Applying Remote Sensing Image Classification Techniques and Area Estimation Methods for Land Use Purposes

Dr. Tarek Eldamaty

Faculty of Engineering & Islamic Architecture – Umm Al-Qura University, KSA

Abstract: ***In this paper, two main techniques of area estimation for land use derived from remotely sensed data have been discussed. The first is conventional technique which refers to pixel counting estimator. The other technique is based confusion matrix area estimators including direct estimator, inverse estimator, additive estimator, bias removal technique and map marginal proportion based estimator. To evaluate the performance of these area estimator techniques a comparative study has been done using three evaluation criteria quantitative error, bias and dispersion. The results show that, all areas which are estimated by based confusion matrix area estimators are more accurate and closer to the true areas than that by conventional technique. Direct and additive estimators produce accurate and consistent results. The map marginal proportion based estimator and inverse estimator also produce accurate results in cases of large size testing samples.***

Keywords: Image classification – Area estimation – Land use – Remote sensing

Abbreviation: *N_{i+} = population number of pixels classified into class i by the classifier (known); N_{+j} = population number of pixels from reference in class j (unknown); N = total number of the pixels in the population (whole scene), (known); n_{ij} = sample number of pixels classified in class i by the classifier, from reference found in class j (known); n_{i+} = sample number of pixels classified into class; n_{+j} = sample number of pixels from reference for class j; n = total number of pixels in the sample (sample size). p_{i+} = the sample marginal posteriori probability for class I; q_{+j} = the sample marginal a priori probability for class j; T = Reference numbers of pixels for a class (unknown); M = Total pixels classified to a class (known), obtained from classified image; A_c = Area estimated by pixel counting for a class c (m^2); nc = Total number of pixels classified for a class c; N = Total number of pixels in the region; A_T = Total area of image (m^2).*

1. Introduction

Accurate area estimation is one of the important applications of remote sensing for different studies as crop areas or forest management strategies etc. The usual approach of classifying all pixels and counting or proportioning the pixels per class are rather inaccurate that it is often necessary to make calibrations on the direct counts in order to obtain better estimates for the marginal areas (Dymond , 1992, Schriever and Congalton, 1995). The calibrations of the marginal area estimates are based on the utilization of the sample confusion matrices. In this paper conventional technique and based confusion matrix area estimators have been studied.

Pixel counting as an area estimator is often proposed in remote sensing projects run by the private sector for public administrations, mainly in developing countries. The estimates are acceptable only if spectral signatures are clearly discriminated and image classification is very accurate (Gallego, 2004). However, because of classification error, the area derived from pixel counting is usually biased (Gallego, 2004, Stehman, 2005). Because the pixel counting is based on a complete census of the region, the bias of this pixel count area is viewed as a “measurement bias” rather than as an “estimator bias” (Stehman, 2005). A confusion matrix provides the classification error information that allows for adjusting the area obtained from pixel counting to account for this measurement bias (Stehman, 2009).

The direct estimator of the marginal areas has been suggested by Card (1982) for image classification and it has been theoretically compared with the inverse estimator by Jupp (1989). Also direct estimator was called "inverse estimator" by Czaplewski and Catts (1992) and Walsh and Burk (1993). The direct estimator uses the sample commission (a posteriori) probability matrix to estimate the population commission probability matrix.

The inverse estimator is based on the inverse of the sample omission probability matrix therefore it is called the inverse correction method. On another hand Czaplewski and Catts (1992) called it as "classic correction method". This method had been widely used in remote sensing applications (Prisley and Smith, 1987, Hay, 1988, Yuan, 1997).

The additive estimator, has been suggested by Dymond (1992) is based on the idea that the true number of pixels in a class can be estimated by the classified number of pixels in the same class plus a correction term. The parameters required are the off-diagonal elements of error matrix.

Bias Removal Technique is based on the assumption that if error of commission and error of omission are equal then area estimates are accurate that is because the main source of error when crisp or soft classified remotely sensed images are used for area estimation by pixel or proportion counting are the misclassification of pixels. Therefore, these area estimates are biased. The bias of pixel or proportion counting estimator is approximately obtained from the difference between the commission and omission error (Gallego, 2004).

