

Paweł Hanus*

Estimation of Parameters of Statistical Models Used in Real Estate Cadastre**

1. Introduction

In the scope of detailed surveying regarding geodetic property management services, geodetic and legal activities associated with the basic processes of real estate management are often performed.

These include:

- delimitation of a property,
- subdivision of a property,
- merging and division of a property,
- expropriation of a property, in the case a change of the owner regards only a part of a property.

The above processes refer to land property, in which the extent of property rights is described by the boundaries determined and fixed on the ground by an authorized surveyor [2].

It should be added that the main purpose of these activities is primarily a geometric change in the property boundaries, which has the effect of a change in the legal characteristics of the properties – i.e. a change in the range of ownership rights to these properties. Setting the new ones, or revising the existing property boundaries always requires a field survey, alignment of these measurements, determining the components of shifting boundary markers, correcting the position of boundary markers, cartographic visualization, and the preparation of appropriate formal and legal documentation. These procedures always result in significant land information changes in spatial and legal aspects. Therefore, they must be recorded both in the land and buildings registers (the cadastre) as well as in land and mortgage registers.

* AGH University of Science and Technology, Faculty of Mining Surveying and Environmental Engineering, Department of Geomatics, Krakow, Poland

** This work has been made within scientific research program of Geomatics Department no. 11.11.150.006

Record of a change in the position of the property boundary markers, marked in the cadastre as record parcel (hereinafter referred to as the plot), also brings about changes in the previously determined values of the plot surface area, as recorded in the land register [3]. Quite often we have to deal with a problem that model conditions in the form of functional relations determining the surface area of the plot are directly imposed on the coordinates of the boundary markers, thus the resulting necessity to estimate these models.

It is worth noting that we are quite often facing a discrepancy between the surface area of the plot disclosed in the land register and the surface resulting from the cadastral map. In the case when determining the position of the boundary markers of a plot does not meet current standards¹ (detail of the first-order accuracy), the regulation on land and buildings register [4] imposes leaving the existing surface area of the plot in the land register, often calculated with graphical methods and rounded to 1 are.

Therefore, the purpose of this paper is mainly to identify the problem of recording spatial information regarding the position of boundary markers and the surface area of the plot in the cadastre, as well as to present such a serious problem as the estimation of parameters of statistical models, which can be applied in the cadastre.

2. Functional Conditions for Geodesic Values Defining Property Boundaries

In solving problems related mainly to the division, merging and division, as well as the demarcation of a property, and other forms of determining boundaries, conditions on the determined coordinates of the boundary markers in the form of functional relations on the surface area of the plots which constitute the property in question are often imposed. These relations constitute the conditions which are to be met by the agreed (model) coordinate values of the boundary markers of the analyzed properties.

Let every value X_i , representing the coordinates of the boundary markers meet the stochastic model, i.e.:

$$\hat{X}_i = X_i + dx_i \quad (1)$$

where \hat{X}_i stands for the model value of value X_i , since $E(X_i) = \hat{X}_i$, and dx_i is a random component.

¹ The most popular method of capturing data of the coordinates of boundary points is the cartographic method, depending on vectorization of scanned and calibrated cadastral maps. Standard deviation of such determined coordinates of boundary point for cadastral map in scale 1:2000 one can take at level of some 0.6 m [2].

If between the values of the coordinates of the boundary markers there are functional conditions, these conditions are exactly satisfied by the model values \hat{X}_i , thus:

$$S(\hat{X}_1, \hat{X}_2, \dots, \hat{X}_n) = w \quad (2)$$

where is the value of the function calculated for the model coordinate values of the plot boundary markers.

