Calibration of precipitation estimation algorithm with particular emphasis on the Pomeranian region using high performance computing

Tomasz BIELIŃSKI, Andrzej CHYBICKI

Gdańsk University of Technology
Faculty of Electronics, Telecommunications and Informatics
Gabriela Narutowicza 11/12 80-233 Gdańsk, Poland
tomasz.bielinski@pg.gda.pl, andrzej.chybicki@eti.pg.gda.pl

Fast and accurate precipitation estimation is an important element of remote atmosphere monitoring, as it allows, for example, to correct short-term weather forecasts and the prediction of several types of meteorological threats. The paper presents methodology for calibrating precipitation estimation algorithm based on MSG SEVIRI sensor data, and Optimal Cloud Analysis product available via EumetCast transmission. Calibration is performed in a predefined coastal zone area, and utilizes the parallelized gradient computing method. In order to perform and validate results of estimation, reference weather radar data was provided by the Meteorology and Hydrology Research Institute (pol. Instytut Meteorologii i Gospodarki Wodnej - IMGW). The research was conducted using the Tryton supercomputer - the high performance computing environment of Gdansk University of Technology.

Keywords: SEVIRI, MSG, Gradient method, weather radars

1. Introduction

Multispectral satellite observations allow for monitoring of various phenomena related to the atmosphere, and the Earth’s surface. Appropriate processing of the data acquired by spaceborn sensors is of high interest, as it ensures appropriate accuracy of the obtained observations, and allows for estimating various physical quantities of environment systems. Therefore, in order to increase the accuracy of obtained measurements, methods of atmospheric, geometric and radiometric corrections are usually applied in the post-processing phase of the satellite data flow chain. However, appropriate correction of obtained observations is often not sufficient for obtaining the desired accuracy of particular satellite products like: land surface temperature, sea surface temperature, precipitation, or others.

Therefore, one of the possible approaches to deal with this problem is a process known as parameterization, which relies on finding the best possible (for instance, in the sense of mean squared error) estimation that best fits conditions for the analysed set of observations. In the paper, the methodology for calibrating precipitation estimation algorithm based on MSG SEVIRI sensor and Optimal Cloud Analysis product available via EumetCast transmission is presented. Calibration is performed in a predefined coastal zone area. In order to perform and validate results of estimation, reference weather radar data provided by Meteorology and Hydrology Research Institute (IMGW) was utilized. Due to the fact that processing of analyzed data requires large computational and memory resources, that exceed capabilities of standard PC's, the computations for this research were performed in the dedicated high performance computing environment of Gdansk University of Technology, namely, using the Tryton supercomputer [1].

2. Materials and methods

2.1. Algorithm of precipitation estimation

Precipitation estimation, considered in this paper, is based upon two parameters that can be obtained from satellite sensors. The first parameter, Condensed Water Path (CWP) describes the amount of water contained in a cloud column [3], the second parameter Cloud Top Temperature (CTT), and the relation between these two values, is defined as follows [2]:

$$R_E = \frac{c}{H}\left(\frac{CWP \ - CWP_0}{CWP_0}\right)^{\alpha}, \tag{1}$$

where *CWP* is the amount of liquid or solid water per m^2 of cloud [g/m^2] product delivered via EumetCast transmission, *H* is the height of rainfall column [km] retrieved from level 1.5 product [4], *c* is the scaling constant [mm km / h] and CWP_0 is threshold value of *CWP*, determining whether the considered cloud fragment is classified as precipitating [g/m^2],α – exponent.

Height of rainfall column is calculated with following formula:

$$H = \frac{CTT_{max} - CTT}{\gamma} + H_{min}, \tag{2}$$

where CTT is cloud top temperature, CTT_{max} is maximum cloud top temperature in square 128x128 pixels of SEVIRI image, γ is decrease of atmosphere temperature related to altitude (assumed to be constant and equal to 6.5 K/km) and H_{min} is the minimal height of rain column for very thin cloud. Usually this constant is equal to 0.7 km [2].

2.2. Data and sensors

To estimate precipitation using the formula mentioned in the previous section; two types of information are needed. First one is CWP, which can be inferred from cloud properties. In this case, cloud properties product is obtained from Optimal Cloud Analysis (OCA) dataset provided via EumetCast. OCA product is calculated by Eumetsat using data from SEVIRI sensor (Fig. 1).