Map marginal proportions based estimator approach adopted for estimating the area of land cover classes by modifying of the approach was suggested by Card (1982).

2. Methods and Materials Dataset

In this paper two classified remotely sensed images with five land cover categories have been used. First image in this study is from Indian Remote Sensing (IRS) satellite LISS II sensor at spatial resolution of 36 x 36 m used as classified map. Second image at fine spatial resolution (6 m) obtained from IRS 1C PAN sensor has been used (with topographical maps or topo-sheet (number 53G/13) at 1:50,000 scale ,1973, and existing field surveyed map at 1:1000 scale ,1992) as reference map. These two images are classified by maximum likelihood classifiers (MLC) by Ibrahim (2004). Since the purpose of this study is not on classification accuracy, assuming one map as reference will not bias any following result. In this way a clear idea of the exact area proportion of each class in reference map is available. The class categories and their proportion on both maps are listed in table (1)

Table 1: Class categories and their proportion in hard classification

Land cover type	Proportion in reference map (%)	Proportion in classified map (%)	Difference (%)
Built-up land	23.7	22	1.7
Grass land	28.6	38.8	10.2
Trees	29.7	24	5.7
Agriculture	8. 5	9.8	1.3
Barren land	9.6	5.4	4.2

Sampling Scheme and Sampling Fraction

In this study two sampling schemes are used simple random sampling and systematic sampling. Input hard classified and reference data were sampled using these two schemes with fraction ranged from 0.1 to 1 of whole population by an increment of 0.1 to construct error matrices for every sampling fraction. These error matrices have been used by estimators to estimate different classes' area.

Models of Conventional and Based Confusion Matrix Estimators

Before a description of each technique is given, some notations that will be used subsequently in each technique have been described. These notations are similar to those used by Yuan (1997). Two types of error matrices shall be used: population error matrix (Cp) and sample error matrix (Cs). Another two matrices are dependent on (Cp and Cs) may be used (Ps) and (Qs) as the sample commission and omission probability matrices.

The population error matrix (Cp) and sample error matrix (Cs) can be defined as,

$$Cp = \begin{pmatrix} N_{11} & N_{12} \cdots\cdots & N_{1r} \\ N_{21} & N_{22} \cdots\cdots & N_{2r} \\ \vdots & \vdots & \vdots \\ N_{r1} & N_{r2} & N_{rr} \end{pmatrix}$$

$$Cs = \begin{pmatrix} n_{11} & n_{12} \cdots\cdots & n_{1r} \\ n_{21} & n_{22} \cdots\cdots & n_{2r} \\ \vdots & \vdots & \vdots \\ n_{r1} & n_{r2} & n_{rr} \end{pmatrix}$$

Let's define:

$$N_{+j} = \sum_{i=1}^{r} N_{ij} \qquad N_{i+} = \sum_{j=1}^{r} N_{ij}$$

$$N = \sum_{i=1}^{r} N_{+j} = \sum_{j=1}^{r} N_{i+} = \sum_{i=1}^{r}\sum_{j=1}^{r} N_{ij}$$

$$n_{+j} = \sum_{i=1}^{r} n_{ij} \qquad n_{i+} = \sum_{j=1}^{r} n_{ij}$$

$$n = \sum_{i=1}^{r} n_{+j} = \sum_{j=1}^{r} n_{i+} = \sum_{i=1}^{r}\sum_{j=1}^{r} n_{ij}$$

Where,
N_{ij} =population number of pixels being classified in class i by the classifier, from reference found in class j (unknown).
N_{i+} = population number of pixels classified into class i by the classifier (known).
N_{+j} = population number of pixels from reference in class j (unknown).
N = total number of the pixels in the population (whole scene), (known).
n_{ij} = sample number of pixels classified in class i by the classifier, from reference found in class j (known).
n_{i+} = sample number of pixels classified into class.
n_{+j} = sample number of pixels from reference for class j. and
n = total number of pixels in the sample (sample size).