In general, conditions (2) occur in non-linear form, therefore using the first derivative of Taylor series, they are transformed to a linear form [1], i.e.:

$$\sum_{i=1}^n \frac{\partial S}{\partial X_i} dx_i + S(X_1, X_2, \dots, X_n) = w \quad (3)$$

The best approximation of the model values of the coordinates of the boundary markers are the values of these coordinates determined based on surveys, as after taking into account the relation $\hat{X}_i = X_i - dx_i$, it is shown that the differentials to the model values are the increments of the coordinates of the boundary markers. Thus, the condition (3) can be written as:

$$\sum_{i=1}^n \left(\frac{\partial S}{\partial X_i} \right)_{X_i = \hat{X}_i} \times dx_i + S(X_1, X_2, \dots, X_n) = w \quad (4)$$

After transferring a known value of the S function to the right and after adopting a denotation:

$$w - S(X_1, X_2, \dots, X_n) = t \quad (5)$$

conditional equation of the form (4) will take the following form:

$$\sum_{i=1}^n \frac{\partial S}{\partial X_i} dx_i = t \quad (6)$$

Denoting the partial derivatives in the above relation by b_i , any conditional equation for the coordinates of the boundary markers takes the following symbolic form:

$$\sum_{i=1}^n b_i \times dx_i = t \quad (7)$$

In the geodetic and legal activities related to real estate management, the functional condition for the coordinates of markers in closed figures is most often used, which defines the surface area of the plot.

Using a Gaussian model for a double area of the closed figure in the following form:

$$2S = \sum_{i=1}^n (x_{i+1} - x_{i-1}) y_i \quad (8)$$

it can be concluded that the surface area S of each plot is uniquely identified by the coordinates of the boundary markers.

If we make a differentiation of the function (8) against the individual coordinates of the boundary markers, then we obtain a linear form of this condition in the form of the following dependence:

$$\sum_{i=1}^n (x_{i+1} - x_{i-1}) dy_i + \sum_{i=1}^n (y_{i-1} - y_{i+1}) dx_i = 2S - \sum_{i=1}^n (x_{i+1} - x_{i-1}) y_i \quad (9)$$

Condition (9) includes the coordinates (x_i, y_i) of the boundary markers determined on the basis of surveys, as well as differentials dx_i, dy_i to these coordinates, which are of a random nature.

The surface area S of the plot is determined in the land register or in the documentation derived from geodetic and legal actions.

The solution of the system of conditional equations in the form of (9), which leads to the estimation of the differentials to the coordinates of the boundary markers, and then to determining the model values of these coordinates, constitutes the main objective of the problem posed in this paper.

3. Estimation of Model Coordinates of Border Markers

Let the values of X_1, X_2, \dots, X_n , representing the coordinates of the boundary markers determined by surveying, meet the Gauss–Markov model in its simplest form:

$$\hat{\mathbf{X}} = \mathbf{IX} + \mathbf{d}\hat{\mathbf{X}} \quad (10)$$

which means that to each value of the coordinate X_1 of the boundary marker corresponds the model value of this coordinate, decreased by a random component dx_i .

We assume that on the coordinates of the boundary markers functional conditions are imposed in the form of (9), which should be clearly satisfied by the model values \hat{X}_1 .

When the functional conditions appear in the linear form (7), their form in matrix notation will be as follows:

$$\mathbf{B} \times \mathbf{dX} = \mathbf{t} \quad (11)$$

Estimation of differentials to the coordinates of the markers is carried out by OLS, taking into account the Lagrange function of the following form [1]:

$$\Psi = (\mathbf{X} - \mathbf{I}\hat{\mathbf{X}})^T \times (\mathbf{X} - \mathbf{I}\hat{\mathbf{X}}) + 2\mathbf{K}^T (\mathbf{B} \times \mathbf{dX} - \mathbf{t}) = \mathbf{dX}^T \times \mathbf{dX} + 2\mathbf{K}^T (\mathbf{B} \times \mathbf{dX} - \mathbf{t}) = \min. \quad (12)$$

Necessary conditions for the existence of a minimum of this function take the form:

$$\frac{\partial \Psi}{\partial \mathbf{X}} = 2 \times \mathbf{dX} + 2\mathbf{B}^T \times \mathbf{K} = \mathbf{0} \quad \Rightarrow \quad \mathbf{dX} + \mathbf{B}^T \times \mathbf{K} = \mathbf{0} \quad (13)$$

$$\frac{\partial \Psi}{\partial \mathbf{K}} = 2\mathbf{B} \times \mathbf{dX} - 2\mathbf{t} = \mathbf{0} \quad \Rightarrow \quad \mathbf{B} \times \mathbf{dX} = \mathbf{t} \quad (14)$$

