# Comparison of Spectral Indices and Principal Component Analysis for Differentiating Lodged Rice Crop from Normal Ones

Zhanyu Liu[1,2,*], Cunjun Li[3], Yitao Wang[4], Wenjiang Huang[3], Xiaodong Ding[1], Bin Zhou[1], Hongfeng Wu[4], Dacheng Wang[3], and Jingjing Shi[5]

[1] Institute of Remote Sensing and Earth Sciences, Hangzhou Normal University, Hangzhou 311121, China

[2] Key Laboratory of Urban Wetland and Region Change in Zhejiang Province, Hangzhou 311121, China

[3] Beijing Research Center for Information Technology in Agriculture, Beijing 100097, China

[4] Institute of Scientific and Technological Informatics, Heilongjiang Academy of Land Reclamation Sciences, Harbin 150036, China

[5] Institute of Agricultural Remote Sensing & Information Technology, Zhejiang University, Hangzhou 310029, China

liuzhanyu@zju.edu.cn

**Abstract.** Hyperspectral reflectance of normal and lodged rice caused by rice brown planthopper and rice panicle blast was measured at the canopy level. Over one decade broad- and narrow-band vegetation indices (VIs) were calculated to simulate Landsat ETM+ with *in situ* hyperspectral reflectance. Principal component analysis (PCA) was utilized to obtain the front two principal components (PCs). Probabilistic neural network (PNN) was employed to classify the lodged and normal rice with VIs and PCs as the input vectors. PCs had 100% of overall accuracy and 1 of Kappa coefficient for the training dataset. While PCs had the greatest average overall accuracy (97.8%) and Kappa coefficient (0.955) for the two testing datasets than VIs consisting of broad- and narrow-bands. The results indicated that hyperspectral remote sensing with PCA and artificial neural networks could potentially be applied to discriminate lodged crops from normal ones at regional and large spatial scales.

**Keywords:** Hyperspectral remote sensing, Lodged rice, Principal component analysis (PCA), Vegetation indices (VIs), Artificial neural networks (ANN).

## 1 Introduction

Detecting plant health condition plays an important role in controlling disease and insect stress in precise pest management (PPM), which always results in yield loss and poor quality [1]. Although plant pest stress is predominantly concentrated in patches around stress centers, it is still widespread to spray agrochemicals indiscriminately over the entire field in practice [2]. It is necessary to accurately assess the pest stress distribution and damage caused by disease and insect pests [3]

* Corresponding author.

D. Li and Y. Chen (Eds.): CCTA 2011, Part II, IFIP AICT 369, pp. 84–92, 2012.

so that agrochemicals can be applied to reduce environmental pollution due to agrochemicals in the stressed patches.

Lodging at the late growth stage is caused by insect and fungal disease in paddy rice field, which often leads to yield loss and quality decrease [4]. Accurate distribution information of lodging rice is helpful to spray agrochemicals and harvest

**Table 1.** Broadband and narrowband spectral indices used in this study

| *Abbreviation* | *Formula* | *Reference* |
|---|---|---|
| DVI | $NIR-R$ | Richardson and Everitt [11] |
| RVI | $NIR/R$ | Jordan[12] |
| NDVI | $(NIR-R)/(NIR+R)$ | Rouse *et al.* [13] |
| $SAVI_{L=0.5}$ | $1.5(NIR-R)/(NIR+R+0.5)$ | Huete[14] |
| OSAVI | $1.16(NIR-R)/(NIR+R+0.16)$ | Rondeaux *et al.* [15] |
| $MSAVI_2$ | $\frac{2NIR+1-\sqrt{(2NIR+1)^2-8(NIR-R)}}{2}$ | Qi *et al.* [16] |
| TVI | $60(NIR-G)-100(R-G)$ | Broge *et al.* [17] |
| IPVI | $NIR(NIR+R)$ | Crippen [18] |
| TDVI | $\frac{1.5(NIR-R)}{\sqrt{NIR+R+0.5}}$ | Bannari *et al.* [19] |
| RDVI | $\frac{NIR-R}{\sqrt{NIR+R}}$ | Roujean and Breon [20] |
| EVI | $2.5\frac{NIR-R}{NIR+6.5R-7.5B+1}$ | Liu and Huete [21] |
| GEMI | $\eta(1-0.25\eta)-(R-0.125)/(1-R)$ <br> $\eta=\frac{2(NIR^2-R^2)+1.5NIR-0.5R}{NIR+R+0.5}$ | Pinty and Verstraete[22] |
| BI | $0.2909B+0.2493G+0.4806R+0.5568NIR+0.4438SWIR_1+0.1706SWIR_2$ | Crist *et al.* [23] |
| GVI | $-0.2728B-0.2174G-0.5508R+0.7721NIR+0.0733SWIR_1-0.1648\ SWIR_2$ | Crist *et al.* [23] |
| WI | $0.1446B+0.1761G+0.3322R+0.3396NIR-0.621SWIR_1-0.4186\ SWIR_2$ | Crist *et al.* [23] |
| NDWI | $(\lambda_{860nm}-\lambda_{1240nm})/(\lambda_{860nm}+\lambda_{1240nm})$ | Gao [24] |
| LSWI | $(\lambda_{860nm}-\lambda_{1640nm})/(\lambda_{860nm}+\lambda_{1640nm})$ | Xiao *et al.* [25] |
| NMDI | $\frac{\lambda_{860nm}-(\lambda_{1640nm}-\lambda_{2130nm})}{\lambda_{860nm}+(\lambda_{1640nm}-\lambda_{2130nm})}$ | Wang *et al.* [26] |