The sample commission and omission probability matrices Ps and Qs can be defined based on the sample error matrix as,

$$P_s = \begin{pmatrix} p_{11} & p_{12} \cdots\cdots & p_{1r} \\ p_{21} & p_{22} \cdots\cdots & p_{2r} \\ \vdots & \vdots & \vdots \\ p_{r1} & p_{r2} & p_{rr} \end{pmatrix} \quad Q_s = \begin{pmatrix} q_{11} & q_{12} \cdots\cdots & q_{1r} \\ q_{21} & q_{22} \cdots\cdots & q_{2r} \\ \vdots & \vdots & \vdots \\ q_{r1} & q_{r2} & q_{rr} \end{pmatrix}$$

Where,
$p_{ij} = n_{ij} / n_{i+}$ $\qquad q_{ij} = n_{ij} / n_{+j}$
$p_{i+} = n_{i+} / n$ $\qquad q_{+j} = n_{+j} / n$
p_{ij}: is the sample commission (a posteriori) probability for a pixel being classified as in class i given the condition that it is from class j,
q_{ij}: is the sample omission (a priori) probability for a pixel being from class j given the condition that it is classified as in class i,
p_{i+}: is the sample marginal posteriori probability for class i
q_{+j}: is the sample marginal a priori probability for class j.

To simplify the above notations, two vectors may be defined as:

$$M = (N_{1+}, N_{2+}, \ldots, N_{r+})^T$$
$$T = (N_{+1}, N_{+2}, \ldots, N_{+r})^T$$

Where,
T- Reference numbers of pixels for a class (unknown).
M - Total pixels classified to a class (known), obtained from classified image.

The task of an area estimation technique is to obtain accurate area estimates of T based on knowledge of *M* and the sample error matrix *Cs*. All the error matrix based area estimation techniques require this information as the input. The population omission and commission probability matrices P_p

and Q_p are defined based on the population error matrix (Cp).

Pixel Counting Estimator

In pixel counting estimator, pixels for each land cover are counted and used directly for area estimation. In hard classification, pixel counting estimator is given by Gallego (2004):

$$A_c = \frac{n_c A_T}{N}$$

Where,
A_c: Area estimated by pixel counting for a class c (m²).
nc –Total number of pixels classified for a class c
N: Total number of pixels in the region.
A_T: Total area of image (m²).

Direct Estimator

The direct estimator uses the sample commission (a posteriori) probability matrix to estimate the population commission probability matrix. The direct estimator for the area estimation can be given by Yuan (1997):

$$T^{(d)} = (N_1^{(d)}, N_2^{(d)}, \ldots\ldots\ldots, N_r^{(d)})^T$$

In matrix form,

$$T^{(d)} = Ps\ M$$

$T^{(d)}$ is multiplied with the area of a pixel to get area estimates in terms of m^2.

Inverse Estimator

Inverse estimator can be given by Hay (1988):

$$T^{(i)} = (N_1^{(i)}, N_2^{(i)}, \ldots\ldots\ldots, N_r^{(i)})^T$$

In matrix form

$$M = Q_s^T\, T^{(i)}$$

The inverse estimates ($T^{(i)}$) can be obtained by inverting Q_s^T if its inverse exists. The error matrix used in accuracy assessment is $r\ x\ r$ so that it is generally invertible. But if diagonal entry for any class is zero then that matrix may not be invertible, and therefore, the zero entries may have to be substituted with very small numbers

$$T^{(i)} = (Q_s^T)^{-1} M$$

As, $(Q_s^T)^{-1}$ has some negative elements therefore the estimator, $\boldsymbol{T}^{(i)}$ may also have some negative elements. The negative elements may be replaced with zero for area estimate of that class (Yuan, 1997). $\boldsymbol{T}^{(i)}$ is multiplied with the area of a pixel to get area estimates in terms of m^2.

