# DETERMINING THE SAMPLE SIZE REQUIRED FOR THE MATERIALS QUALITY ANALYSIS IN FORM OF SHEET USING IMAGE ANALYSIS

**ABABEI STEFAN[1*], ABABEI OANA[2]**

[1] *"Vasile Alecsandri" University of Bacau, Calea Marasesti 156, Bacau, 600115, Romania*

[2]*Yokogawa Europe B.V. - Romania Branch 6 Dimitrie Pompeiu Bvd, Building E Novo Park 2, Romania*

**Abstract.** On-line control of great dimensions sheet form products, e.g. paper sheet, involves the existence of some sensors which are moved across the sheet. Consequently, some parameters, e.g. impurities amount on surface unit, must be measured off line with a representative sample; the dimensions of the sample are surfaces impurities amount functions. This paper presents a preliminary analysis method of surface impurities amount in the purpose of determining the sample dimensions.

**Keywords**: image analysis, impurities amount, sample dimensions

## 1. THE USEFULNESS OF IMAGE ANALYSIS IN MONITORING THE PRODUCTS QUALITY IN FORM OF SHEETS

In the production of materials in form of sheets, the surveillance of these products quality based on the analysis of their area is used more widely. The impurities present inside the materials causes certain spots with random sizes and distributions on the surface of these produced materials, this is the case mainly for paper but also for textile materials or metal plate.

When the surface is very large it is difficult to create a detailed technically feasible analysis of the whole surface or it is not economically justified. In this case, a detailed analysis is made for a sample or a fraction of the created material.

Strictly speaking in the production of paper it is necessary to show that, with the development of computer technology, several systems and methods for automatically calculating the total area of spots and methods for estimating the errors occurred as the difference between estimation and actual total area covered by them have been created.

In [1] it is shown that the errors from different methods applied in the same areas are negligible compared with errors caused by random distribution of spots in the sample. Given that the errors caused by the random distribution of spots in the sample can only be reduced by increasing the scanned area, the problem to be solved is: how much area should be scanned in order to get an estimated error of acceptable amplitude.

As the main sources of errors in using image analysis to estimate the quality of the products in sheets can be mentioned:

[*] Corresponding author, email: sababei@ub.ro

- the errors introduced by device used for image acquisition;
- errors caused by the settings made by the operator for the equipment used;
- the errors introduced by the procedure used to select the sample or the fraction analyzed in order to determine the quality of the product.

## 2. THE ESTIMATION OF THE SAMPLE AREA TO BE ANALYZED

In TAPPI Test Method T437/90 ('impurity in paper and paper bands') a $10m^2$ area is indicated to be analyzed in order to obtain significant results in specification of the impurity content of the cellulose and paper.

The domain containing the estimated values is called confidence interval and we will note it from now on with CI Very often it is used a confidence level of 95% to determine the confidence interval, meaning that only one of 20 measurements will be outside the confidence interval [2]. Similar expressions can be determined for any value of the confidence level.

Often it is practical to express confidence interval CI, as a value relative to the value that is estimated. For example, if the estimated value is 200ppm and CI is 20ppm, then the size of the relative confidence interval is 10%, we will note from now on the relative confidence interval with CI%. To estimate the $A_S$ area that should be scanned to get a CI% of a certain size, when the average spot size $A_M$ and the content of impurities in ppm are given it is possible to use the following equation:

$$A_S = A_M \cdot 1000000 \cdot K_{CI\%}[ppm] \tag{1}$$

where $k_{CI\%}$ in *ppm* is a *CI%* function and independent of the $A_M$ ,$A_S$ values.

Table 1 shows the values for the coefficient $k_{CI\%}$ for the most common values for *CI%*.

Table 1. Values for the coefficient $k_{CI\%}$ for different CI%.

| CI% | 5 | 10 | 20 | 25 | 30 | 50 | 75 | 100 |
|---|---|---|---|---|---|---|---|---|
| $k_{CI\%}$ | 4799 | 1199 | 299 | 191 | 132 | 48 | 22.3 | 13.0 |

A simple straight line for each value for CI% results when the scanned surface is represented by the content of impurities in ppm in a double logarithmic representation (Figure 1). [2]