*$B$,$G$,$R$,$NIR$,$SWIR_1$ and $SWIR_2$ represent surface spectral reflectance averaged over ranges of wavelengths in Landsat ETM+, viz.TM1 (450-515nm), TM2 (525-605nm), TM3 (630-690nm), TM4 (775-900nm), TM5 (1.55-1.75μm ) and TM7 (2.09-2.38μm ), respectively. $\lambda_{i\ nm}$ denotes surface spectral reflectance at *i* nanometer.

in agricultural field management. Remote sensing has been proved to be more effective than ground surveys for detecting plant health condition in numerous studies [5]. Perhaps vegetation indices (VIs) is the most popular technique of remote sensing to predict agricultural crop biophysical variables such as leaf area index, chlorophyll content/concentration, above-ground biomass, and percent vegetation cover and so forth [6]. In addition to VIs, principal component analysis (PCA) has been extensively studied as a data compression technique [7].

Furthermore, the application of artificial intelligent methods including artificial neural networks (ANN) and support vector machines (SVM) in plant stress detection has also been reported recently[8-9]. For example, SVM technology was used to detect weed and nitrogen stress in corn by Karimi *et al* [8]. Shi *et al.* [9] applied SVM model to discriminate the health status in rice and great classification was obtained. However, few studies have been conducted to resolve classification problems by VIs and PCA [10].

The main goal of this research is to evaluate and compare the performances of broad- and narrow-band VIs (Table 1) and PCA in differentiating the normal rice crop from the lodged ones stressed by rice brown planthopper or rice panicle blast.

## 2 Methodology

### 2.1 Study Sites and Materials

There were two sampling sites in this research. One site was located in Friendship Farm (120°43′E, 46°39′N), Heilongjiang Province. The rice crop cultivar was Suijing 4, which was naturally infected with rice panicle blast. The other site was in Jiuxian Village (119°37′E, 29°48′N), Tonglu County, Zhejiang Province. And the rice crop cultivar was Hybrid rice 718, which was stressed by rice brown planthopper.

### 2.2 Hyperspectral Measurement

Canopy hyperspectral reflectance were measured using an ASD FieldSpec Pro FRTM Spectroradiometer (Analytical Spectral Devices Inc., Boulder, CO, USA) throughout the whole spectral range from 350 nm to 2 500 nm. The spectral resolution of the instrument varied from 3 nm (<1 000 nm) to 10 nm (>1 000 nm), while the spectra were interpolated by the spectrometer software in 1nm interval. Therefore, each measurement generated a spectrum ranging between 350 nm and 2500 nm at 1nm increment.

The fiber optic sensor with a 25° instantaneous field of view (IFOV) was pointed to the paddy rice canopy about 1 m height at the nadir position. Reference panel measurements were made at the beginning of each set of canopy measurements. The reference panel painted by BaSO4 is a Lambert surface with a reflectance of no less than 99%. These radiance units were in turn converted to reflectance using scans of the BaSO4 white reference panel.

The hyperspectral measurement date was taken on 24th August and 28th September of 2007 in Heilongjiang and Zhejiang Province, respectively. The number of reflectance spectra of normal paddy rice were 12 and 35 in the two sites, respectively, and the number of reflectance spectra of lodged paddy rice were 10 and 35 in the two sites, respectively.

