Shafii et al., Int J Appl Exp Math 2016, 1: 106
http://dx.doi.org/10.15344/ijaem/2016/106

International Journal of
Applied & Experimental Mathematics

Research Article Open Access

Application of Bayesian Methods for Assessing Detection Accuracy in Remote Sensing

Bahman Shafii* and William J. Price

Statistical Programs, P.O. Box 442337, University of Idaho, Moscow, ID, 83844-2337, USA

Abstract

Digital imagery and remote sensing have become popular and accessible tools in many scientific research fields. Accuracy of classification, the degree of agreement between classification and ground truth, is traditionally quantified by an error matrix and summarized using agreement measures such as Cohen's kappa. The kappa statistic, however, can be shown to be a transformation of the marginal proportions commonly referred to as omissional and commissional error rates. Alternative estimation methods for these agreement measures include binomial, bootstrap and Bayesian techniques. In this study, we develop a Bayesian estimation method for omissional and commissional errors and discuss its utilization in variability assessment and inference. We will also show how additional or sequential information may be incorporated to improve the estimation situation. Techniques are demonstrated using previously published data.

Publication History:

Received: June 10, 2016
Accepted: August 04, 2016
Published: August 06, 2016

Keywords:

Bayesian estimation, Agreement measures, Error rates

Introduction

In remote sensing, accuracy of classification is traditionally assessed by the comparison of classified pixels with ground truth using agreement measures such as Cohen's kappa. Conventionally, statistical inferences concerning the agreement measures have been based on asymptotic normality assumptions. While asymptotic methods may produce satisfactory results in certain instances, they fail to account for the underlying distribution of the classified data. This can result in poor and inconsistent inferences regarding the classification accuracy.

Accuracy of classification is often represented in the form of an error matrix [1, 2]. The rows of this table (i=1, 2, 3, ..., C) represent the computer or human generated classification and the columns (j=1, 2, 3, ..., C) denote the reference or ground truth categories:

Ground Truth

	1	2	3	...	C	
1	x11	x12	x13	...	x 1c	N1.
2	x21	x22	x23	...	x 2c	N2.
3	x31	x32	x33	...	x 3c	N3.
⋮	⋮	⋮	⋮	⋮	⋮	⋮
C	xc1	xc2	xc3	...	xcc	NC.
	N.1	N.2	N.3	...	N.C	N

where, x_{ii} is the number of pixels correctly classified in category i, $N_{i.}$ and $N_{.i}$ are the corresponding marginal totals for classification and ground truth, respectively, and $N = \Sigma N_{i.} = \Sigma N_{.i}$.

Various methods have been suggested for assessing the degree of ground truth agreement for each category. Common measures include conditional kappa, a general index of agreement [3]:

$$\hat{\kappa}_I = \frac{\left(x_{ii}/N_{.i} - N_{i.}/N \right)}{\left(1 - N_{i.}/N \right)},$$

the omissional error rate, measuring the proportion of pixels incorrectly omitted from a classification:

$$\hat{O}_I = 1 - x_{ii}/N_{.i},$$

and the commissional error rate, measuring the proportion of pixels erroneously committed to a classification category [4]:

$$\hat{C}_I = 1 - x_{ii}/N_{i.},$$

where $x_{ii}/N_{.i}$ and $x_{ii}/N_{i.}$ are commonly referred to as producer's and user's accuracies [5].

The kappa statistic has been suggested as a means of assessing the degree of agreement in remotely sensed data because it equally weighs both omissional and commissional errors [6]. Remote sensing, however, presents a unique situation for conditional kappa in which, for a given image classification, the marginal ground truth totals, $N_{.i}$, as well as classified totals for each class, $N_{i.}$, are constant. Under these conditions, (1) becomes a simple monotonic function of the omissional error rate [7]. It should also be noted that, although kappa treats misclassifications equally, in many cases it may be important to distinguish between the error types [8]. For these reasons, it will be more advantageous to carry out accuracy assessment based on the later two measures, namely $\hat{O}_I$ and $\hat{O}_I$

Using Bayesian estimation and maximum entropy, Shafii et al, [9] developed a methodology that may be used for estimation and inference regarding the aforementioned agreement measures. Furthermore, it has been shown that the Bayesian estimation technique is superior to that of the exact binomial, and that due to its ability to incorporate prior information, it can be more advantageous than the parametric bootstrapping technique [10].