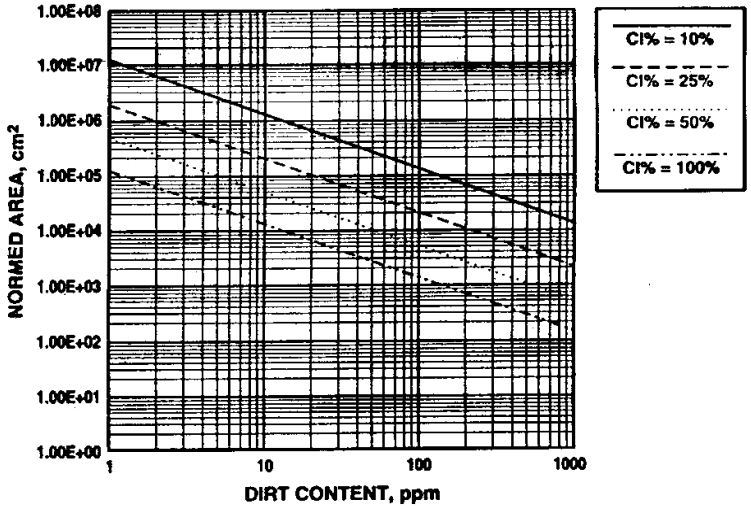


Fig. 1. Diagram for determining the scanned area.

The following are experimental charts for estimating the scanned area when the content of impurities (in ppm), the confidence interval CI% and estimation of the average size of spots are given, as well as for determining the confidence interval when the scanned area and content of impurities in ppm are known. Such a chart is shown in Figure 1.

This chart can be used as follows:

a) To estimate the scanned area for a population of spots with any average value of sizes

1. It is estimated the impurity content in ppm by an image analysis;
2. The confidence interval CI% is chosen;
3. The normalized area is read from the chart;
4. The average spot size $A_M$ is estimated in $mm^2$;
5. The estimated area to be scanned is calculated using the formula:

$$A_S = A_N \cdot A_M \tag{2}$$

b) To estimate the confidence interval when the scanned area and the content of impurities in ppm are known:

1. Calculate the normalized area with $A_N=A_S/A_M$;
2. An estimation for the length of the confidence interval can be found in the intersection of the normalized areas horizontal line and the impurities in ppm vertical line. An interpolation can be done to determine an interval unspecified in the chart;
3. To express *ppm±CI* the relationship *ppm± (CI%*ppm/100%)* is used.

It must be made clear that the number of samples (i.e. sheets) estimated needs to be more than 6 in order that the statistical estimates to be valid.

When there is no initial information about the estimated parameters, it is necessary to obtain this information from a small initial sample size $T_0$. With this information an initial model for the parameters can be obtained and therefore an approximate value of the sample. Obtaining the initial estimation for the impurities content started by analyzing samples of size A4 (297x210 mm) of the tested product. These samples were scanned using a MUSTEK SCANNER with a color scale consisting of 256 gray tones. The image was saved in BMP format as file patrat.bmp. an application in C++ was made in order to achieve the following:

- converts image file into a numeric matrix with a number of rows and columns equal to the width and height in pixels of the original image (saved in file patrat.bmp).the value corresponding from the color table attached to the image is assigned to each array element in the original image file;
- the array is saved in a file called matrfig.m. This file can be further processed to rebuild the original image using matlab programming environment;
- reads the color table attached to the image file, converts it into a numerical array and saves it into a file called culori.m. This file will also be used afterwards in rebuilding the original image. The color array has 256 rows and 3 columns. Each line corresponds to a gray tone (between 0 and 255) and each line contains color indexes rgb (red, green, blue) corresponding to the color intensity to those in the corresponding tone line. the minimum intensity is 0 and the maximum intensity is 1, each intermediate intensity being obtained from the absolute intensity of the color in eight-bit related to 255;
- calculates the number of locations in the image array that have the color index less than 155 (they are considered spots caused by impurity content) and displays its value relative to the total number of bits as the impurity content;
- the number of spots n = N and the average size of spots $A_M$ are calculated to determine the average size of the spots $A_M$ in order to use the chart for Figure 1, starting from the same scanned image. The algorithm used is a recursive algorithm that finds all points that are adjacent.

## 3. CONCLUSIONS

Six sheets size A4 are scanned and based on their analysis by using the algorithm presented, the sample size was established according to the chart in Figure 1. It was found after analyzing the entire batch, that the estimated errors based on the estimated sample don't exceed 5% from the preliminary ones.

**REFERENCES**

[1] Ababei, Șt**.,** Determinarea mărimii eşantionului necesar estimării cu precizie impusă a impurităților de suprafață din materialele electroizolante celulozice. SIELMEC 99, Chişinău, 1999, p. 101-106.
[2] Jesus, J., Sampling error in measurement of dirt in pulp and paper. TAPPI Journal, vol. 79, no. 5, 1996, p. 222-229.