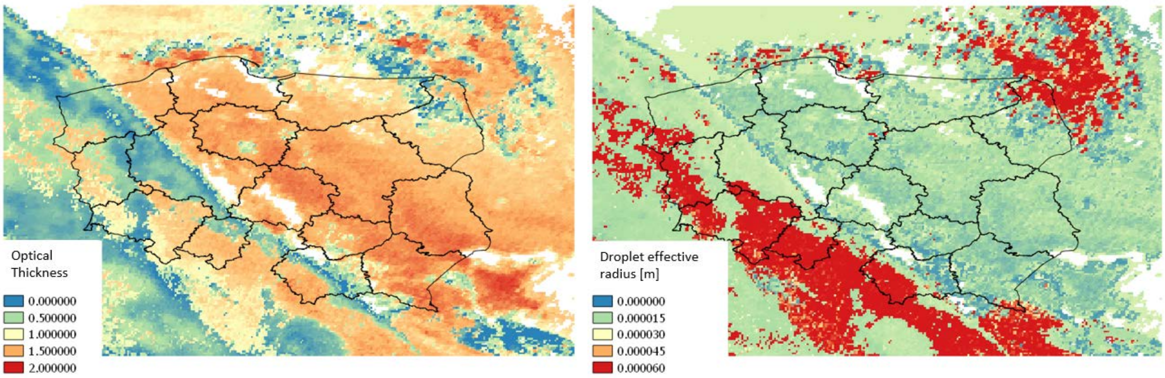

Fig. 1. Optical Thickness (left) and Droplet effective radius (right) as part of Optimal Cloud Analysis Product.

Second input data type is rainfall column height (H). This parameter can be calculated using Cloud Top Temperature (CTT). CTT can be obtained using 10.8 um thermal channel provided from SEVIRI sensor (Fig. 2).

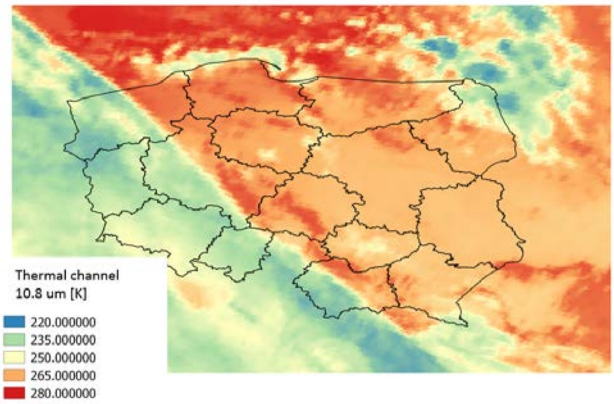

Fig. 2. SEVIRI 10.8 um thermal channel.

As shown above, the precipitation estimation algorithm mentioned in the previous section can be exclusively based on data from SEVIRI sensor [4]. SEVIRI (Spinning Enhanced Visible and Infrared Imager) sensor is an integral part of Meteosat-8, Meteosat-9, Meteosat-10 and Meteosat-11 satellites. Data used in this research is mainly provided by Meteosat-10. Effective resolution of SEVIRI sensor over Poland oscillates about 3.5 km x 7.5 km per pixel. New observations are provided every 15 min.

The calibration process requires reference data. In the case of precipitation estimation algorithm, data from weather radar is used. Radar datasets, provided by IMGW *(pol. Instytut Meteorologii i Gospodarki Wodnej - IMGW)*, cover the entirety of Poland and have a resolution of 1 km x 1 km per pixel. IMGW provides several radar observation products, as well as raw data. For calibration of satellite data, CMAX product, which represents radar reflectivity maximum values from the area of localized meteorological structure, was chosen [5]. Example, weather radar dataset visualization was presented in Fig. 3. Every dataset is produced for 10 min periods.

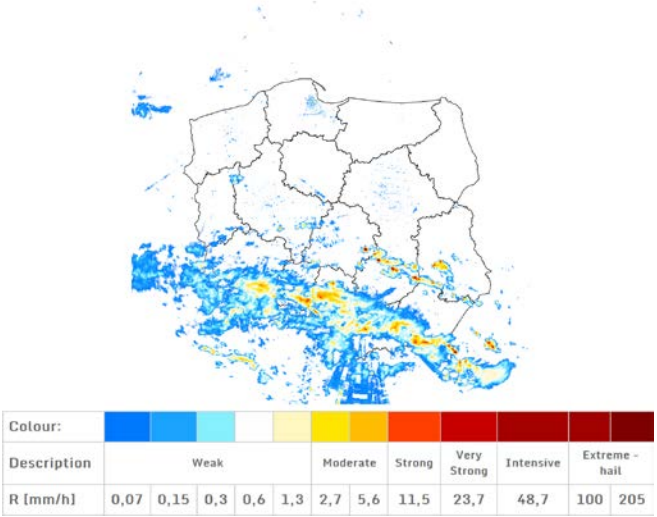

Fig. 3. Weather data CMAX product rescaled to precipitation intensity.