In this paper, we review the Bayesian estimation technique for omissional and commissional error rates and illustrate its inferential use for image variability assessment and comparison. Furthermore,

*Corresponding Author: Dr. Bahman Shafii, Statistical Programs, P.O. Box 442337, University of Idaho, Moscow, ID, 83844-2337 USA, E-mail: bshafii@uidaho.edu

Citation: Shafii B, Price WJ (2016) Application of Bayesian Methods for Assessing Detection Accuracy in Remote Sensing. Int J Appl Exp Math 1: 106. doi: http://dx.doi.org/10.15344/ijaem/2016/106

Citation: Shafii B, Price WJ (2016) Application of Bayesian Methods for Assessing Detection Accuracy in Remote Sensing. Int J Appl Exp Math 1: 106. doi: http://dx.doi.org/10.15344/ijaem/2016/106

we will demonstrate how additional information may be utilized to improve the estimation situation.

Methods

If the area and location being imaged are held constant, the marginal totals for ground truth, $N_{.i}$ are fixed and the diagonal elements of the error matrix, x_{ii}, can be considered as binomial variates:

$$\Phi_i = \binom{N_{.i}}{x_{ii}} \mathrm{p}_{ii}^{\,x_{ii}}(1-p_{ii})^{N_{.i}-x_{ii}} \tag{4}$$

where x_{ii} and $N_{.i}$ are as given above, and p_{ii} is the true proportion of correctly classified pixels. The Bayesian perspective for ($\hat{O}_i$) may then be developed using (4) as a likelihood and assuming a prior distribution for pii. Using a constant non-informative prior [9], $\pi(p_{ii})$, the posterior distribution for $p_{ii}|x_{ii}$ becomes:

$$\pi\left(p_{ii}|x_{ii}\right) \propto \pi(p_{ii})\cdot\Phi_i = A\cdot\Phi_i \tag{5}$$

The omissional error rate given in (2) may then be formed as a monotonic function of x_{ii} and the distribution of ($\hat{O}_l$) may be derived from (4) and (5) using the following transformation:

$$\Phi_i = p(x_{ii}=b) = p\left(1-{x_{ii}}/{N_{.i}}\right) = p\left(1-{b}/{N_{.i}}\right) = p(\hat{O}_l = b') \tag{6}$$

where b and b' are constant values.

The commissional error rate is a function of binomial variates involving sums and ratios:

$$\hat{C}_l = 1-{x_{ii}}/{N_{i.}} = 1-\frac{x_{ii}}{x_{ii}+\sum x_{ii}} \tag{7}$$

where Σx_{ij} is the sum of the i^{th} row elements over j, excluding the case i=j. The xij are independent and distributed as:

$$\Phi_{ij} = \binom{N_{.j}}{x_{ij}} p_{ij}^{\,x_{ij}}(1-p_{ij})^{N_{.j}-x_{ij}} \tag{8}$$

Analytical derivation of a posterior distribution for (7) is troublesome, however a numerical derivation is possible using posterior distributions based on (4) and (8). To simplify the computations, an initial distribution for the inverse of x_{ii} / $N_{i.}$ given by:

is generated. The resulting values are subsequently reinverted to obtain the final distribution. It should be noted, however, that this solution is restricted to $x_{ii} > 0$. This does not pose a problem as this is a degenerate case where inferential evaluation of $\hat{C}_l$ (or $\hat{O}_l$) would not be meaningful. Estimates, moments, and probability intervals may then be derived using these posterior distributions.

Methods for pair-wise comparisons of independent estimates of error rates can also be developed. For example, let $\omega(\hat{O}_k)$and $\varphi(\hat{O}_l)$ be the posterior distributions for omissional errors $\hat{O}_k$ and $\hat{O}_l$, respectively. Then the joint distribution of $\hat{O}_k$ and $\hat{O}_l$ under independence is defined as $\tau(\hat{O}_k, \hat{O}_l) = \omega(\hat{O}_k)\,\varphi(\hat{O}_l)$ and the distribution of the difference, $m(\hat{d}_{kl})$; $\hat{d}_{kl} = \hat{O}_k - \hat{O}_l$, is given by a transformation of variables:

$$m\left(\hat{d}_{kl}\right) = \int\tau\left(\hat{O}_k, -\hat{d}_{kl}\right)\delta_{\hat{O}_l} = \int\omega\left(\hat{O}_k\right)\varphi(\hat{O}_k - \hat{d}_{kl})\,\delta_{\hat{O}_l} \tag{9}$$

The posterior distributions given above and their associated inferential results may be derived through numerical integration and interpolation.

