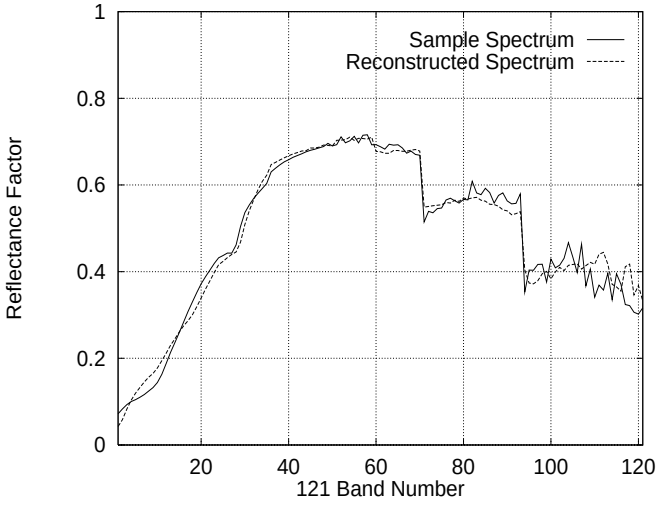


Figure 5. The original reflectance spectra and reconstructed reflectance spectra from the three standard pattern decomposition coefficients (left side) and from four coefficients (right side).

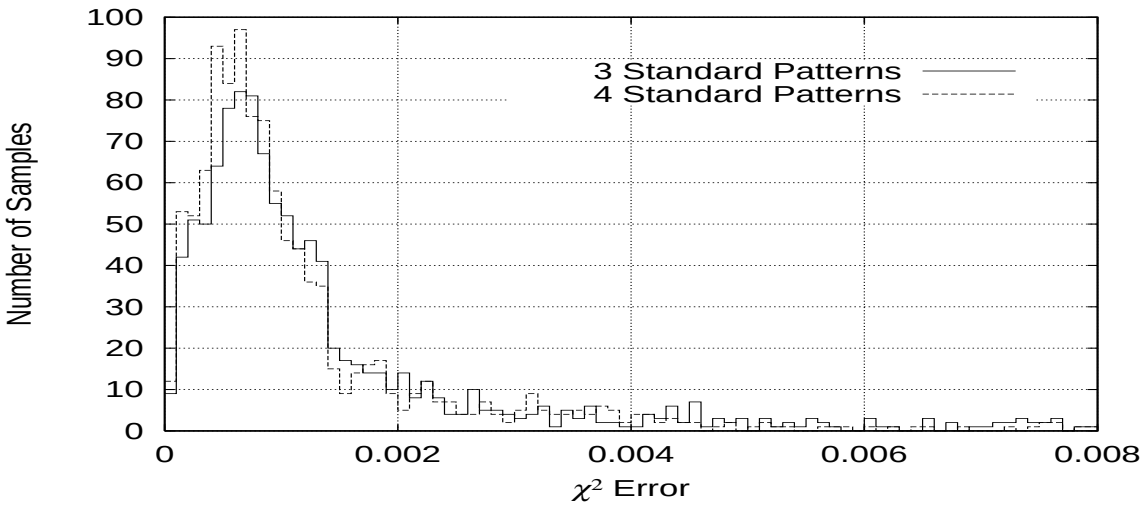


Figure 6. Frequency distribution of χ^2 for PDM with three standard patterns and for PDM with additional supplementary pattern.

Table 2. The cumulative contribution rate of the principal components.

the number of axis	principal value	cumulative rate
1	77.97	0.6444
2	24.77	0.8491
3	12.05	0.9486
4	2.05	0.9656
5	1.32	0.9765