The system of equations (13) and (14) can be written in a block form:

$$\begin{bmatrix} \mathbf{I} & \mathbf{B}^T \\ \mathbf{B} & \mathbf{0} \end{bmatrix} \times \begin{bmatrix} \mathbf{dX} \\ \mathbf{K} \end{bmatrix} = \begin{bmatrix} \mathbf{0} \\ \mathbf{t} \end{bmatrix} \quad (15)$$

The solution of the above system of equations requires the determination of the inverse of the block matrix, which shall be saved in the following form:

$$\begin{bmatrix} \mathbf{I} & \mathbf{B}^T \\ \mathbf{B} & \mathbf{0} \end{bmatrix}^{-1} = \begin{bmatrix} \mathbf{C}_1 & \mathbf{C}_2 \\ \mathbf{C}_3 & -\mathbf{C}_4 \end{bmatrix} \quad (16)$$

Following the execution of multiplication, which leads to the inverse of the block matrix, which is:

$$\begin{bmatrix} \mathbf{I} & \mathbf{B}^T \\ \mathbf{B} & \mathbf{0} \end{bmatrix} \times \begin{bmatrix} \mathbf{C}_1 & \mathbf{C}_2 \\ \mathbf{C}_3 & -\mathbf{C}_4 \end{bmatrix} = \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{I} \end{bmatrix} \quad (17)$$

we obtain a system of four matrix equations, which leads to the following dependences:

$$\begin{bmatrix} \mathbf{C}_1 & \mathbf{C}_2 \\ \mathbf{C}_3 & -\mathbf{C}_4 \end{bmatrix} = \left[\begin{array}{c|c} \mathbf{I} - \mathbf{B}^T (\mathbf{B}\mathbf{B}^T)^{-1} \mathbf{B} & \mathbf{B}^T (\mathbf{B}\mathbf{B}^T)^{-1} \\ \hline (\mathbf{B}\mathbf{B}^T)^{-1} \mathbf{B} & -(\mathbf{B}\mathbf{B}^T)^{-1} \end{array} \right] \quad (18)$$

Having taken into account (18), the sought estimators can be calculated by the following equation:

$$\begin{bmatrix} \mathbf{dX} \\ \mathbf{K} \end{bmatrix} = \begin{bmatrix} \mathbf{C}_1 & \mathbf{C}_2 \\ \mathbf{C}_3 & -\mathbf{C}_4 \end{bmatrix} \times \begin{bmatrix} \mathbf{0} \\ \mathbf{t} \end{bmatrix} \quad (19)$$

Estimator of the vector of unknowns (differentials to the coordinates of the boundary markers) is expressed by the following formula:

$$\mathbf{dX} = \mathbf{C}_2 \times \mathbf{t} = \mathbf{B}^T (\mathbf{B}\mathbf{B}^T)^{-1} \times \mathbf{t} \quad (20)$$

The estimator determined in such a way has this property that the sum of the squares of the differentials to the coordinates of the boundary markers takes the smallest value.

The covariance matrix of the vector (\mathbf{dX}) of the differentials to the coordinates of the boundary markers will be determined by the formula:

$$\text{Cov}(\mathbf{dX}) = \sigma^2 (\mathbf{I} - \mathbf{C}_1) = \sigma^2 \mathbf{B}^T (\mathbf{B}\mathbf{B}^T)^{-1} \mathbf{B} \quad (21)$$

On the basis of the matrix (21) it is possible to determine the variances for the surface area of the analyzed plots.

Assuming a linear form (9) of the vector of the surface area of the plot, the variance of the surface area is expressed by the formula:

$$\sigma^2(S) = \mathbf{F}_s \times \text{Cov}(\mathbf{dX}) \times \mathbf{F}_s^T \quad (22)$$

In this case, the matrix \mathbf{F}_s corresponds to matrix \mathbf{B} .

The value of the standard deviation estimated from the above dependence is the basis to calculate the reliability coefficient of the determination or regulation of the property boundaries, or:

$$\lambda_s = \frac{\sigma(S)}{S} \quad (23)$$

To assess the reliability of the determination or defining property boundaries, the following coefficient degrees are proposed:

- $(1 - \lambda_s) > 0.99$ – very high reliability,
- $0.99 \geq (1 - \lambda_s) > 0.98$ – high reliability,
- $0.98 \geq (1 - \lambda_s) > 0.97$ – relatively high reliability,
- $0.97 \geq (1 - \lambda_s) > 0.96$ – sufficient reliability,
- $0.96 \geq (1 - \lambda_s) > 0.95$ – acceptable reliability,
- $0.95 \geq (1 - \lambda_s)$ – reliability unacceptable – surveying and estimation of model coordinates of the boundary markers must be repeated.

