Andrzej Kwinta*

Outlier Identification Method for Horizontal Strain on the Mining Areas

1. Introduction

The results of geodetic measurements provide information regarding the course of deformation process occurring because of the mining activity. Every analysis of deformations and theoretical considerations base on the results of measurements. The most important indicator of deformation taken into account in the analysis of adverse influence of the deformation on the objects on the surface is horizontal strain. Analysis of the distribution of the deformations determined by geodesic methods on the measurement lines indicates large oscillation of the recorded values. It is connected with occurrence of many factors disturbing deformation process. An image of deformations shows all types of errors occurring in the problems connected with measurement results' analysis [1]. Part of errors appearing during measurement can be eliminated during this measurement or during initial analysis of the results. However, there are many factors that cannot be unequivocally identified what causes lasting disturbance of the image of deformation process.

In case of dynamic courses of the process observed with use of geodesic methods a problem of error identification appears. A basic problem is to assume a criterion which would allow unequivocally eliminating values detached from deformation image disturbed by accidental factors. Measuring and working on the results are connected with errors (accidental, systematic, instrument, personal, etc.). In case of observing deformation process an obtained image is affected not only by the factors connected with realization of measurements but also by a whole group of factors disturbing the process itself. That is why it seems justified to change the nomenclature from errors to disturbances.

* Department of Geodesy, University of Agriculture, Krakow

Process of eliminating gross disturbances or outlier values for the horizontal strain should be performed at first, after equalization of the observation and determination of the indicator's value. In certain cases, after drawing a graph one can immediately identify outlier or "suspicious" values. However, one can often meet the following problem: if this particular disturbance is accidental or if it is just a mistake? There is no unequivocal answer to that question and every such case should be considered individually. The present work presents the solution allowing estimation of the disturbances in order to eliminate outlier values from the image of deformation process. Proposed solution bases on the theoretical considerations which were in compliance with Knothe theory [5] and determined value of random distribution of the process course [8]. Presented theoretical considerations are applied in calculations performed for the practical example.

2. Types of Disturbances

Image of the strain caused by underground mining activity is influenced by many factors that cannot be unequivocally identified. Various types of disturbing factors affect the source (mining exploitation), a medium (rock mass, foundation of points) and applied technologies of the measurement. Because of that the disturbances can be considered in terms of their type or features of the given disturbance's source. Basic types of the disturbances are presented later in this work.

Accidental disturbances affect the image of strain in the most significant way. Deformation process, caused by underground mining activity, is a classic deterministic-probabilistic process. Mining activity determines deformation process and large number of the accidental factors causes significant disturbances of the deformations of this process. It results in creation of the substantial scattering values. Basing on the many years of research the deformation indicators' values of random distribution were established [7, 8].

Systematic disturbances which (in assumption) can be taken into account thanks to proper mathematical formulas, significantly affect deformation indicators at mining areas as well. The systematic errors become particularly significant in reference to the horizontal and vertical dislocations. Error of tie, i.e. assuming measurement reference point in relative measurements as constant point while it undergoes deformation movements can cause very serious formal consequences. Introduction of the GPS system caused limitation of the errors of tie and possibility of performing absolute measurements [4].

Gross disturbances (deviations, mistakes) can be usually identified by repeating measurements (levelling measurement with loosening an instrument). In

the classic statistical results analysis the appropriate statistical tests are performed [6] what allows identifying values of the gross errors. However, in such cases it is necessary to repeat the measurements many times and assume standard distribution of the obtained results. In case of deformations on the mining areas it is connected with necessity of approximating the image with some theoretical curve.

3. Criterion for Estimating Outlier Values

Before determining identification criterion for the outlier values, let us introduce a few formal assumptions. Let us assume, that Knothe-Budryk [5] theory describes deformation process caused by the underground mining activity in the satisfying way. The values of accidental disturbance (i.e. parameters of the random scattering) are known for certain mining-geologic conditions for the activity.

Let us introduce the following definition: *If a difference of horizontal strains determined on two neighbouring sections of measurement line exceeds established value of a criterion then an outlier value of a strain appears and is to be removed.*

$$\Delta\varepsilon_{i,j+1} > K_\varepsilon \qquad i = 1, 2, \ldots, n-1 \qquad (1)$$

where:

$\Delta\varepsilon_{i,j+1}$ – strain difference on the two neighbouring measurement sections,
K_ε – established estimation criterion of the outlier values.