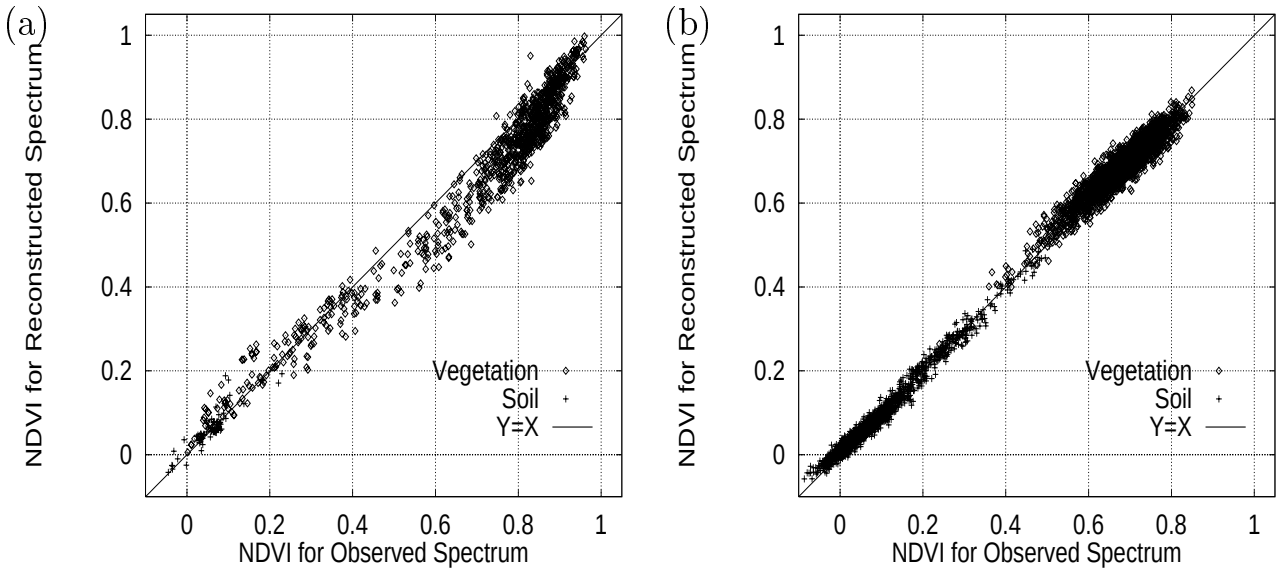


Figure 7. Reproducibility of NDVI for ground measurement data (a) and AMSS data (b). Horizontal axis shows NDVI calculated from original spectra and vertical axis shows NDVI calculated from reconstructed spectra.

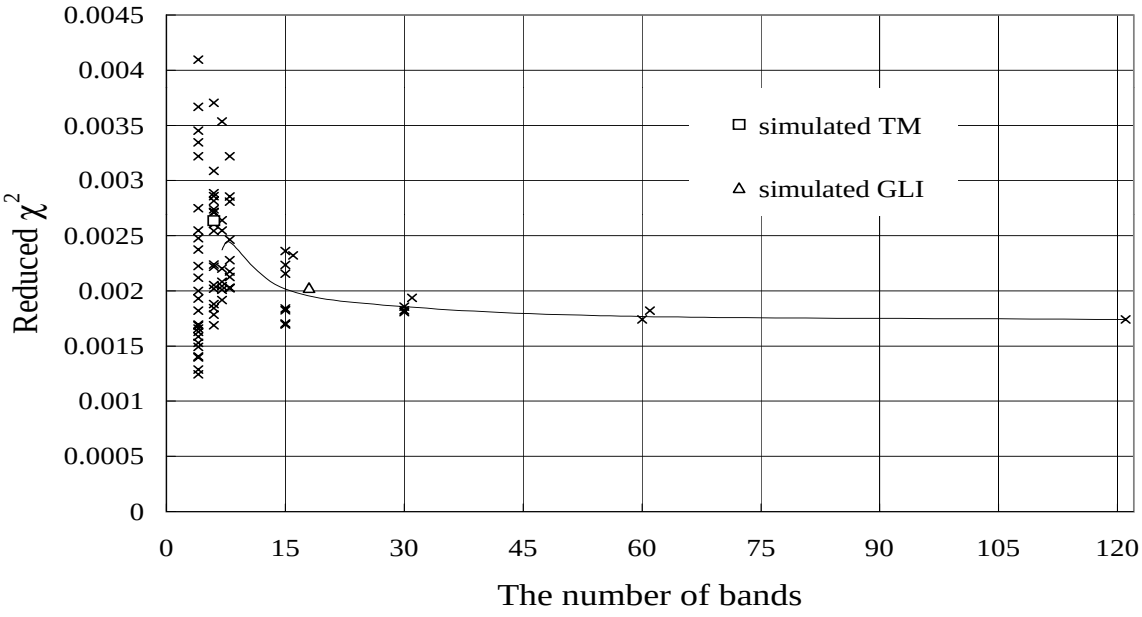


Figure 8. Reduced χ^2 as a function of the number of bands.