Additive Estimator

The additive estimator for the area estimation can be given by Yuan (1997):

$$\boldsymbol{T}^{(a)} = (N_1^{(a)}, N_2^{(a)}, \ldots\ldots\ N_r^{(a)})^T$$

In final form:

$$T^{(a)} = N_{i+} + \left(\sum_{i=1}^{r} n_{ji} - \sum_{j=1}^{r} n_{ij} \right) \frac{N}{n}$$

Where,

$$i = 1, 2, \ldots\ldots., r$$

The correction is the difference between the sum of elements in the r^{th} column and the sum of elements in the r^{th} row in the sample error matrix (C_s) multiplied by *N/n*. $T^{(a)}$ is multiplied with the area of a pixel to get area estimates in terms of m2. As the subtraction is used in the computation, some of these estimates may be negative, but not all of them. The negative values are replaced with zero for computation of area estimate of that class (Yuan, 1997).

Bias Removal Technique

This estimator is based on the assumption that if error of commission and error of omission are equal then area estimates are accurate.
Where:
Bias = Error of commission – Error of omission
Error of commission = $1 - n_{ii} / n_{i+}$
Error of omission = $1 - n_{ii} / n_{+j}$

If bias is positive, it means commission error is larger therefore area may be over estimated. If bias is negative it means omission error is larger therefore area may be under estimated.

Once the bias is computed, the area may be estimated by Gallego (2004):

$$A_i = (1 - Bias)\, N_{i\,+}\, a$$

Where,
A_i is the actual area of a land cover class and *a* is the area of a pixel (m2)

This method produces accurate results if the bias is smaller than commission and omission errors.

Map Marginal Proportions Based Estimator

From the known map marginal proportions (π_i) and sample error matrix, true map proportions (t_i) are computed. Then multiplying the true map proportion (t_i) with total image area (A_T), area estimate of class i is obtained. Map marginal proportions refer to relative areas of each land cover class. It can be obtained from other existing ancillary data. If map marginal proportions are not available from ancillary data then these may be computed from sample error matrix by Card (1982):

$$\pi_i = \frac{n_{+j}}{n}$$

Where, i, j = 1…….r

A two step procedure is required in this technique to compute true map proportions.

In first step, cell probabilities are computed.

In the next step, the sum of cell probabilities for a land cover class is computed to produce true map proportions.

Cell probabilities can be computed as:

$$t_{ij} = \sum_{i=1}^{r} \frac{\pi_j n_{ij}}{n_{i+}}$$

The sum of cell probabilities is obtained as shown in Table (2).

Table 2: Cell probabilities as calculated from error matrix and map marginal proportions

Classified Image	Reference Image				
	Class 1	Class 2	Class 3	…….	Class r
Class 1	t_{11}	t_{12}	t_{13}	…….	t_{1r}
Class 2	t_{21}	t_{22}	t_{23}	…….	t_{2r}
………	…….	…….	…….	…….	…….
Class r	t_{r1}	t_{r2}	t_{r3}	…….	t_{rr}
True proportion	t1	t2	t3	…….	tr

Where,

$$t_j = \sum_{i=1}^{r} t_{ij}$$

Actual area estimate of a class is then computed from,

$$A = t_j A_T$$

Conceptually, this approach weighs each sample from a classified image according to the number of samples from that class and the proportions of the classified image assigned to the class or (π_{i} / n_{i+}).The distribution of the samples from class i among the reference classes j, as contained in the n_{ij}, determines how the area in class i of classified image is redistributed (Woodcock and Gopal, 2000).

Evaluation Criteria

In order to evaluate the Confusion matrix area estimators and conventional techniques total quantitative error, dispersion and bias are used. The first criteria are the total quantitative error which is the sum of difference between the area proportions of land cover category on two maps. the second criteria is average absolute bias, which is the consistent difference between estimated proportions and their true values while the third criteria is the average dispersion (i.e., converse of precision), which can be described by the size of the covariance matrix for each multivariate estimate.