Let us determine horizontal strain along selected calculation profile according to the Knothe–Budryk theory. Let us take into account exploitation field generating flat strain, i.e. so called "infinite halfplane" (Fig. 1).
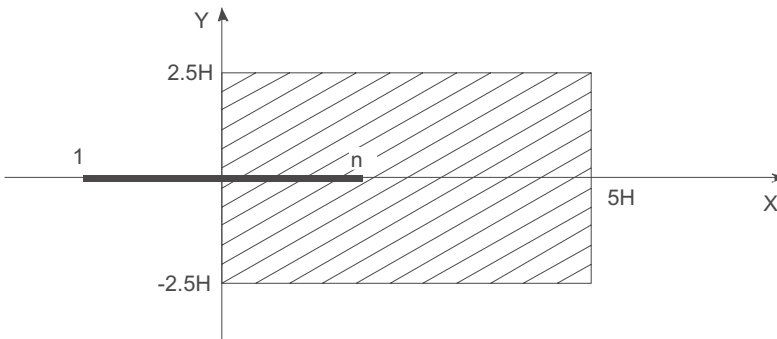

Fig. 1. Diagram of exploitation field assumed in the calculations

According to the figure 1, the formula for horizontal strain in Knothe–Budryk theory, for the selected line, can be written as follows

$$\varepsilon(x) = \frac{-2\pi w_{max} B}{r^2} \frac{x}{r} \exp\left(-\pi \frac{x^2}{r^2}\right) \tag{2}$$

where:

$\varepsilon(x)$ – horizontal strain,

w_{max} – max. subsidence,

$B = -\dfrac{r}{\sqrt{2\pi}}$ – factor of proportionality between vertical and horizontal deformations,

r – radius of Knothe theory's range.

In order to simplify the notation let us assume the following

$$\varepsilon_0 = \frac{-2\pi w_{max} B}{r^2} \tag{3}$$

Using (2) and (3), let us determine first derivative of the deformations

$$\frac{d\varepsilon(x)}{dx} = \varepsilon_0 \left[\frac{1}{r}\exp\left(-\pi\frac{x^2}{r^2}\right) + \frac{x}{r}\left(\frac{-2\pi x}{r^2}\right)\exp\left(-\pi\frac{x^2}{r^2}\right)\right] \tag{4}$$

Taking into account definition of function's derivative, using (4) we obtain formula of an increment of deformation between two points (on the section of lenght b) that has the following form

$$\Delta\varepsilon(x) = \varepsilon_0 \frac{b}{r}\left(1 - \frac{2\pi x}{r^2}\right)\exp\left(-\pi\frac{x^2}{r^2}\right) \tag{5}$$

In order to determine maximum increment of the deformation in the given exploitation conditions let us look for maximum value using (5). First derivative can be written in the following way

$$\frac{d\Delta\varepsilon(x)}{dx} = \varepsilon_0 \frac{b}{r}\left[\left(1 - \frac{2\pi x}{r^2}\right)\left(-\frac{2\pi x}{r^2}\right)\exp\left(-\pi\frac{x^2}{r^2}\right) + \left(\frac{4\pi x}{r^2}\right)\exp\left(-\pi\frac{x^2}{r^2}\right)\right] \tag{6}$$

According to the necessary and sufficient conditions for existence of function's (5) extremum, after comparing dependence (6) to zero and defining variation of derivative signs one can determine maximum which is equal to

$$\Delta\varepsilon_{max} = \Delta\varepsilon(x=0) = \frac{-2\pi w_{max}}{r^2}\frac{B}{r}\frac{b}{r} \tag{7}$$

According to the classic Knothe–Budryk theory, assuming factor of proportionality B and taking into account that [3]:

$$T_{max} = \frac{w_{max}}{r} \tag{8}$$

and

$$\varepsilon_{max} = 0.6T_{max} \tag{9}$$

then

$$\Delta\varepsilon_{max} \approx 1.5\frac{b}{r}\varepsilon_{max} \tag{10}$$

The maximum value of strain increment given above (10) is connected with performed theoretical considerations. If the deformation process would run without accidental disturbances, the dependence (10) could be the criterion for elimination of the outlier values. However, in case of real processes one has also to take into account accidental dislocations (random scattering) appearing in the image [2]. Let us assume coefficient of variation for the deformations in the following form [7]

$$M_\varepsilon = \frac{\sigma_\varepsilon}{\varepsilon_{max}}100\% \tag{11}$$

Thus, the coefficient of random scattering is a ratio of standard deviation to the maximum value of the indicator. For many years this coefficient has been investigated and there are many publications in which its value was defined for average conditions of mining exploitation [8]. Assuming that random disturbance is constant in the whole observed profile, the increment of strain should meet the following inequality

$$\left|\Delta\varepsilon_{i,i+1}^p\right| + 2\sigma_\varepsilon \tag{12}$$

where $\Delta\varepsilon_{i,i+1}^p$ means increment of average deformations along selected profile.

Thus, the final criterion for estimation of the outlier horizontal strain basing on the dependencies (10)–(12) will take the following form

$$K_\varepsilon = \left(1.5\frac{b}{r} + 2M_\varepsilon\right)\varepsilon_{max} \tag{13}$$

In this way the criterion for filtering the outlier values of horizontal strain was defined and it is taken into account in the computational algorithm.