In the presence of auxiliary or sequential classification data for a common area or image, the Bayesian paradigm can provide a mechanism for incorporating this additional information into the estimation process. If the underlying binomial parameters of the new and the preceding data matrices can be considered equivalent and generated from the same process, the posterior distribution derived for the original data set may be considered as a prior distribution for the new data and the posterior distribution in (5) augmented to become:

$$\pi(p_{ii}\mid x_{ii}) \;\propto= A\cdot\Phi_k\cdot\Theta_l \tag{10}$$

where $A\cdot\Phi_k$ is the previous posterior distribution defined in (5) and Θ_l is the binomial likelihood (4) assumed for the new data. The term $\Phi k\cdot\Theta l$ can be rewritten as:

$$\Phi^*_i = \binom{N^*_{.i}}{x^*_{ii}} P_{ii}^{x^*_{ii}}(1-p_{ii})^{N^*_{.i}-x^*_{ii}} \tag{11}$$

where $N_{.i}^* = N_{.i} + N_{.i}'$ is the sum of the marginal ground truth totals for the original and new data, respectively, and $x_{ij}^* = x_{ii} + x_{ii}'$ is the corresponding sum of correctly classified pixels. This new likelihood is then used to update the posterior distributions for the specified error rates.

All estimations and computations were carried out using custom C codes and SAS (2012). Program codes are available from the authors at: http://webpages.uidaho.edu/cals-statprog/IJAEM2016/index.html.

Demonstration

The data used for the purpose of demonstration were taken from Congalton et al. [11]. Two photo-interpreters of equal skill were employed to evaluate the same aerial photographs of a forested area. The results of the classification are summarized in two error matrices given in Tables 1a and 2a. Shafii et al. [9] have used this data to demonstrate the Bayesian estimation for conditional kappa (1) as well as procedures by which to compare the estimated conditional kappas for a given category between the two interpreters. Here, we will concentrate our effort on the Bayesian estimation and comparison of omissional and commissional errors for each of the interpreters, individually and combined.

The estimation results for the first photo-interpreter (I) are provided in Table 1. b. Accuracy of classification was relatively high for both pine and oak categories as indicated by the estimated values of the omissional (.3396 and .4063, respectively) or commissional (.4262 and .3968, respectively) error rates [12]. Whereas the classification accuracy for the cedar category was poor ($\hat{O}_2$=.7180) to moderate ($\hat{C}_2$=.3889), and that of cottonwood was poor overall ($\hat{O}_4$=.7143, and $\hat{C}_4$=.9048). The 95% probability limits for the above estimates suggested a reasonable estimation for all the categories except for cottonwood where the upper bounds on both omissional and commissional errors approached 1 (complete misclassification). The posterior probability distributions of omissional and commissional error rates for all the specified categories for interpreter I are given in Figure 1. Note that in both cases, the probability distributions of oak and pine categories appeared to be symmetrical. However, probability distributions for cottonwood were skewed and dispersed. This constitutes a case, i.e. small sample size and asymmetrical distribution for which inferences based on asymptotic results should be avoided.

Suppose that now additional data are provided from a second interpreter of the same aerial photograph. The estimation results for this interpreter (II) are provided in Table 2. b. Conclusions similar to those of interpreter I can be drawn regarding the classification accuracy of the four categories. Again, pine and oak had reasonable

Citation: Shafi B, Price WJ (2016) Application of Bayesian Methods for Assessing Detection Accuracy in Remote Sensing. Int J Appl Exp Math 1: 106. doi: http://dx.doi.org/10.15344/ijaem/2016/106

accuracy of detection, whereas the detection accuracies for cedar and cottonwood were poor to moderate. Posterior probability distributions of pair-wise differences (for both error types) may be utilized to assess the similarity of detection accuracy between the two interpreters. As an example, the distributions of pair-wise differences for omissional error rates in pine and cottonwood categories were centered on and encompassed zero (Figure 2). This indicates statistical similarity of these agreement measures between the two interpreters. The same was true for all the other categories and error types.

