ESTIMATION OF INTERPRETATION ERROR IN REMOTE SENSING

Risto Kuittinen, Yu Xiaowei, Panu Sotkas
Finnish Geodetic Institute
Helsinki, Finland

Abstract

The goal of the interpretation of satellite image must be a result which gives the location of the objects and the accuracy of the location and characteristics. A system is presented which gives a possibility to estimate the interpretation errors in different parts of the interpretation process as well as to determine the accuracy of the result.

1. INTRODUCTION

Remote sensing has two components, measurement of object radiance with a sensor, and interpretation of the signal recorded. Both include many sources of error that significantly affect the accuracy of the outcome. The goal of the remote sensing process must be a result which gives the location of the objects and the accuracy of the location and characteristics. This problem has received very little notice in the literature, references to it usually being restricted to the testing of interpretation results. However, when testing interpretation results we have to take into account that there are information on different kinds of target, namely

a) very permanent objects, areas which can be verified years after imaging (e.g. geological features),
b) nonpermanent objects, which can be verified within some months of imaging (e.g. forest stands),
c) rapidly changing objects, which can be verified only by making in situ measurements (e.g. sea ice, water quality) and
d) objects difficult to verify at all (e.g. cloud types).

Different methods are needed for estimating interpretation errors in each of the previous cases. What selected depends on whether we are determining the accuracy of the algorithms used or of the interpretation result. However, we do not always have any practical means of using in situ measurements, and in these cases the error has to be estimated from information obtained from satellite data only.

The interpretation process usually requires the interpretation of very complex data (several channels and targets) to yield a simple output, and different types of input have to be classified in the same class /10/.

The interpretation proceeds in the following steps:

a) Identification of the object.
b) Interpretation of the object characteristics.

c) Combining results from steps a) and b) to make thematic maps or obtain statistics.
d) Computing other variables or making forecasts on the basis of remotely sensed data and other data (mainly field measurements).

The requirement for this kind of interpretation is that remotely sensed and other data can be combined in space and time. Remotely sensed data can give representative areal averages (statistics) but not very accurate point values. Other data comprise are mainly point observations, and the computations needed to determine areal averages. The most useful input data for models are the areal averages of variables.

Error estimation of the interpretation process is highly desirable for the following reasons:

- During the interpretation process error estimates can be used to improve the accuracy of the supervised interpretation system.
- Error estimates for the whole interpretation process can be used as such when there are no opportunities for testing the accuracy of the product.
- In cases where the final product can be tested with field measurements it is useful to make a well-planned field test to verify the accuracy of the product.

It is important to see that all the test material itself has some errors; thus the comparison do not give totally right estimates for the accuracy of the interpretation.

2. THE INTERPRETATION PROCESS AND ERROR ESTIMATION PROCEDURE

Error estimation for the remote sensing process can be made in three steps. The first two supervise the interpretation process itself and the third one gives a reliable estimate of the accuracy of the product. Figure 1 shows the interpretation process when a hierarchical interpretation method is used. In non-hierarchical methods, the interpretation and error estimations are made at one level only.

2.1 Interpretation process for object recognition

The following graph describes a hierarchical interpretation process on three levels but the process can also have only one level. The information required for error estimation comprises the classifier, the distribution of the data and the distribution of errors, because different estimation methods are used for different classifiers. If there are no computation methods to estimate the error, it must be estimated by testing the interpretation result. The results of this process are the areas of different classes.

LEVEL 1.

1. Estimate parameters for object detection algorithms or classifiers
2. Compute errors for all or selected pairs of classes

2.1 Errors accepted, continue (3)

2.2 Errors not accepted, return to estimate parameters (1)

3. Interpretate the whole data
4. Estimate errors and/or determine the accuracy of the interpretation result.

LEVEL 2.

The only data to be interpreted in this phase are those in the areas of the classes which have to be divided into more detailed classes.
Procedure as on level 1.

LEVEL 3.

Procedure as on level 1.
5. Compute the errors of the whole process.
6. Determine the accuracy with field investigations.