4. Verification of the Proposed Algorithm against a Numerical Example

In order to verify the proposed formulas to estimate the model coordinates of the boundary markers, a property was analyzed, which consists of three parcels, determined by nine boundary markers.

For the markers 3, 5, 7 and 8, error-free coordinate values were assumed. The coordinates of the markers 1, 2, 4 and 6 have the approximate coordinates, the value of which can be changed by a random component dx_i . The surface area of the analyzed plots calculated from the coordinates amounts to 5050, 4725 and 9060 m^2 , respectively. The analyzed situation has been presented in the following figure 1.

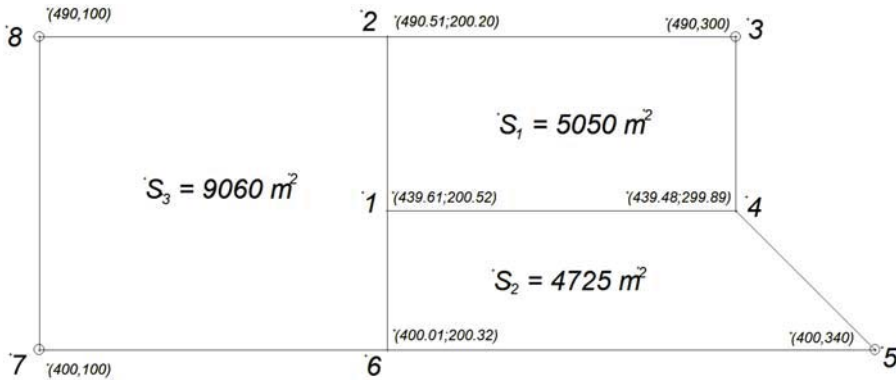


Fig. 1. Configuration of the analyzed record parcels

Surface areas of the plots as defined during establishing of the land and buildings register in the seventies of the last century and disclosed in the land register are:

$$\begin{aligned} \hat{S}_1 &= 5000 m^2, \\ \hat{S}_2 &= 4800 m^2, \\ \hat{S}_3 &= 9000 m^2. \end{aligned}$$

For the purpose of this article, these surface areas were adopted as reference areas. Such an assumption is needed in order to determine the reliability of surface areas of the plots disclosed in the land and buildings register. The double difference between the surface areas disclosed in the land and buildings register and these calculated on the basis of the coordinates are as follows:

$$\begin{aligned} 2\hat{S}_1 - \sum_i (x_{i+1} - x_{i-1})y_i &= 100 m^2, \\ 2\hat{S}_2 - \sum_i (x_{i+1} - x_{i-1})y_i &= -150 m^2, \\ 2\hat{S}_3 - \sum_i (x_{i+1} - x_{i-1})y_i &= 120 m^2. \end{aligned}$$

Conditional equations in the form:

$$\sum_{i=1}^n (x_{i+1} - x_{i-1}) dy_i + \sum_{i=1}^n (y_{i-1} - y_{i+1}) dx_i = 2\hat{S} - \sum_{i=1}^n (x_{i+1} - x_{i-1}) y_i$$

take the following form:

$$99.69dx_1 + 51.03dy_1 - 99.48dx_2 + 50.39dy_2 + 99.48dx_4 - 50.39dy_4 = 100,$$

$$-99.57dx_1 + 39.47dy_1 - 139.48dx_4 - 39.61dy_4 + 139.48dx_6 + 39.61dy_6 = -150,$$

$$-0.12dx_1 - 90.50dy_1 - 100.52dx_2 - 50.39dy_2 + 100.52dx_6 - 39.61dy_6 = 120.$$

In the matrix notation the above equations are as follows:

$$\begin{bmatrix} 99.69 & 51.03 & -99.48 & 50.39 & 99.48 & -50.39 & 0 & 0 \\ -99.57 & 39.47 & 0 & 0 & -139.48 & -39.61 & 139.48 & 39.61 \\ -0.12 & -90.50 & -100.52 & -50.39 & 0 & 0 & 100.52 & -39.61 \end{bmatrix} \times \begin{bmatrix} -100 \\ 150 \\ -120 \end{bmatrix}.$$