### 2.3 Hyperspectral Data Preprocessing

Hyperspectral reflectance was smoothed with a five-step moving average to suppress instrumental and environmental noise [27]. The hyperspectral reflectance represented the domains of 400 ~ 2 350 nm, and the missing segments corresponding to strong instrument and environment noise were not considered for further analysis [10].

The total datasets collected from the two paddy rice fields were divided into three different subsets, namely, one training dataset and two testing subsets. The training dataset (n=52) was used to calibrate the artificial neural network (ANN) classification models with three quarters of the canopy hyperspectral reflectance data collecting on 28th September of 2007, and one quarter was applied to validate the ANN classification models as the first testing dataset (n=18). While the hyperspectral reflectance data collecting on 24th August of 2007 was applied as the second testing dataset (n=22).

### 2.4 Analytical Techniques

Principal component analysis (PCA) has been used as a data compression technique for preserving total variance in transformation and minimizing mean square approximate errors [7]. This technique is very suitable for hyperspectra data for high dependence and autocorrelation in adjacent wavebands [10]. In this study, the front two principal component spectra (PCs) were derived from the canopy hyperspectral reflectance, which could explain over 98% of information variation.

Probabilistic neural networks (PNN) are a novel and composite neural network with radial basis function and competitive neural network. The neural architecture of PNN (Fig.1) consists of a radial basis layer and competitive layer with the exception of input layer [28]. In this study, PNN has been used to identify the lodged rice crop from normal ones at the canopy level. The neural nodes in the input layer were the broad- and narrow-band VIs and PCs. The competitive layer is also called as the

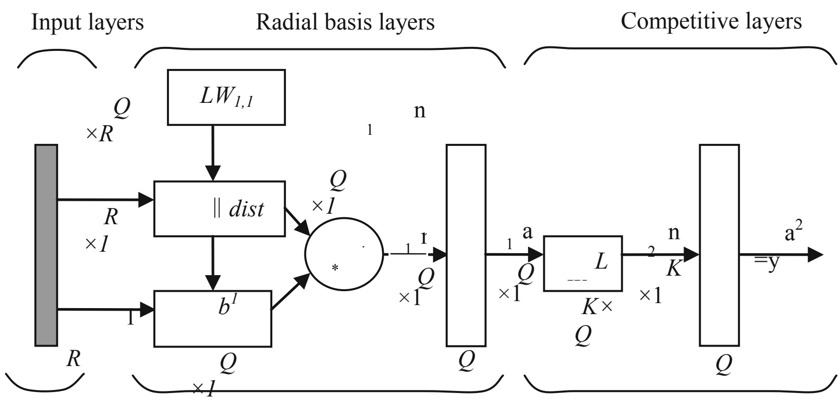


**Fig. 1.** Neural architecture for probabilistic neural networks (PNN)

output layer in practice. The neural nodes in the output layer were the classification of rice crops, viz. lodged and normal. The objective of the training process is examined to find the probabilistic mode between neural node of input layer and different classification of output layer. Additional detailed information on PNN can be found in the literature by Sivandm *et al.* [28].

The performance of VIs and PCA for classifying the lodged and normal paddy rice was evaluated with two statistical parameters, namely overall accuracy and Kappa coefficient [29].

PNN and PCA were conducted with Matlab functions in Matlab 7.0 (R14). Hyperspectral reflectance data preprocessing was made in ViewSpec Pro (V5.6.10) and VIs were derived from canopy hyperspectral reflectance in Microsoft Office Excel 2003.

# 3 Results and Discussions

## 3.1 Canopy Hyperspectral Reflectance

The spectral response properties of rice crop to disease and insect stresses were very significant for identifying lodged rice from normal ones in PPM using hyperspectral remote sensing technique. Fig.2 showed the canopy hyperspectral reflectance curves of normal rice and lodged rice caused by rice brown planthopper. Six wavelength intervals indicated the spectral differences (Table 2) because the photon transferring process was different between normal and lodged rice crop. Maybe it's because after the rice plant was damaged and lodged due to disease and insect stresses, which resulted in evident changes in the arrangement structure of rice panicles, leaves and stems.