4. Computational Algorithm

Criterion for identification of the outlier values of the horizontal strain (presented in the previous chapter) can be used to build appropriate computational algorithm. Obtained solution can be a base for initial analysis of the calculated results before taking into account some theoretical model. It can be independent criterion as well.

Formed computational algorithm will have an iterative character. Following steps take into account detailed data related to the real course of the process. Realization of the presented algorithm can be listed as follows:

1. Preparation of data: calculating increments of strain along measurement line, determining criterion according to the dependence (13).
2. Identification of the outlier values of strain.
3. Removing outlier values.
4. Approximation of the horizontal strain with theoretical model – determining random scattering (M_ε).
5. Repeating 1–4 until all of the outlier values are removed.

According to the presented algorithm the calculations for the practical example were performed.

5. Computational Example

Presented considerations should be verified by practical example. For that purpose a measurement line consisting of 54 measurement points (53 sections) was chosen. Average length between points was 48 m and total length of the line reached approximately 2500 m. At depth of about 1100 m under measurement line a mining exploitation is located. It causes compressing and tensile strains. Maximum value of the strain was –2.60 mm/m. Figure 2 presents horizontal strain on the analyzed measurement line.

According to the presented theoretical considerations and proposed iterative computational algorithm the outlier disturbances were determined. Basing on the measured deformations their increments were calculated (Fig. 3). Initial value of the identification criterion for the outlier values was assumed taking into account that random scattering for the horizontal strain equals 20% of the maximum value. According to that the beginning value of the criterion was equal

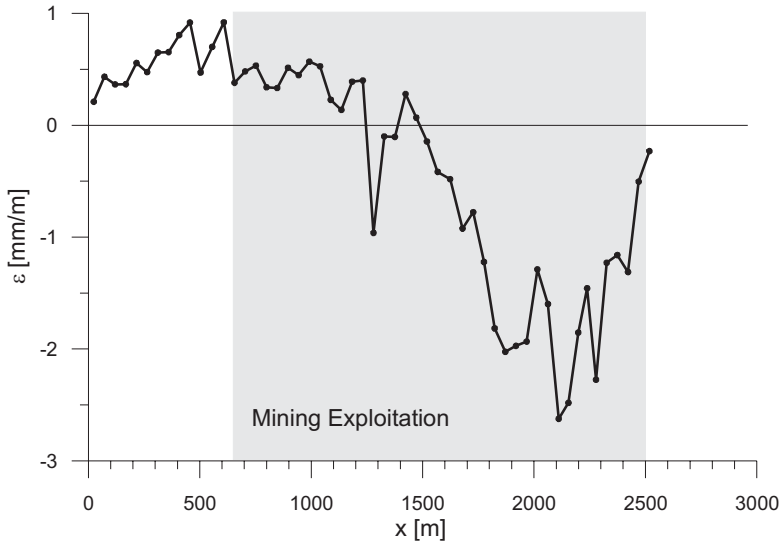$$K_\varepsilon^0 = \pm 1.3 \text{ mm/m}.$$

Fig. 2. Distribution of the horizontal strain along measurement line

In figure 3 a graph of strain increment and identification criterion for the outlier values were marked. Comparing obtained results one has to state that in one case the assumed criterion was exceeded.
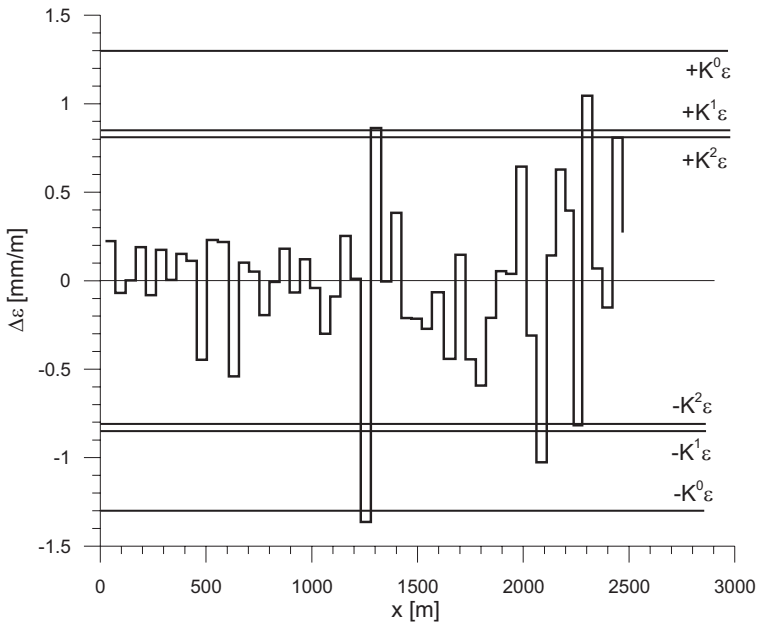