[B] **[t]**

Using the formulas [1]:

$$\mathbf{B}^T (\mathbf{B}\mathbf{B}^T)^{-1} \mathbf{B} \Rightarrow \mathbf{Q},$$

$$\mathbf{I} - \mathbf{Q} \Rightarrow \mathbf{C},$$

$$d\mathbf{X} = \mathbf{C}_2 \times \mathbf{t} = \mathbf{B}^T (\mathbf{B}\mathbf{B}^T)^{-1} \times \mathbf{t},$$

elements of the following matrixes have been calculated:

$$\mathbf{B}^T (\mathbf{B}\mathbf{B}^T)^{-1} = \begin{bmatrix} 0.002064 & -0.001117 & 0.000122 \\ 0.002934 & 0.002438 & -0.003706 \\ -0.002758 & -0.000572 & -0.002696 \\ 0.002088 & 0.001110 & -0.002036 \\ 0.001500 & -0.002126 & 0.000451 \\ -0.002296 & -0.001700 & 0.000665 \\ 0.001258 & 0.002698 & 0.002245 \\ 0.000822 & 0.001318 & -0.001651 \end{bmatrix},$$

$$d\mathbf{X} = \begin{matrix} dx_1 & dy_1 & dx_2 & dy_2 & dx_4 & dy_4 & dx_6 & dy_6 \\ 0.39 & -0.52 & -0.51 & -0.20 & 0.52 & 0.11 & -0.01 & -0.31' \end{matrix}$$

$$\mathbf{B}^T (\mathbf{B}\mathbf{B}^T)^{-1} \mathbf{B} = \begin{bmatrix} 0.3170 & 0.0502 & -0.2176 & 0.0979 & 0.3612 & -0.0598 & -0.1436 & -0.0491 \\ 0.0502 & 0.5814 & 0.0807 & 0.3346 & -0.0482 & -0.2444 & -0.0325 & 0.2434 \\ -0.2176 & 0.0807 & 0.5453 & -0.0031 & -0.1946 & 0.1616 & -0.3507 & 0.0841 \\ 0.0979 & 0.3346 & -0.0031 & 0.2078 & 0.0529 & -0.1492 & -0.0497 & 0.1246 \\ 0.3612 & -0.0482 & -0.1946 & 0.0529 & 0.4458 & 0.0086 & -0.2513 & -0.1021 \\ -0.0598 & -0.2444 & 0.1616 & -0.1492 & 0.0086 & 0.1830 & -0.1702 & -0.0937 \\ -0.1436 & -0.0325 & -0.3507 & -0.0497 & -0.2513 & -0.1702 & 0.6020 & 0.0180 \\ -0.0491 & 0.2434 & 0.0841 & 0.1246 & -0.1021 & -0.0937 & 0.0180 & 0.1176 \end{bmatrix},$$

The covariance matrix for the estimated parameters is calculated by the formula:

$$\text{Cov}(d\hat{\mathbf{X}}) = \sigma_0^2 \times \mathbf{B}^T (\mathbf{B}\mathbf{B}^T)^{-1} \mathbf{B},$$

where σ_0^2 we take at the average variance level of determining the coordinates of the boundary markers.

Assuming $\sigma_0 = 0.05$ m, we obtain:

$$\sigma_0^2 = \frac{S_p [\text{Cov}(x; y)]}{2n} = 0.0025.$$

From the diagonal of the matrix $\text{Cov}(d\hat{\mathbf{X}})$ we obtain a variance (mean squared error) for each coordinate. These errors are as follows:

- $\sigma(dx_1) = 0.03$ m,
- $\sigma(dy_1) = 0.04$ m,
- $\sigma(dx_2) = 0.04$ m,
- $\sigma(dy_2) = 0.02$ m,
- $\sigma(dx_4) = 0.03$ m,
- $\sigma(dy_4) = 0.02$ m,
- $\sigma(dx_6) = 0.04$ m,
- $\sigma(dy_6) = 0.02$ m.