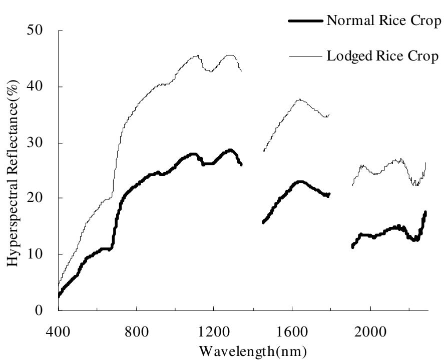


**Fig. 2.** Canopy hyperspectral reflectance of normal rice crop and lodged rice crop stressed by rice brown planthopper

As shown in Fig.2, the spectral curve shape of lodged rice crop was similar to that of the normal ones through the entire range (400~2350 nm), but the spectra amplitude was different. The hyperspectral reflectance of lodged rice crops increased about 75.7%, 70.9%, 79.2%, 62.3%, 64.7% and 76% in the blue-green (450~515 nm), green (525~605 nm), red (630~690 nm), near-infrared (775~900 nm), shortwave infrared (1550~1750 nm and 2090~2350 nm), respectively (Table 2).

**Table 2.** Average canopy hyperspectral reflectance of rice crops at six wavelength intervals representing the Enhanced Thematic Mapper Plus (ETM+) of Landsat-7 (Unit: %)

| Spectrum range (nm) | Site 1 (rice brown planthopper) | | Site 2 (rice panicle blast) | |
|---|---|---|---|---|
| | Healthy (n=35) | Lodged (n=35) | Healthy (n=12) | Lodged (n=35) |
| Blue-Green (450~515) | 5.9c | 10.4a | 3.3d | 7.6b |
| Green (525~605) | 9.6d | 16.4b | 10.9c | 17.2a |
| Red (630~690) | 11.4b | 20.5a | 8.2c | 18.1a |
| NIR (775~900) | 23.3c | 37.7b | 45.7a | 46.7a |
| $SWIR_1$ (1550~1750) | 22.0b | 36.2a | 20.5c | 41.6a |
| $SWIR_2$ (2090~2350) | 14.2c | 25.0b | 13.4c | 35.3a |

*: Reflectance values within a row followed by the same letter are not significantly different from each other by Duncan test at α = 0.05.

## 3.2 Results of Probabilistic Neural Network (PNN) Classifier

The overall accuracy and Kappa coefficient of broad-band and narrow-band spectral indices and PCA for discriminating lodged rice from normal ones based on PNN were

**Table 3.** Comparison of the two accuracy measures for VIs and PCA based on PNN (Unit: %)

| Dataset types | Training dataset | | Testing dataset 1 | | Testing dataset 2 | |
|---|---|---|---|---|---|---|
| Input vectors | OA | KC | OA | KC | OA | KC |
| DVI | 96.2% | 0.923 | 88.9% | 0.778 | 90.9% | 0.814 |
| RVI | 100% | 1 | 61.1% | 0.222 | 63.6% | 0.214 |
| NDVI | 88.5% | 0.769 | 72.2% | 0.444 | 63.6% | 0.214 |
| $SAVI_{L=0.5}$ | 94.2% | 0.885 | 55.6% | 0.111 | 68.2% | 0.319 |
| OSAVI | 94.2% | 0.885 | 66.7% | 0.333 | 72.7% | 0.421 |
| $MSAVI_2$ | 96.2% | 0.923 | 61.1% | 0.222 | 63.6% | 0.214 |
| TVI | 90.4% | 0.808 | 66.7% | 0.333 | 63.6% | 0.214 |
| IPVI | 94.2% | 0.885 | 66.7% | 0.333 | 63.6% | 0.214 |
| TDVI | 100% | 1 | 77.8% | 0.556 | 63.6% | 0.214 |
| RDVI | 94.2% | 0.885 | 66.7% | 0.333 | 59.1% | 0.108 |
| EVI | 94.2% | 0.885 | 66.7% | 0.333 | 72.7% | 0.421 |
| GEMI | 94.2% | 0.885 | 44.4% | -0.111 | 72.7% | 0.421 |
| BI | 100% | 1 | 100% | 1 | 77.3% | 0.56 |
| GVI | 86.5% | 0.731 | 88.9% | 0.778 | 81.8% | 0.621 |
| WI | 94.2% | 0.885 | 88.9% | 0.778 | 81.8% | 0.621 |
| NDWI | 92.3% | 0.846 | 44.4% | -0.111 | 63.6% | 0.214 |
| LSWI | 94.2% | 0.885 | 33.3% | -0.333 | 68.2% | 0.319 |
| NMDI | 84.6% | 0.692 | 72.2% | 0.444 | 27.3% | -0.467 |
| 2 PCs | 100% | 1 | 100% | 1 | 95.5% | 0.909 |