3. Results and Discussion

All result; of this paper illustrated in table (3) and (4) and in Charts (1), (2), (3) and (4). In tables (3) and (4) the area percentage estimated by software program; Area Estimation Methods via Remotely Sensed Data (AEMRSD) is recorded. First row (Proportion in reference map) illustrate the true area percentage for deferent classes. Second row (Proportion in classified map) illustrate the area percentage classified for different classes. Below rows (Average area percentage estimated by different estimators) illustrate the average of area estimated for sampling from 10 % to 100 % by different estimators. In this table total quantitative errors are used to evaluate different estimators as first evaluation criteria

Output of Hard Classified Data with Simple Random Sampling

In this case hard classified image and hard reference image are used, error matrices generated with simple random sampling scheme and the area were investigated by different estimators for different classes in table (3).

Table (3): Average area percentage for different classes estimated by different estimators (Random Sampling)

Land cover type		Built-up land	Grass land	Trees	Agriculture	Barren land
Proportion in reference map (%)		23.7	28.6	29.7	8.5	9.6
Proportion in classified map (%)		22	38.8	24	9.8	5.4
Average area percentage estimated by deferent estimators	Additive Estimator	23.7	28.5	29.7	8. 5	9.6
	Bias Removal Technique	23.3	30.6	27.6	8.7	7.5
	Direct Estimator	23.7	28.6	29.7	8.5	9.6
	Map M. Proportion based Estimator	23.6	23.7	31.8	7.6	13.3
	Inverse Estimator	23.8	28.2	29.7	8.4	9.9
	Pixel Counting	21.9	38.9	23.9	9.9	5.5

The results in table (3) show that the areas estimated from the different methods are not the same. For all classes the total quantitative error (first evaluation criteria) between the proportions of land cover categories, are the difference in the mean of total area estimated and reference map, reduced from 23.4% with pixel counting estimator to 0.2 % in additive Estimator, 6.8 % in bias removal technique, 0.0 % in direct Estimator, 11, 7 % in map marginal proportion based Estimator and 0.9% in inverse estimator.

The direct estimator has smallest total quantitative error 0.1 % and pixel counting estimator have largest total quantitative error 23.35%. Generally the area estimated from confusion matrix area estimators is much closer to the true value than that estimated by conventional methods (pixel counting). From table (3) it is obvious that the average estimate from the direct estimator was virtually unbiased for all classes compared with those taken from the inverse estimator this also suggested by Czapleski and Catts (1992). Comparison of the individual area estimates for each class showed that the direct estimator produced the most stable area estimates with mean values closer to the true proportions while map marginal proportion based Estimator produced the most unstable area estimates.

Output of Hard Classified Data with Systematic Sampling

In this case the error matrices have been generated with systematic sampling scheme and the area is investigated by different estimators for different classes. The area percentage estimated by software program is recorded in table (4) using systematic sampling with hard classified image.

Table 4: The Average area percentage for different classes estimated by different estimators (Systematic sampling)

Land cover type		Built-up land	Grass land	Trees	Agriculture	Barren land
Proportion in reference map (%)		23.7	28.6	29.7	8.5	9.6
Proportion in classified map (%)		22	38.8	24	9.8	5.4
Average area percentage estimated by deferent estimators	Additive Estimator	23.7	28.6	29.6	8.5	9.7
	Bias Removal Technique	23.2	30.8	27.5	8.7	7.5
	Direct Estimator	23.6	28.6	29.7	8.4	9.6
	Map Marginal Proportion based Estimator	23.5	23.6	31.9	7.6	13.3
	Inverse Estimator	23.6	28.5	29.4	8. 6	9.9
	Pixel Counting	21.9	38.8	24.0	9.8	5.4

The results show that the total quantitative error between the proportions of land cover categories reduced from 23.2% with pixel counting estimator to 0.1 % in Additive Estimator, 7.2 % in Bias Removal Technique, 0.1 % in Direct Estimator, 11,8 % in Map Marginal Proportion based Estimator and 0.7% in Inverse Estimator.