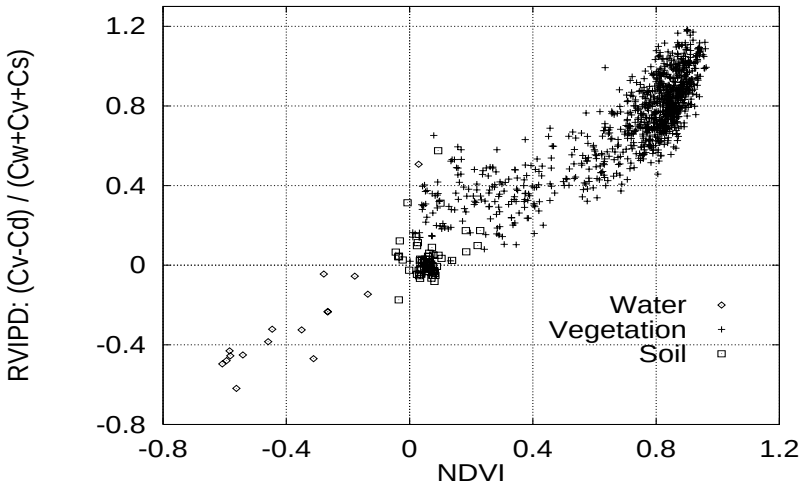


Figure 9. The relation between RVIPD and NDVI.

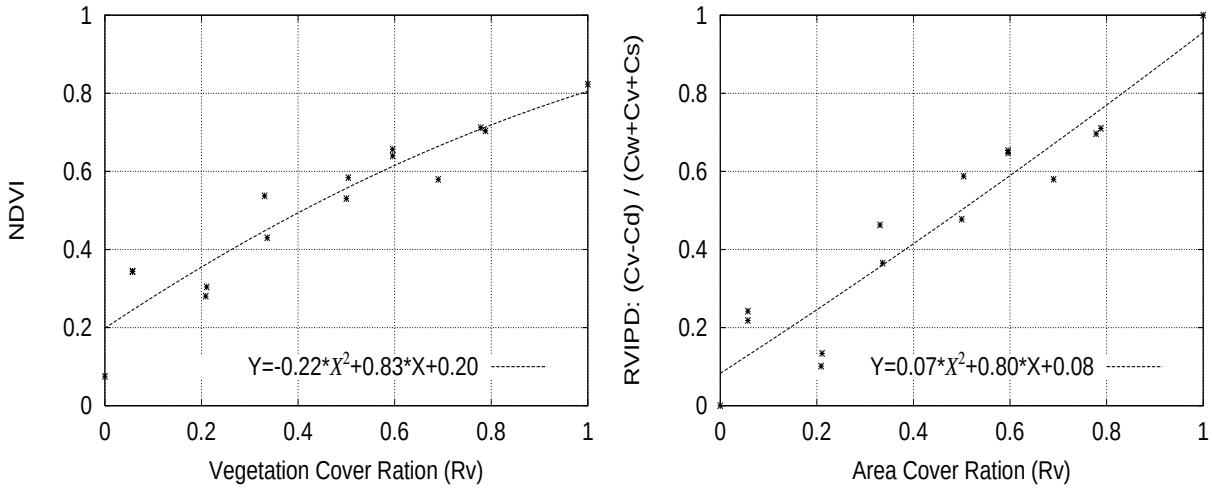


Figure 10. Relationships between the ared cover ratio and vegetation indices.