3. Algorithms of parameterisation

One of many methods used to perform formula or algorithm parameterisation is linear regression. This method is well-known, and in general can be characterized by low computational cost. However, standard form of linear regression is suitable for estimating linear regression between two variables. In the case of more complicated formulas, it is insufficient. Therefore, in such situations, the family of gradient method could be a good solution.

3.1. Gradient method

Gradient method is one of many artificial intelligence (AI) algorithms used in the problem of minimization of the goal function. The basic version of this method uses gradient of goal function to calculate new points in the domain, which should be closer to global or local minimum of goal function. This could be expressed by the following equation:

$$\boldsymbol{x}_{n+1} = \boldsymbol{x}_n - a\nabla f(\boldsymbol{x}_n), \tag{3}$$

where f is goal function, x is some point in domain of goal function, and a is a scaling constant.

However, in some situations, convergence of pure gradient method is too weak, therefore, to accelerate the process, a hessian matrix of goal function, obtained by multiplying its reciprocal by the gradient, can be applied. Note that, for some particular areas of domain, a hessian matrix is singular, or near to singularity, and cannot be successfully reversed. In such cases, regularisation similar to the Levenberg method [6] is applied, resulting with the following formula:

$$\boldsymbol{x}_{n+1} = \boldsymbol{x}_n - a\left(\lambda I + \mu\nabla^2 f(\boldsymbol{x}_n)\right)^{-1}\nabla f(\boldsymbol{x}_n), \tag{4}$$

where a, λ and μ are scaling coefficients. Sometimes in research, an extended version of this method was applied:

$$\boldsymbol{x}_{n+1} = \boldsymbol{x}_n - \alpha(\lambda I + \mu \nabla^2 f(\boldsymbol{x}_n) + \phi|\nabla f(\boldsymbol{x}_n)|I)^{-1} \nabla f(\boldsymbol{x}_n), \tag{5}$$

where ϕ is a scaling constant. Gradient normalisation fragment ($\phi|\nabla f(\boldsymbol{x}_n)|I$) was used in calibration test for synthetic data; however, it did not give positive results for real data. Even, in some cases, it made the search process more unstable.

In the case of precipitation estimation, the goal function was defined as follows:

$$E_{\underline{R}} = \frac{\sum_{R \in \underline{R}} E_R}{|\underline{R}|} \tag{6}$$

where:

$$E_R = (R_E(t_R) - R_R(t_R))^2, \tag{7}$$

where E_R – squared error is calculated for each pixel using equation (6),$R_E(t_R)$ is estimated precipitation, and $R_R(t_R)$ is reference precipitation from radar data. Goal function in this form allows to relatively easy designation of its gradient and hessian. Form of gradient and hessian operator was defined as follows:

$$\nabla_P = \left[\frac{\partial}{\partial c} \quad \frac{\partial}{\partial CWP_0} \quad \frac{\partial}{\partial \alpha}\right]^T, \tag{8}$$

$$\nabla_P^2 = \nabla_P \nabla_P^T = \begin{bmatrix} \frac{\delta^2}{\delta c^2} & \frac{\delta^2}{\delta c \delta CWP_0} & \frac{\delta^2}{\delta c \delta \alpha} \\ \frac{\delta^2}{\delta CWP_0 \delta c} & \frac{\delta^2}{\delta {CWP_0}^2} & \frac{\delta^2}{\delta CWP_0 \delta \alpha} \\ \frac{\delta^2}{\delta \alpha \delta c} & \frac{\delta^2}{\delta \alpha \delta CWP_0} & \frac{\delta^2}{\delta \alpha^2} \end{bmatrix}, \tag{9}$$