* OA and KC denoted the overall accuracy and Kappa coefficient, respectively.

shown in Table 3. PCA had the highest average overall accuracy (97.8%) and Kappa coefficient (95.5%) than the broad-band and narrow-band spectral indices for the two testing datasets, DVI and BI followed. Among the spectral indices, DVI had relative higher average overall accuracy (90.9%) and Kappa coefficient (0.796). Maybe it was because that DVI was more sensitive to changes of photosynthesis pigment (i.e. chlorophyll and carotenoid) of rice organs due to lodging caused by rice brown planthopper or rice panicle blast at the late growth stage.

## 4 Conclusions

The results of our study showed that the spectral indices based on distance from isoline to soil line such as DVI, BI, GVI and WI had higher overall accuracy and Kappa coefficient than those spectral indices based on ratios of spectral bands such as RVI, NDVI, SAVI series, and so on. Ratio-based spectral indices suppressed the spectral difference between different spectral bands, which made it more difficult to discriminate lodged rice from normal ones than distance-based spectral indices that amplified the spectral difference between different spectral bands [30]. The front several principal components (PCs) always accounted for most proportion of the variance of the original hyperspectral reflectance dataset [10], and retained more information than the broad-band and narrow-band spectral indices, and then PCs had the higher overall accuracy and Kappa coefficient than spectral indices.

The results indicated that hyperspectral remote sensing with PCA and ANNs has considerable potential in discriminating the lodging rice from normal ones as a supplementary and even alternative technique against spectral indices at the regional and even large spatial scales. However, there was only two kinds of stress status (i.e., healthy and lodging) in our research. Different lodging angles bring about different damage severity in practice. Future study is needed to include more stress status and extrapolate from ground to airborne and spaceborne platforms.

**Acknowledgements.** This research was supported by the China Postdoctoral Science Special Foundation Project (201003712), the Start Research and Fund Profect of Hangzhou Normal University (2011QDL23), the National Basic Research Program (973) of China (2010CB126200), the Agro-Industry R&D Special Fund of China (200903051), and the National Hi-Tech Research and Development Program (863) of China (2007AA10Z205).

## References

1. Everitt, J.H., Escobar, D.E., Summary, K.R., Davis, M.R.: Using airborne video, global positioning system, and geographical information system technologies for detecting and mapping citrus blackfly infestations. Southwest. Entomol. 19(2), 129–138 (1994)
2. McCartney, H.A., Fitt, B.D.L.: Dispersal of foliar fungal plant pathogens: mechanisms, gradients and spatial patterns. In: Gareth Jones, D. (ed.) Plant Disease Epidemiology, pp. 138–160. Kluwer Publishers, London (1998)