Given the similarity of additional data provided by interpreter II, it was feasible to consider a combined estimation of the omissional and commissional error rates. The combined error matrix is given in Table 3a, and the corresponding estimation results are provided in Table 3b. The interpretation of results remained relatively unchanged as compared to that made for interpreter I. The pine and oak categories showed acceptable detection accuracies, while the accuracies for the cedar and cottonwood categories were poor to moderate. Although the estimated values of omissional and commissional error rates

		Interpreter I Ground Truth				
		1	2	3	4	Total
Classified	1	35	14	11	1	61
	2	4	11	3	0	18
	3	12	9	38	4	63
	4	2	5	12	2	21
	Total	53	39	64	7	163

1 = Pine 2 = Cedar 3 = Oak 4 = Cottonwood

Category	$\hat{O}_i$	S.E.	LL	UL
Pine	0.3396	0.0635	0.2271	0.4758
Cedar	0.7180	0.0702	0.5612	0.8341
Oak	0.4063	0.0600	0.2947	0.5290
Cottonwood	0.7143	0.1491	0.3495	0.9154

Category	$\hat{C}_i$	S.E.	LL	UL
Pine	0.4262	0.0329	0.3524	0.5326
Cedar	0.3889	0.0473	0.2796	0.6590
Oak	0.3968	0.0327	0.3186	0.4991
Cottonwood	0.9048	0.0146	0.8241	0.9882

Table 1: Error matrix and Bayesian estimates of omissional, $\hat{O}_i$, and commissional, $\hat{C}_i$ errors, along with their corresponding standard errors (S.E.) and lower (LL) and upper (UL) 95% percentile limits for interpreter I computed empirically from the estimated probability density.

		Interpreter II Ground Truth				
Classified		1	2	3	4	Total
	1	32	15	5	3	55
	2	7	8	5	0	20
	3	7	8	38		2
	4		6	2	7	15
	Total	52	38	63	6	159

1 = Pine 2 = Cedar 3 = Oak 4 = Cottonwood

Category	$\hat{O}_i$	S.E.	LL	UL
Pine	0.3846	0.0657	0.2654	0.5213
Cedar	0.7895	0.0652	0.6351	0.8891
Oak	0.3968	0.0603	0.2857	0.5210
Cottonwood	0.8333	0.1443	0.4211	0.9634

Category	$\hat{C}_i$	S.E.	LL	UL
Pine	0.4182	0.0332	0.3390	0.5263
Cedar	0.6000	0.0434	0.4639	0.8122
Oak	0.3091	0.0323	0.2351	0.4277
Cottonwood	0.9655	0.0065	0.9112	0.9992

Table 2: Error matrix and Bayesian estimates of omissional, $\hat{O}_i$, and commissional, $\hat{C}_i$ errors, along with their corresponding standard errors (S.E.) and lower (LL) and upper (UL) 95% percentile limits for interpreter II computed empirically from the estimated probability density.

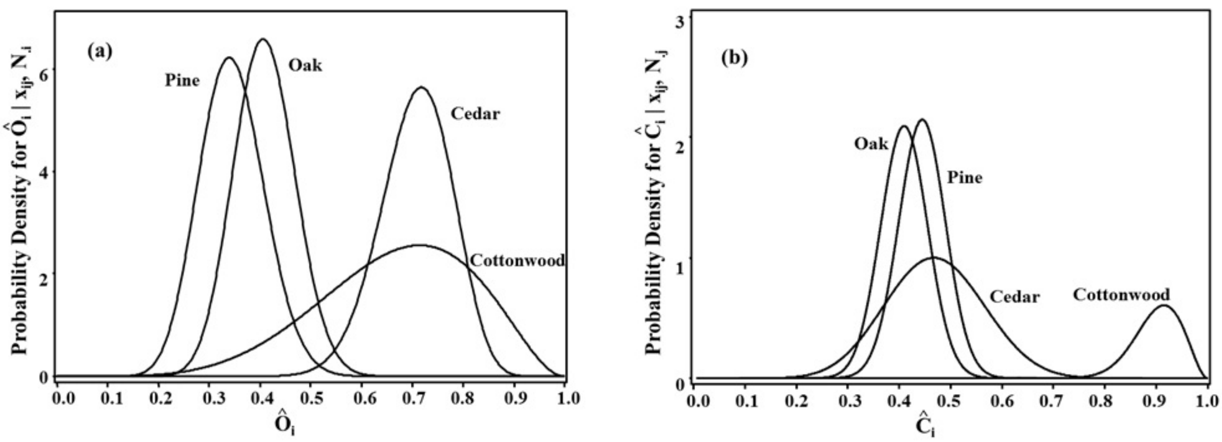

Figure 1: Posterior probability distributions of omissional (a) and commissional (b) errors associated with the four forest categories for interpreter I.