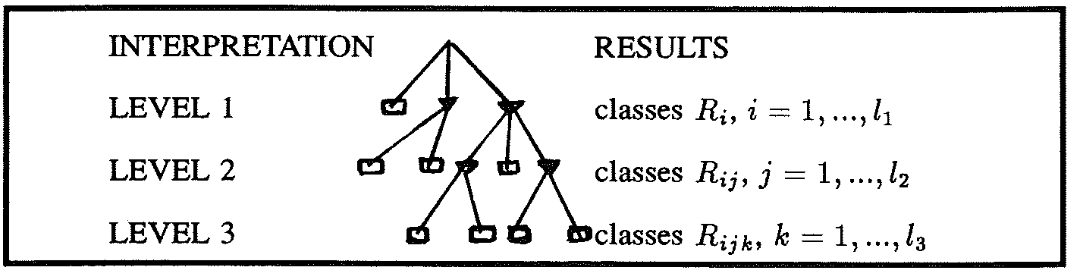

Fig. 1 A hierachical interpretation method of three levels

2.2 Interpretation of object characteristics

1. Develop the algorithm
2. Validate the algorithm
3. Test accuracy of validation
   3.1 accepted, go to 4.
   3.2 not accepted go to 1.
4. Interpretate the data using validated algorithms
5. Estimate the error of the product
   - estimate error for the areas
   - estimate error for algorithms
6. Check in the field if possible or necessary

3. RADIANCE MEASUREMENT

The result of this part of the remote sensing process should be a measurement (target radiance or digital number of the pixel) of known accuracy. For numerical interpretation it is important to know the accuracy because we can then determine the uncertainty of the estimated parameters used in the classifiers, algorithms and models. A measured signal consists of the object radiance, background radiance, noise (affected by the sensor) and disturbance (affected by the atmosphere). Uncorrected satellite images include systematic errors, that are mainly caused by irradiance variations, sensor calibaration, etc. Random errors also exist, but they are almost impossible to detect. Gross errors (e.g. striping) are easy to detect but they are difficult to correct because the original radiance information is usually missing. Only inter- or extrapolation can be used to obtain data to replace those missing.

Sensor calibration

Careful sensor calibration eliminates systematic errors of data. It is difficult for the ordinary data user to obtain calibration information. Results concerning the accuracy of some satellite-borne sensors have been reported. Nianzeng (1991) /7/ has studied the NOAA/AVHRR data and he estimates an uncertainty of 7.5 % in the gain. The gain coefficients showed a reduction of 4% in channel 1 and of 13% in channel 2 during a 19 month period. Similar results have been obtained for Landsat data /8/.

Atmospheric effects

Atmosphere is the main reason for reducing the quality of the data. Using atmospheric models it is possible to estimate the actual object radiances from the measured radiances. Once the atmospheric corrections have been made, the images taken at different times can be compressed and sometimes the contrasts in the image can be improved. Atmospheric corrections can change the pixel values by 20%.

Irradiance variations

When data of non-sunsynchronous satellites are used the effects of irradiance variations in the images must be corrected. This correction is especially important in cases in which object characteristics are determined but it is also useful in object recognition. During daylight hours the irradiance variations can affect the image variations by as much as 30 - 40%. However variations depend to some extent on the object dependent /4/.

Measurement geometry and object location

The location of the object in the imaged area and the scanning geometry affect the radiances measured. If the scan angle is large, over 10°, the effect of scanning should be corrected. The magnitude of the variation caused by scanning varies in general between 10 and 30% /4/,/9/.

Topography

In areas where the sun is not at the nadir, topography changes the irradiance in the terrain. Changes in pixel values can be of the order of 10%. These factors have a considerable effect on the contrasts of the image and thus also on the interpretation accuracy.

In conclusion the factors mentioned above can cause errors as great as 50 - 60% in the image if they are not corrected. If corrections are made the accuracy of the radiances can be about 5 - 10% depending on whether single image or images taken at different times are studied.

4. ESTIMATION OF OBJECT DETECTION ERRORS

4.1 Parametric methods

The accuracy of this part of the interpretation process depends on the accuracy of the data and the design of the classifier. First, the error estimates are determined for all pairs of classes to be interpreted. Thereafter an overall error can be determined. The results of this estimation are figures (numbers) which describe the accuracy of the areas of different interpreted classes.

Since only a finite number of samples are obtainable in practice, and therefore in most solutions the parameters of the classifier are considered as random variables. In these cases the probability of error is actually a quantity that evaluates the performance of the classifiers, and the estimation of the probability of the error becomes important.

The methods for error estimation fall into one of the two following categories:

- Parametric methods in which the form of distribution of the error probability is known. In this case the error estimation problem becomes a problem of approximating of the error distribution function using the known statistics of class populations.
- Bounds on probability of error where it is difficult to estimate the parameters of distribution of probability of error even the form is known. The bounds on probability of error are derived in this case.