3. Pedigo, L.P.: Closing the gap between IPM theory and practice. J. Agri. Entomol. 12, 171–181 (1995)
4. Sōgawa, K.: The rice brown planthopper: feeding physiology and host plant interactions. Ann. Rev. Entomol. 27, 49–73 (1982)
5. West, J.S., Bravo, C., Oberit, R., Lemaire, D., Moshou, D., McCartney, H.A.: The potential of optical canopy measurement for targeted control of field crop diseases. Ann. Rev. Phytopathol. 41, 593–661 (2003)
6. Elvidge, C.D., Chen, Z.K.: Comparison of broad-band and narrow-band red and near-infrared vegetation indices. Remot. Sens. Environ. 54(1), 38–48 (1995)
7. Holden, H., LeDrew, E.: Spectral discrimination of healthy and non-healthy corals based on cluster analysis, principal component analysis, and derivative spectroscopy. Remot. Sens. Environ. 65(2), 217–224 (1998)
8. Karimi, Y., Prasher, S.O., Patel, R.M., Kim, S.H.: Application of support vector machine technology for weed and nitrogen stress detection in corn. Comput. Electron. Agr. 51(1-2), 99–109 (2006)
9. Shi, J.J., Liu, Z.Y., Zhang, L.L., Zhou, W., Huang, J.F.: Hyperspectral recognition of rice damaged by rice leaf roller based on support vector machine. Ric. Sci. 23(3), 331–334 (2009)
10. Liu, Z.Y., Wang, D.C., Li, B., Huang, J.F.: Discrimination of lodged rice based on visible/near-infrared spectroscopy. J. Infra. Milli. Wav. 28(5), 321–324 (2009)
11. Richardson, A.J., Everitt, J.H.: Using spectral vegetation indices to estimate rangeland productivity. Geocarto Int. 1, 63–77 (1992)
12. Jordan, C.F.: Derivation of leaf area index from quality of light on the forest floor. Ecol. 50(4), 663–666 (1969)
13. Rouse, J.W., Haas, R.H., Schell, J.A., Deering, D.W.: Monitoring vegetation systems in the great plains with ERTS. In: 3rd ERTS Symposium, NASA, Washington, USA, pp. 48–62 (1973)
14. Huete, A.R.: A soil-adjusted vegetation index (SAVI). Remot. Sens. Environ. 25(3), 295–309 (1988)
15. Rondeaux, G., Steven, M., Baret, F.: Optimization of soil-adjusted vegetation indices. Remot. Sens. Environ. 55(2), 95–107 (1996)
16. Qi, J., Chehbouni, A., Huete, A., Kerr, Y., Sorooshian, S.: A modified soil-adjusted vegetation index (MSAVI). Remot. Sens. Environ. 48(2), 119–126 (1994)
17. Broge, N.H., Leblanc, E.: Comparing prediction power and stability of broadband and hyperspectral vegetation indices for estimation of green leaf area index and canopy chlorophyll density. Remot. Sens. Environ. 76(2), 156–172 (2000)
18. Crippen, R.E.: Calculating the Vegetation Index Faster. Remot. Sens. Environ. 34(1), 71–73 (1990)
19. Bannari, A., Asalhi, H., Teillet, P.M.: Transformed difference vegetation index (TDVI) for vegetation cover mapping. In: Proceedings on CD-Rom, paper I2A35, International Geoscience and Remote Sensing Symposium, Toronto, Ontario, Canada, p. 1508 (2002)
20. Roujean, J.L., Boreon, F.M.: Estimating PAR absorbed by vegetation from bidirectional reflectance measurements. Remot. Sens. Environ. 51(3), 375–384 (1995)
21. Liu, H.Q., Huete, A.R.: A feedback based modification of the NDVI to minimize canopy background and atmospheric noise. IEEE Transaction on Geosci. Remot. Sens. 33, 457–465 (1995)
22. Pinty, B., Verstraete, M.M.: GEMI: A non-linear index to monitor global vegetation from satellites. Vegetatio. 101(1), 15–20 (1992)

23. Crist, E.P., Laurin, R., Cicone, R.C.: Vegetation and soils information contained in transformed thematic mapper data. In: International Geoscience and Remote Sensing Symposium (IGARSS 1986), Zurich, Switzerland, pp. 1465–1470 (1986)
24. Gao, B.C.: NDWI: A normalized difference water index for remote sensing of vegetation liquid water from space. Remot. Sens. Environ. 58(3), 257–266 (1996)
25. Xiao, X., Boles, S., Frolking, S., Salas, W., Moore III, B., Li, C.: Observation of fl ooding and rice transplanting of paddy rice fields at the site to landscape scales in China using VEGETATION sensor data. Int. J. Remot. Sens. 23(15), 3009–3022 (2002)
26. Wang, L., Qu, J.: NMDI: A Normalized Multi-band Drought Index for Monitoring Soiland Vegetation Moisture with Satellite Remote Sensing. Geophy. Research Lett. 34, 204–208 (2007)
27. Kobayashi, T., Kanda, E., Kitada, K., Ishiguro, K., Torigoe, Y.: Detection of rice panicle blast with multispectral radiometer and the potential of using airborne multispectral scanners. Phytopathol. 91(3), 316–323 (2001)
28. Sivandm, S.N., Sumathi, S., Deepa, S.N.: Introduction to neural networks using Matlab 6.0. Tata Mcgraw-Hill Publishing Company Limited (2006)
29. Congalton, R.G.: A review of assessing the accuracy of classification of remotely sensed data. Remot. Sens. Environ. 37(1), 35–46 (1991)
30. Asrar, G.: Theory and applications of optical remote sensing, pp. 119–125. John Wiley & Sons, Inc., New York (1989)