Citation: Shafii B, Price WJ (2016) Application of Bayesian Methods for Assessing Detection Accuracy in Remote Sensing. Int J Appl Exp Math 1: 106. doi: http://dx.doi.org/10.15344/ijaem/2016/106

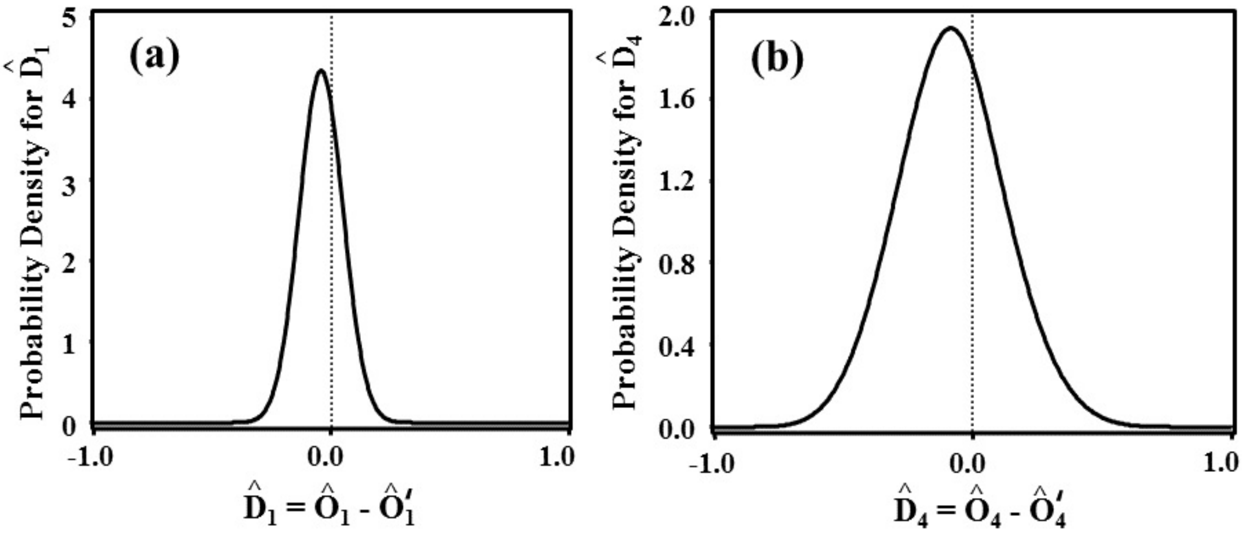

Figure 2: Posterior probability distributions of pair-wise differences, $\hat{D}_i$, in omissional error rates between the two interpreters for pine (a) and cottonwood (b) categories.

Classified	Combined Ground Truth					
		1	2	3	4	Total
	1	67	29	16	4	116
	2	11	19	8	0	0
	3	19	17	76	6	6
	4		8	27	3	12
	Total	105	77	127	13	322

1 = Pine 2 = Cedar 3 = Oak 4 = Cottonwood

Category	$\hat{O}_i$	S.E.	LL	UL
Pine	0.3396	0.0635	0.2271	0.4758
Cedar	0.7180	0.0702	0.5612	0.8341
Oak	0.4063	0.0600	0.2947	0.5290
Cottonwood	0.7143	0.1491	0.3495	0.9154

Category	$\hat{C}_i$	S.E.	LL	UL
Pine	0.4262	0.0329	0.3524	0.5326
Cedar	0.3889	0.0473	0.2796	0.6590
Oak	0.3968	0.0327	0.3186	0.4991
Cottonwood	0.9048	0.0146	0.8241	0.9882

Table 3. Error matrix and Bayesian estimates of omissional, $\hat{O}_i$, and commissional, $\hat{O}_i$ errors, along with their corresponding standard errors (S.E.) and lower (LL) and upper (UL) 95% percentile limits for both interpreters combined as computed empirically from the estimated probability density.