There are two types of errors in the two-class problem. One results from misclassifying class R_i samples (omission, when R_i is studied) denoted by $P_e(j|i)$, the other results from misclassifying class R_j samples (comission, when R_i is studied) denoted by $P_e(i|j)$. For multiclass problems pairwise errors can be treated in similar manner.

Minimum distance classifier

For this classifier, the upper bound of the error probability are determined by van Otterloo (1978) /12/ by computing:

$$P_e(i) \leq \frac{N}{D_i^2},$$

where $P_e(i)$ is the upper bound of error for class R_i , N is the number of features in the feature space and $D_i = minD_{ij}$, $(i \neq j)$ is a measure of the 'radial neighbourhood' of the class. D_{ij} is determined by minimizing $d(X, M_i)$ subject to the constraint $d(X, M_i) = d(X, M_j)$.

$$d^2(X, M_i) = (X - M_i)^T \Sigma_i^{-1} (X - M_i),$$

where M_i is the mean vector of the class R_i, Σ_i is the covariance matrix of class R_i and X is the feature vector.

Lapsa (1979) /5/ also gave an upper bound for the accuracy involving absolute 'k-moments, Q^k' for one dimension which is defined by

$$P_e(i) \leq (\frac{k}{k+1})^{\frac{k}{k+1}} \frac{Q^k}{c_i^k} \quad (c_i > 0, k \geq 0).$$

where $Q^k = E(|x|^k) = \int |x|^k p(x)dx$.

Maximum likelihood classifier

The decision rule for the Maximum likelihood classifier is the following:

$$\frac{p(X|R_i)}{p(X|R_j)} > \frac{P(R_j)}{P(R_i)} \Longrightarrow X \in R_i.$$

where $p(X|R_i)$ is class conditional density function and $P(R_i)$ is a *priori* probability for class R_i.

By taking minus logarithm at the both sides of the inequation, we have minus-logarithm likelihood ratio $h(X)$:

$$h(X) = ln\, p(X|R_j) - ln\, p(X|R_i) < ln\, P(R_i) - ln\, P(R_j) \Longrightarrow X \in R_i.$$

By assuming that $h(X)$ is normally distributed, we may obtain a rough approximation of the error probability. The expected value and variance of $h(X)$ can be estimated /2/ and thus error probability can be calculated as follows:

$$P_e(j|i) = \int_t^{\infty} p(h|R_i)dh = \int_{(t-\eta_i)/\sigma_i}^{\infty} (2\pi)^{-1/2} exp(-\xi^2/2)d\xi,$$

$$P_e(i|j) = \int_{-\infty}^{t} p(h|R_j)dh = \int_{(\eta_j-t)/\sigma_j}^{\infty} (2\pi)^{-1/2} exp(-\xi^2/2)d\xi,$$

where $t = ln\{P(R_i)/P(R_j)\}$, η_k and σ_k are the estimated mean value and variance of $h(X)$ in class R_k (k=i,j), respectively.

Because in practice the above integration is very difficult, an upper bound on error probability are used. This can be determined in the following way /2/: for any $0 \leq s \leq 1$,

$$P_e(j|i) \leq [P(R_j)/P(R_i)]^s exp[-\mu(s)],$$

$$P_e(i|j) \leq [P(R_i)/P(R_j)]^{1-s} exp[-\mu(s)],$$

$$P_e(ij) = P(R_i)P_e(j|i) + P(R_j)P_e(i|j) \leq [P(R_i)]^{1-s} P(R_j)^s exp[-\mu(s)].$$

Where $\mu(s)$ is the *chernoff* distance. The optimum value of s can be obtained from the curve of error bound versus s.

For normal distribution, an exact expression for $\mu(s)$ can be written as

$$\mu(s) = \frac{s(1-s)}{2}(M_i - M_j)^T[s\Sigma_i + (1-s)\Sigma_j]^{-1}(M_i - M_j) + \frac{1}{2}ln\frac{|s\Sigma_i + (1-s)\Sigma_j|}{|\Sigma_i|^s|\Sigma_j|^{1-s}}$$

which expression can be used in practical calculations.