The calculation of the variance of each plot surface areas was carried out according to the formula:

$$\text{Cov}(S) = \sigma_0^2 F_s B^T (BB^T)^{-1} B F_s^T.$$

Hence the covariance matrix for the surface areas of the analyzed plots takes the following form:

$$\text{Cov}(S_1, S_2, S_3) = \sigma_0^2 \begin{bmatrix} 37413.0 & -19791.5 & 2830.4 \\ -19791.5 & 53519.3 & 8891.5 \\ 2830.4 & 8891.5 & 32506.9 \end{bmatrix}.$$

The standard deviations of the surface areas of each plot shall have the values:

$$\sigma(S_1) = 9.7 \text{ m}^2,$$

$$\sigma(S_2) = 11.6 \text{ m}^2,$$

$$\sigma(S_3) = 9.0 \text{ m}^2.$$

The reliability of the surface areas of the analyzed plots with the assumed error of the marker position at the level of $\sigma_0 = 0.05 \text{ m}$ is so high because $1 - \lambda_s > 0,99$.

With a limit inaccuracy of determining the boundary marker coordinates at the level of $\sigma_0 = 0.45 \text{ m}$ (mean value for the attribute BPP = 3 from the range of 0.31–0,60 m), $\sigma_0^2 = 0.2025$ we obtain:

$$\sigma(S_1) = 87.0 \text{ m}^2,$$

$$\sigma(S_2) = 104.1 \text{ m}^2,$$

$$\sigma(S_3) = 81.1 \text{ m}^2.$$

With such parameters, reliability assessment is as follows:

$$1 - \lambda_{s_1} = 0.983 - \text{high reliability},$$

$$1 - \lambda_{s_2} = 0.978 - \text{relatively high reliability},$$

$$1 - \lambda_{s_3} = 0.991 - \text{very high reliability}.$$

5. Completion and Conclusions

The article presents the concept of real estate cadastre attribute estimation using the conditional model. Certain conditions for the surface area of the plot result from a discrepancy in the value of the surface specified in the system of land and buildings register. The surface area disclosed in the land register often does not result from analytical calculations and it is only a surface calculated upon establishing the land register by graphical or mechanical methods, and in rural areas rounded to a full are [3, 6].

A proposal of assessing the reliability of determining the surface area of plots presented in this article is simplified. Assumption of the correctness of certain markers as well as the scale of assessing the reliability were defined for the purpose of the conducted analyses. However, the accepted manner of evaluation can be used both as a method of defining an attribute used to describe record parcels, as well as the factor determining the uncertainty relating to the surface area of the plot disclosed in the land and building register system. Indirectly, it can be used in real estate transactions as an attribute applied to property valuation.

Low reliability of the given surface area of a plot informs its potential buyer of the need to carry out activities aimed at determining the location of the boundary markers. In some cases, such information may be crucial for the actions that the investor will be forced to take. The location of the boundary markers determines the subsequent location of an object in the field, which in the case of small or narrow plots with boundary markers whose coordinates have been determined with insufficient accuracy – can be problematic.

References

- [1] Czaja J.: *Modele statystyczne w informacji o terenie*. Skrypty Uczelniane AGH nr 1465, Wydawnictwa AGH, Kraków 1996.
- [2] Hycner R.: *Podstawy katastru*. Uczelniane Wydawnictwa Naukowo-Dydaktyczne AGH, Kraków 2004.
- [3] Hanus P.: *Ocena przydatności dokumentacji byłego katastru austriackiego dla potrzeb prac geodezyjnych*. AGH, Kraków 2006 (PH.D. thesis).
- [4] *Rozporządzenie Ministra Rozwoju Regionalnego i Budownictwa z dnia 29 marca 2001 roku w sprawie ewidencji gruntów i budynków*. Dz.U. z 2001 r. Nr 38, poz. 454.
- [5] *Ustawa z dnia 17 maja 1989 roku – Prawo geodezyjne i kartograficzne*. Dz.U. z 2010 r. Nr 193, poz. 1287 z późn. zm.
- [6] *Zarządzenie Ministrów Rolnictwa i Gospodarki Komunalnej z dnia 20 lutego 1969 roku w sprawie ewidencji gruntów*. M.P. z 1969 r. Nr 11, poz. 98.