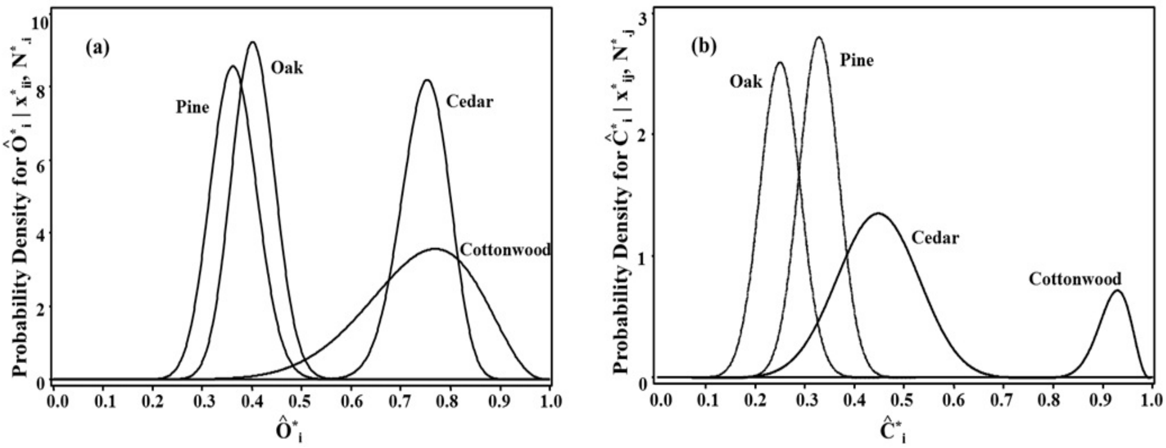

Figure 3: Posterior probability distributions of omissional (a) and commissional (b) errors associated with the four forest categories for both interpreters combined.

were comparable for most of the categories, the estimated standard errors were smaller in all cases. This interpretation is evident from the posterior probability distributions of omissional and commissional error rates for both interpreters combined (Figure 3). Compared to Figure 1, while the magnitude of the most probable values for all categories and both error types remained relatively unchanged, the corresponding distributions appeared narrower and more concentrated on the specified estimated values. This indicates a more precise estimation of the error rates as a result of incorporating the additional information into the estimation situation.

Conclusion

The Bayesian approach provides a sound mechanism for assessing detection accuracy in remote sensing. It avoids asymptotic normality requirements and produces more reliable results based on correct distributional and mathematical assumptions. Since remote sensing applications often utilize images taken sequentially in time, the Bayesian methodology can provide a means of incorporating such additional information into the estimation process, and thereby, improve the reliability and precision of the agreement measures.

Competing Interests

The authors declare that they have no competing interests.

Author Contributions

Both the authors substantially contributed to the study conception and design as well as the acquisition and interpretation of the data and drafting the manuscript.

References

1. Card DH (1982) Using map category marginal frequencies to improve estimates of thematic map accuracy. Photogramm Eng Remote Sens 49: 431-439.
2. Congalton RG, Mead RA (1983) A qualitative method to test for consistency and correctness in photointepretation. Photogramm Eng Remote Sens 49: 69-74.
3. Light RJ (1971) Measures of response agreement for qualitative data: Some generalizations and alternatives. Psychological Bull 76: 365-377.
4. Aronoff S (1982) Classification accuracy: A user approach. Photogram Eng Remote Sens 48: 1299-1307.
5. Congalton RG (1991) A review of assessing the accuracy of classifications of remotely sensed data. Remote Sens Environ 37: 35-46.
6. Rosenfield GH, Fitzpatric-Lins K (1986) A coefficient of agreement as a measure of thematic classification accuracy. Photogramm Eng Remote Sens 52: 223-227.
7. Wackerly DD, McClave JT, Rao PV (1978) Measuring nominal scale agreement between a judge and a known standard. Pyschmetrika. 43: 213-223.
8. Story M, Congalton RG (1986) Accuracy assessment: A user's perspective. Photogramm Eng Remote Sens 52: 397-399.
9. Shafii B, Price WJ, Lass LW, Thill DC (1998) Assessing Variability Of Agreement Measures In Remote Sensing Using A Bayesian Approach. In: proceedings of the American Statistical Association, Section on Bayesian Statistical Science 22-27.
10. Shafii B, Price WJ (2000) A comparison of three estimation methods for assessing the agreement of classified images with ground truth. In: proceedings of the American Statistical Association, Section on Statistical Epidemiology 62-67.
11. Congalton RG, Oderwald R, Mead RA (1983) Assessing landsat classification accuracy using discrete multivariate analysis statistical techniques. Photogramm Eng Remote Sens49: 1671-1678.
12. SAS Institute Inc (2012) SAS Language: Reference, Version 9.4. SAS Institute Inc.