Fig. 3. Distribution of the strain increments and elimination criterion for the outlier values

After discarding the outlier value an approximation of the distribution of horizontal deformations was performed using orthogonal polynomials. For that purpose a proprietary software "wielort.exe" was used. The computations resulted in obtaining value of standard deviation for the analyzed case which was used in next iteration. Value of the criterion was updated:

$$K_\varepsilon^1 = \pm 0.83 \text{ mm/m.}$$

Thanks to the computations two outlier values were removed (Fig. 3). After next iteration the criterion's value was calculated to be at the level of

$$K_\varepsilon^2 = \pm 0.81 \text{ mm/m.}$$

One outlier value, exceeding assumed criterion, was removed from strains distribution. Next iterations introduced no changes. Finally, the computations resulted in removing 4 values of strain exceeding assumed identification criterion of the outlier values. Graphical representation of the results is presented in figure 4.
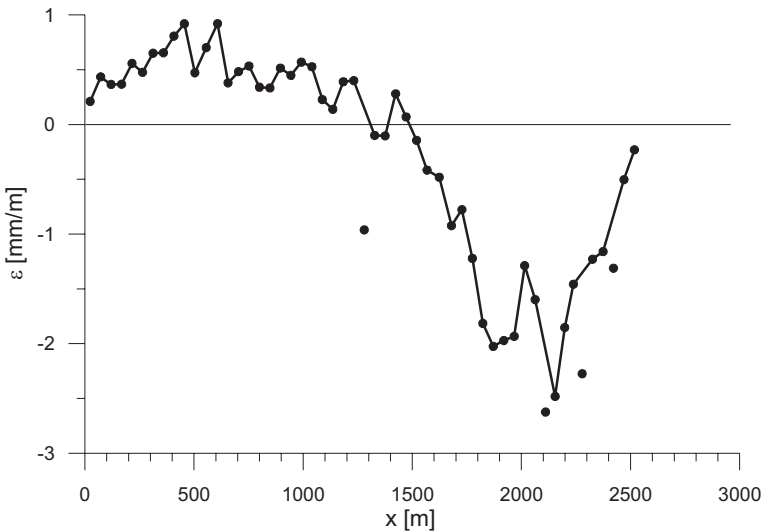

Fig. 4. Distribution of the horizontal strains after removing outlier values

6. Summary

The course of strain indicators recorded by geodesic measurements is highly disturbed. Large amount of factors causing different types of disturbances (accidental, systematic and gross errors) results in significant complication of the de-

formations analysis. Identification of "gross errors" is particularly important in case of large values of random scattering. An image of deformation process without elimination of the outlier values disturbances can result in incorrect conclusions regarding course of the processes. Thus, it is important to introduce criterion allowing elimination of the outlier values.

The present work proposes the criterion of discarding outlier values for horizontal strains, build basing on the maximum increment of the deformations and random scattering of a phenomenon. Presented iterative, computational algorithm was tested in practical example giving satisfying results. It seems to be justified to perform further analyses, to verify the solution in larger number of practical examples and to extend the considerations to other indicators of deformation.

References

[1] Barry B.A.: *Errors in Practical Measurement in Science, Engineering, and Technology*. John Wiley & Sons, New York 1978.

[2] Batkiewicz W.: *The Standard Deviations of Orogen Deformations Subsequent to Exploitation*. Works of Mining-Geodesy Commision PAN in Cracow, Geodesy 10, 1971 (in Polish).

[3] Budryk W.: *Wyznaczanie wielkości poziomych odkształceń terenu*. Archiwum Górnictwa i Hutnictwa, t. 1, z. 1, 1953.

[4] Hejmanowski R., Kwinta A.: *Implementation of GPS satelitary technique for monitoring of point displacements on mining areas*. Proceedings of Second World Mining Environment Congress, vol. 1, 13–16.05.1997, pp. 351–369.

[5] Knothe S.: *Prediction of Influence of Mining Exploitation*. Publ. "Śląsk", Katowice 1984 (in Polish).

[6] Konieczka P., Namieśnik J.: *Quality Assurance and Quality Control in the Analytical Chemical Laboratory: A Practical Approach*. CRC Press, Boca Raton – Londyn – New York 2009.

[7] Popiołek E.: *Statistical Dispersion of Horizontal Deformations of an Area In the Light of Geodetic Observations of the Effects of Mining Exploitation*. AGH University of Science and Technology, Geodesy, vol. 44, 1976 (postdoctoral degree thesis, in Polish).

[8] Stoch T.: *Impact of Geological and Mining Conditions on Stochastic Processes of Terrain Surface Deformation*. AGH University of Science and Technology, Krakow 2005 (Ph.D. thesis, in Polish).