Piecewise linear classifier

The decision function of the linear classifier is the following:

$$h_{ij}(X) = V_{ij}^T X + V_{ij0} \leq 0 \Longrightarrow X \in R_i,$$

$$i, j = 1......n, i \neq j.$$

Because $h_{ij}(X)$ is a linear function of X, it is normally distributed if X is normally distributed. If X is not normally distributed, $h_{ij}(X)$ is close to normal for large values of N, when the central limit theorem is applicable /2/. The estimation of error can be determined as follows:

$$P_e(j|i) = \int_0^\infty p(h|R_i)dh = \int_{-\eta_i/\sigma_i}^\infty (2\pi)^{-1/2}exp(-\xi^2/2)d\xi,$$

$$P_e(i|j) = \int_{-\infty}^0 p(h|R_j)dh = \int_{-\infty}^{-\eta_j/\sigma_j} (2\pi)^{-1/2}exp(-\xi^2/2)d\xi,$$

where η_k and σ_k $(k = i, j)$ are the estimated mean value and variance of h(X) for class R_i and R_j, respectively.

The error estimation and thus also the minimization of the error is made by determining the parameters of the classifier. In practice this is made using the folloving equations /2/:

$$\eta_i = E\{h_{ij}(X)/R_i\} = V_{ij}^T M_i + V_{ij0},$$

$$\sigma_i^2 = Var\{h_{ij}(X)/R_i\} = V_{ij}^T \Sigma_i V_{ij},$$

$$\eta_j = E\{h_{ij}(X)/R_j\} = V_{ij}^T M_j + V_{ij0},$$

$$\sigma_j^2 = Var\{h_{ij}(X)/R_j\} = V_{ij}^T \Sigma_j V_{ij},$$

4.2 Nonparametric methods

Nonparametric methods are needed in cases where it is usually difficult to define a form of the distribution for the underlying population or the distribution is too complex for practical computations. The most popular methods used are resubstitution, the leave-one-out method, the holdout method and the bootstrap method /2/,/3/. Toussaint (1975) /11/ suggested a method which combine two methods, resubstitution and rotation by a weighing function, and McLachlan (1977) /6/ gave a theoretical consideration to the choise of a weighting function. Design samples and independent test data are then required to ensure that an accurate estimation of the probability of error is obtained.

In remote sensing the most widely used nonparametric methods are parallelopiped classifier, which is the following:

$$L_{l_i} < X < L_{u_i} \Longrightarrow X \in R_i, \qquad i = 1, .., n.$$

in which L_{l_i} and L_{u_i} vectors determining a hypercube in the feature space.

The k Nearest Neighbor-method (kNN) is used to determine the denstiy function for a classifier. If there are m_i samples of class R_i within the hypersphere V , the decision rule is the following /2/:

$$m_i > m_j \Longrightarrow X \in R_i, \qquad i \neq j, \quad i = 1, ..., n \quad and \quad j = 1, ..., n.$$

For the parallelopiped classifer and the kNN-method the error estimation $P_e(i)$ for class R_i is the following:

$$P_e(i) = \frac{m_{w_i}}{m_i},$$

where m_{w_i} is the number of missclassified samples.

4.3 Multiclass problem

The algorithms presented only give error estimations for cases with two classes. In practice, though, several classes have to be interpreted simultaneously. The error of the interpretation result can then be determined from the errors of the individual class pairs.

First the classes are ranked from most important to least important. The error estimates $P_e(i|j)$ and $P_e(j|i)$ for the class pairs then produce two matrices in which the elements are also ranked. If only the average error $P_e(ij)$ for each class pair R_i/R_j is estimated, the error sums $P_e(i)$ can be determined as follows:

$$P_e(i) = \sum_{j=i+1}^{n} P_e(ij) \qquad (0 < P_e(ij) < 1) \quad i = 1, ..., n$$

If the errors are weighed by the importance of the classes, the error sum can be determined for each class:

$$P_e(i) = \sum_{j=i+1}^{n} w_{ij} P_e(ij) \qquad (0 < w_{ij} < 1) \quad i = 1, ..., n,$$

where $w_{12} >, ..., > w_{n-1,n}$.

In cases in which omission and commission errors can be estimated by the previously mentioned methods, omission and comission errors produce a matrix in which the error estimations are the elements of the matrix. The errors for class R_i can be determined as follows:
Omission errors:

$$P_{eo}(i) = \sum_{j=1, j \neq i}^{n} P_e(j|i) \qquad i = 1, ..., n$$

Commission errors:

$$P_{ec}(i) = \sum_{j=1, j \neq i}^{n} P_e(i|j) \qquad i = 1, ..., n$$

Weights can also be used if enough background information is available. The usefulness of the parameters, classes and interpretation result can be decided on the basis of these errors.