Although the sampling scheme is changed (from random sampling to systematic sampling)the behavior of estimators is similar the direct estimator still produces the most stable area estimates with mean values closer to the true proportions and map marginal proportion based Estimator still produces the most unstable area estimates.

Bias and dispersion are illustrated in the following charts: Average absolute bias and the average dispersion are plotted in charts (1) through (4) to evaluate different estimators.

The average absolute biases for all estimators under different sampling fraction and random sampling scheme are given in charts (1) and (2). Some important observations can be made on these charts:

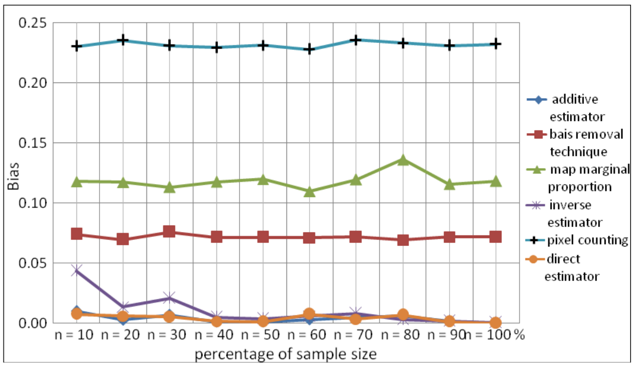

Chart (1): Bias (Hard Data with Random Sampling).

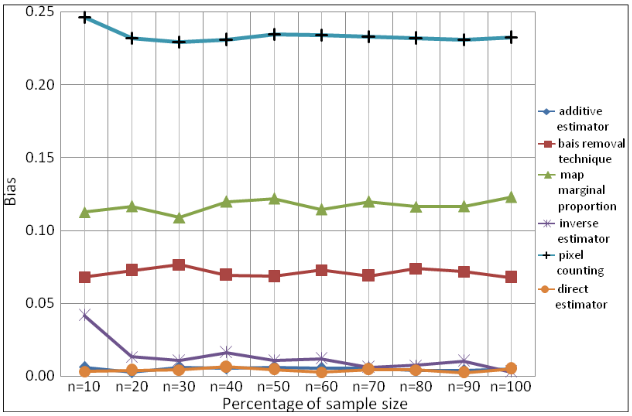

Chart (2): Bias (Hard Data with Systematic Sampling)

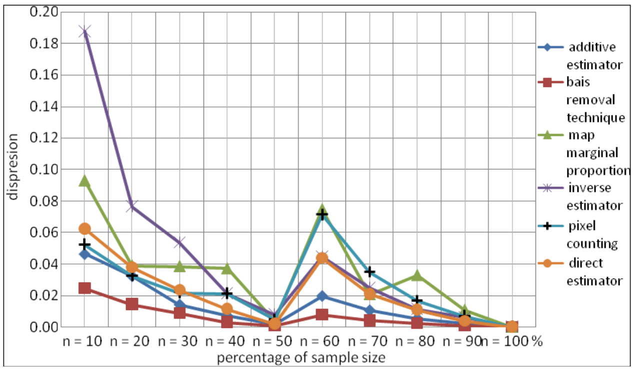

Chart (3): Dispersion (Hard Data with Random Sampling)

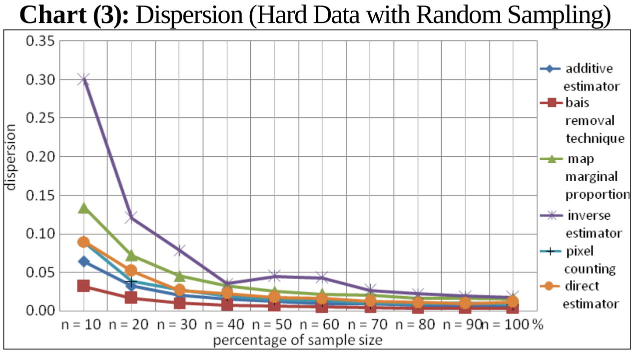

Chart (4): Dispersion (Hard Data with Systematic Sampling)

For all sample size pixel counting estimator (conventional technique) often has the largest bias than all confusion matrix area estimators. Otherwise in confusion matrix area estimator's direct and additive estimators have the smallest bias for all samples fraction and map marginal proportion estimator often has the largest bias. Generally the direct and additive estimators have the smallest bias of all estimators under all sampling fractions.