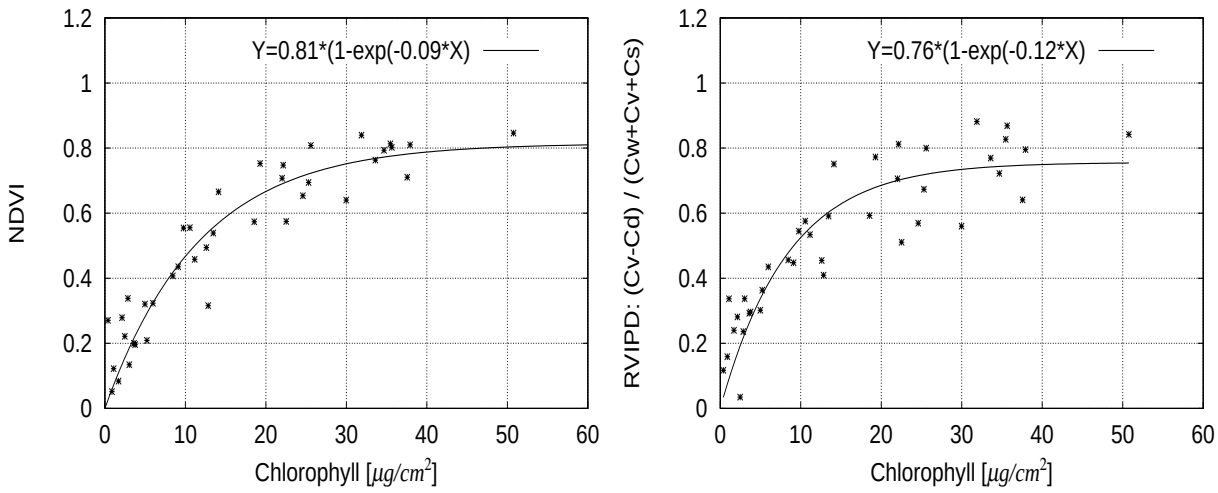


Figure 11. Relationships between the chlorophyll content and vegetation indices.

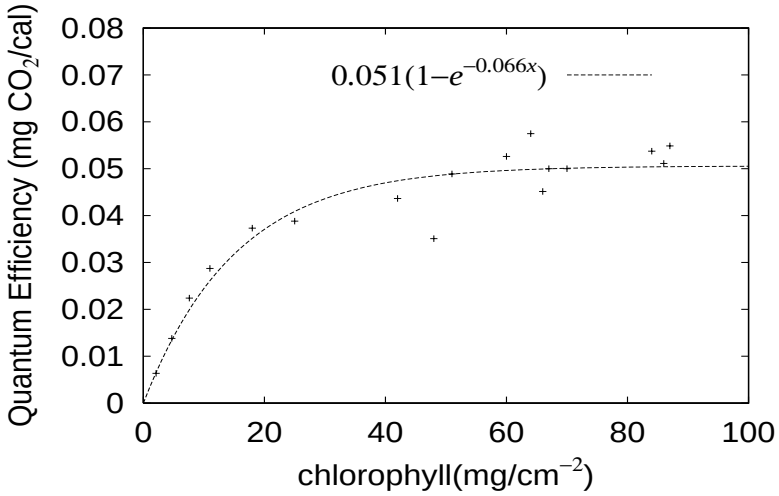


Figure 12. Relationship between the chlorophyll content per unit leaf area and photosynthetic activity (quantum efficiency). (E.K.Gabrielsen[9])

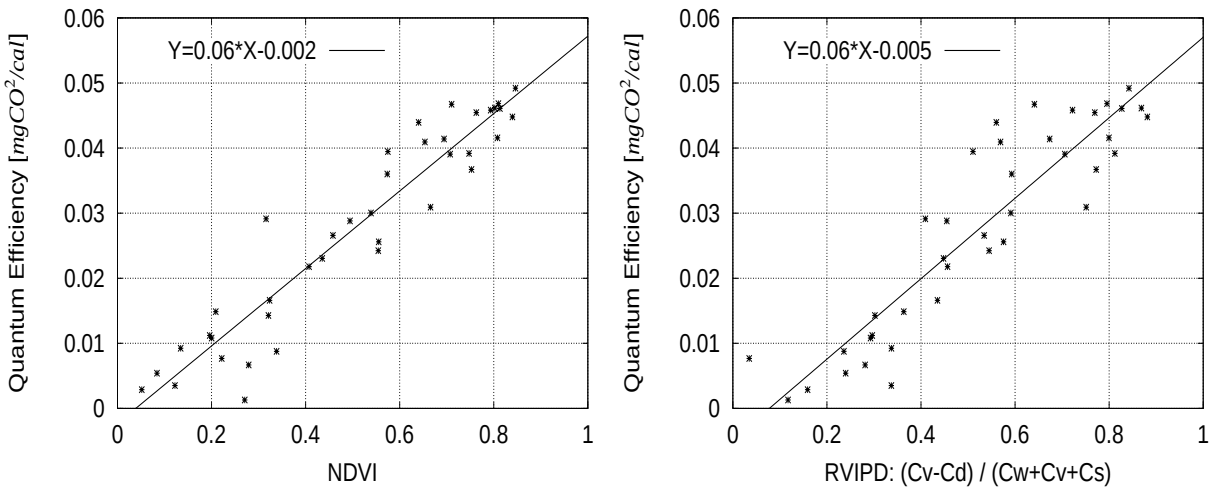


Figure 13. Relationship between the quantum efficiency and vegetation indices.