4.4 Areal error

If a hierarchical multilevel interpretation process is used as presented in Figure 1 the estimates of the area error must be determined from each level and combined with the final result. Each level gives correctly and falsely classified areas: falsely classified areas cannot be corrected on later levels because of the hierarchy of the system, Fig 1.

The error estimate of the areas ΔA_i of different classes R_i can be determined by using the commission of errors $P_{ec}(i)$ as follows:

$$\Delta A_i = P_{ec}(i) A_i$$

The total error ΔA_{tot} for the classification area is then estimated.

$$\Delta A_{tot} = \sum_{i=1}^{n} \Delta A_i$$

All these estimates are necessary in evaluating the accuracy and usefulness of the result.

5. ESTIMATION OF OBJECT CHARACTERISTICS ERRORS

The goal of the interpretation is often to determine the characteristics of the objects in the area of the image. Examples are ice types, timber volumes and the snow water equivalent. In all cases, first the areas of the classes are determined and then their characteristcs using algorithms which convert the digital numbers of the image to the character of the object. Thus the accuracy of the interpretation result has three parts: the accuracy of the area of the object, the accuracy of the data themselves and the accuracy of the algorithm used.

For a single pixel (one measurement) the error of the variable y can be estimated using differentiation of the algorithm f(X), if the standard error of the algorithm parameters Δp and radiance measurement ΔX are known.

$$\Delta y_i = \left[(\frac{\partial f(X_i)}{\partial X})^2 (\Delta X)^2 + (\frac{\partial f(X_i)}{\partial p})^2 (\Delta p)^2 \right]^{1/2}$$

Over an area A_i (which consists of several pixels) the mean error of the variable y is given by Δy:

$$\Delta y = \left[(\Delta A_i \overline{y}_i)^2 + (A_i \Delta \overline{y}_i)^2 \right]^{1/2}$$

in which ΔA_i is the error of the area of the class R_i and $\Delta \overline{y}$ is the mean of the errors of the individual pixels (variables). These calculations are needed for computing the errors of the areal averages, whereas the errors for single pixels are also needed map production.

6. ACCURACY OF THE INTERPRETATION RESULT

The accuracy of the interpretation result can only be tested by making field measurements. These measurements can have two aspects: determining the interpretation or objects accuracy, and determining the characteristics of the objects.

This method only can be used when we can be sure that there are no changes in the studied area between the imaging and testing or that the changes which have taken place can be identified and thus omitted from the analysis. The result of the testing is an error matrix.

In remote sensing, different sampling methods are used to generate the error matrix. It is very important that the testing should cost as little as possible and produce good estimates of accuracy.

a) Random sampling

Many estimates used to assess accuracy are consistent and unbiased when this method is used. The main problem with simple random sampling is under-representation of small classes, with a reasonable sample size.

b) Systematic sampling

The major advantage of this method is that, if the population contains some periodicity, then regular spacing of the sampling units might result in unrepresentative samples. The problem of under-representativness of the small classes is the same as in random sampling.

c) Stratified sampling

In remote sensing the stratification by categories is used because the classification result has already been divided into classes. The classes can be used as a stratum, and the number of test pixels required can be selected from each class direct. Thus the minimum number of test pixels can even be selected from small land use classes. However, this sampling method creates problems for calculations of total accuracy and omission estimates because the data are category weighted not area-weighted. However, these estimates could be 'corrected' if the areas of the classified categories were known.

From these results (error matrix) the following figures can be determined for class R_i:

Interpretation accuracy:

$$(x_{ii}/\sum_{j=1}^{n} x_{ij}) * 100 \qquad i = 1, ..., n$$

Object accuracy:

$$(x_{jj}/\sum_{i=1}^{n} x_{ij}) * 100 \qquad j = 1, ..., n$$

Mean accuracy:

$$\frac{2x_{ii}}{\sum_{j=1}^{n} x_{ji} + \sum_{j=1}^{n} x_{ij}} * 100 \qquad i = 1, ..., n$$