Although changing sampling scheme (random to systematic sampling) chart (2) gives same results in chart (1) in two sampling schemes direct and additive estimators have smallest bias in all estimators and pixel counting estimator has the largest bias it means that no effect of sampling scheme on biases of all estimators.

Charts (3) and (4) illustrate the average dispersion for all estimators direct observation leads to: The dispersions for all estimators approach zero as the sampling fraction (*f*) increases. This means increasing in sampling fraction leads to increasing in precision of area estimation for all estimators. Inverse and map marginal proportion estimators have the largest dispersion in all estimators and bias removable technique and additive estimators have the smallest dispersion in all sampling fraction.

4. Conclusion

From the results of the investigation the following conclusions can be obtained:

- As the size of sample used in area estimation increases the accuracy of area estimates obtained from all techniques in general increases.
- All areas which are estimated by based confusion matrix area estimators are more accurate and closer to the true areas than that by pixel counting estimator.
- The direct and additive estimators produce the most accurate and consistent results than other estimators. Based on the results from simple random sampling and systematic sampling done on evaluation criteria.
- Map marginal proportion based estimator and inverse estimator produce accurate results only in cases of large size testing samples.
- In case of using based confusion matrix area estimators to estimate different class area for land use purposes the direct and additive estimators are recommended.

References

[1] Card, D., (1982). "Using known map category marginal frequencies to improve estimates of thematic map accuracy". Photogrammetric Engineering & Remote Sensing, 48(3): 431−439.
[2] Czaplewski, R. L., & CATTS, G. P. (1992). "Calibration of remotely sensed proportion or area estimates for misclassification error", Remote Sensing of Environment, 39:29−43.
[3] Dymond, J. R., (1992), "How accurately do image classifier estimate area?", International Journal of Remote Sensing, 13:1735-1742.
[4] Dymond, J. R., (1993), "An improved Skidmore/Turner classifier", Photogrammetric Engineering & Remote Sensing, 59(5):623-626.
[5] Gallego, F. J., (2004), "Remote sensing and land cover area estimation", International Journal of Remote Sensing, 25(15): 3019-3047.
[6] Hay, A. M., (1988)," The derivation of global estimates from a confusion matrix", International Journal of Remote Sensing, 9:1395–1398.
[7] Ibrahim, M. A., (2004), "Evaluation of soft classifiers for remote sensing data ", Ph.D. Thesis, Indian Institute of Technology, Roorkee, India.
[8] Stehman, S. V. (2005). "Comparing estimators of gross change derived from complete coverage mapping versus statistical sampling of remotely sensed data". Remote Sensing of Environment, 96:466−474.
[9] Stehman, S. V. (2009). "Model-assisted estimation as a unifying framework for estimating the area of land cover and land-cover change from remote sensing". Remote Sensing of Environment, 113:2455–2462
[10] Woodcock, C. E., and Gopal, S., (2000), " Fuzzy set theory and thematic maps: accuracy assessment and area estimation ", International Journal of Geographical Information Science, 14(2):153–172.
[11] Yuan, D., (1997), " A simulation comparison of three marginal area estimators for image classification", Photogrammetric Engineering and Remote Sensing, 63(4):385–392.

Author Profile

Tarek A. El-Damaty received B.Sc. from the faculty of engineering, Ain Shams University, Egypt, the M.Sc. Degree from the faculty of engineering, Banha University, Egypt and Ph.D. degree from faculty of civil engineering, technical university in Prague; CZ. He is currently working as an Assistant Professor in College of Engineering and Islamic Architecture, Umm Al_Qura University, K.S.A. His research interests include GPS, GIS, Remote sensing and decision-support system.