Kappa coefficient κ of total interpretation:

$$\kappa = \frac{m\sum_{i=1}^{n} x_{ii} - \sum_{i=1}^{n}(\sum_{j=1}^{n} x_{ij} \sum_{j=1}^{n} x_{ji})}{m^2 - \sum_{i=1}^{n}(\sum_{j=1}^{n} x_{ij} \sum_{j=1}^{n} x_{ji})}$$

Kappa coefficients for individual categories:

$$\kappa_{ci} = \frac{mx_{ii} - \sum_{j=1}^{n} x_{ij} \sum_{j=1}^{n} x_{ji}}{m\sum_{j=1}^{n} x_{ij} - \sum_{j=1}^{n} x_{ij} \sum_{j=1}^{n} x_{ji}}$$

$$\kappa_{oi} = \frac{mx_{ii} - \sum_{j=1}^{n} x_{ij} \sum_{j=1}^{n} x_{ji}}{m\sum_{j=1}^{n} x_{ji} - \sum_{j=1}^{n} x_{ij} \sum_{j=1}^{n} x_{ji}}$$

see /1/, where m is the total number of testing samples and x_{ij} $(i, j = 1, ..., n)$ is the element in error matrix.

These are very useful when the accuracies of different interpretations are compared.

7. DISCUSSIONS

Remote sensing methods are developed to get information over large areas fast and with sufficient accuracy. The accuracy of these methods is mainly verified by field measurements all of which have two restrictions, accuracy and areal representativeness. For this reason the extact accuracy of the interpretation result is very difficult to determine.

The large amount of data in one satellite image restricts the possibilities to use very complicated computing methods. The methods presented in previous chapters can be used for satellite data (SPOT, LANDSAT, NOAA etc.). Only partly it is possible to estimate interpretation errors by using the parameters of the classifiers even if supervised methods are used. When unsupervised methods (clustering and neural networks) are used, the only possibility for error estimation is to test the interpretation result by field measurements. For these reasons it is important to develop accurate and effective sampling methods.

The demands for accuracy evaluation increase when the use of remotely sensed data increase. Error estimation possibilities depent on interpretation methods and applications. Thus it is important to develop error estimation in connection with the application.

8. REFERENCES

/1/ Cohen, J.: A coefficient of agreement for nominal scales. Educational and psychological measurement, Vol. 20, No 1, 1960.

/2/ Fukunaga, K.: Introduction to Statistical Pattern Recognition. 1990.

/3/ Hand, D.J.: Recent advances in error rate estimation. Pattern Recognition Letter 4, 1986.

/4/ Kuittinen, R.: Determination of areal snow water equivalent using satellite images and gamma ray spectrometry. Acta Polytechnica Scandinavica. Ci 91. Helsinki, 1988.

/5/ Lapsa, Paul M.: Some statistical bounds for the accuracy of distance-based pattern classification. Pattern Recognition, Vol. 11, No. 2, 1979.

/6/ McLachlan, G.J.: A note on the choice of a weighting function to give an efficient method for estimating the probability of misclassification. Pattern Recognition, Vol. 9, No. 3, 1977.

/7/ Nianzeng,C., Grant,B., Flittner,D., Slater, P., Biggar,S.: Results of calibration of the NOAA-11 AVHRR made by reference to calibrated SPOT imagery at White Sands, N.M. SPIE PROCEEDINGS Vol. 1493, 1991. Calibration of Passive Remote Observing Optical and Microwave Instrumentation.

/8/ Price, J. C.: Calibration of satellite radiometers and the comparison of vegetation indices. Remote Sensing of Environment, No. 21, 1987.

/9/ Royer, A., Vincent, P., Bonn, F.: Evaluation and correction of viewing angle effects on satellite measurements of bidirectional reflectance. Photogrammetric Engineering and Remote Sensing, Vol. 60, No. 12, 1985.

/10/ Sydenham, P.H.: Handbook of Measurement Science. 1982.

/11/ Toussaint, G.T., Sharpe, P.M.: An efficient method for estimating the probability of misclassification applied to a problem in medical diagnosis. Comput. Biol. Med. 4, 1975, pp. 269-278.

/12/ Van Otterloo,P.J., Young, I.T.: A distribution-free geometric upper bound for the probability of error of a minimum distance classifier. Pattern Recognition, Vol. 10, No. 4, 1978, pp. 251-286.

/13/ Vinogradov, B.V.: Errors in the indication of geosystems from remote sensing imagery. Mapping Science and Remote Sensing, 1989.