Application of gradient operator from equation (8) on equation (7) results with:

$$\nabla_P E_{\underline{R}} = \frac{\sum_{R \in \underline{R}} \nabla_P E_R}{|\underline{R}|}, \tag{10}$$

where:

$$\nabla_P E_R = 2(R_E(t_R) - R_R(t_R)) \nabla_P R_R(t_R), \tag{11}$$

And also, applying hessian operator from equation (9) to equation (7) results with:

$$\nabla_P \nabla_P^T E_{\underline{R}} = \frac{\sum_{R \in \underline{R}} \nabla_P \nabla_P^T E_R}{|\underline{R}|}, \tag{12}$$

where:

$$\nabla_P \nabla_P^T E_R = 2\left[\nabla_P R_E(t_R)\nabla_P^T R_E(t_R) + \left(R_E(t_R) - R_R(t_R)\right)\nabla_P \nabla_P^T R_R(t_R)\right], \qquad (13)$$

Because gradient and hessian of precipitation estimation is time and spatialy dependent, as is shown; therefore, they are also computed during the calibration process.

3.2. Data flow and verification methodology

Fig. 4 describes the general, iterative, approach to calibration using gradient method. Stop condition evaluation checks for two sub-conditions. First one is value of goal function; if is lower than assumed threshold, then stop condition is satisfied. Second sub-condition checks if the maximum number of iterations has been reached.

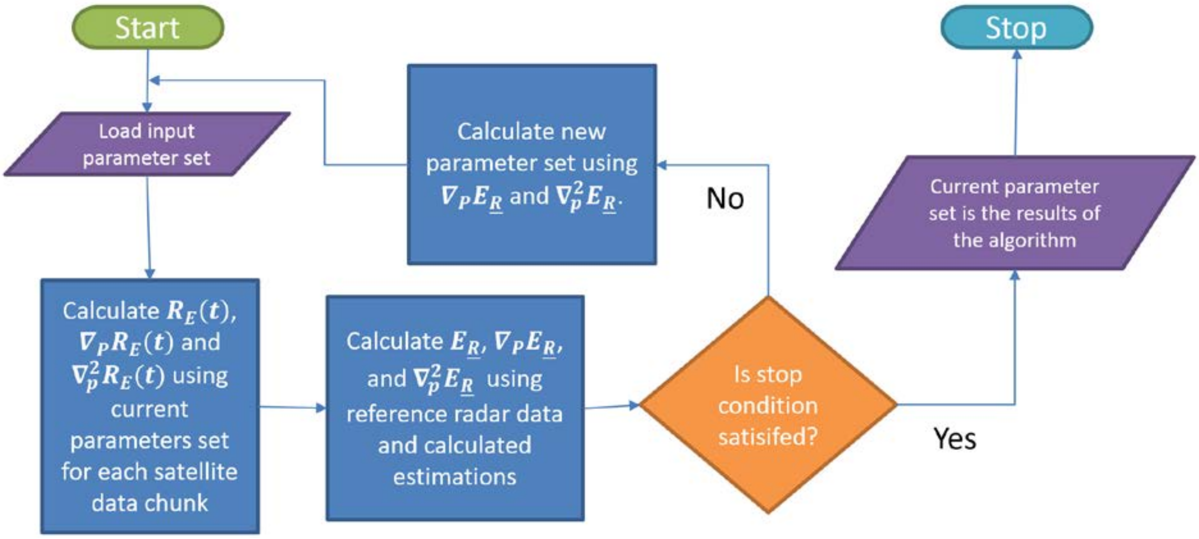

Fig. 4. Gradient method data flow.

To verify whether the method is correct, the following experiment was conducted. In the verification procedure, in place of reference radar data, precipitation estimation with known parameters values was used. Reference data was calculated using equation (1) for the interval from 21.06.2015 to 23.09.2015, with the use of the following parameters:

- c = 1 mm·km/h
- CWP_0 = 18 g/m^2
- α = 1.6

Gradient algorithm was initiated with the following parameters:

- c = 1 mm·km/h
- CWP_0 = 17 g/m^2
- α = 1.6

In this experiment, the next step was calculated with the following formula:

$$0.5(10^{-5}I + 0.7\nabla_p^2 E_{\underline{R}} + 0.3|\nabla_p E_{\underline{R}}|I)^{-1}\nabla_p E_{\underline{R}}, \qquad (14)$$

After 300, the "stop" condition was generated, and final results were as follows:

- c = 0.997935 mm*km/h

- CWP_0 = 17.9878 g/m^2
- α = 1.6054
- Final mean square error: 1.38 · 10^{-11} mm^2/h^2

This experiment shows that the gradient method with the mentioned modification is convergent, as the obtained resulting error is very small (Fig. 5).

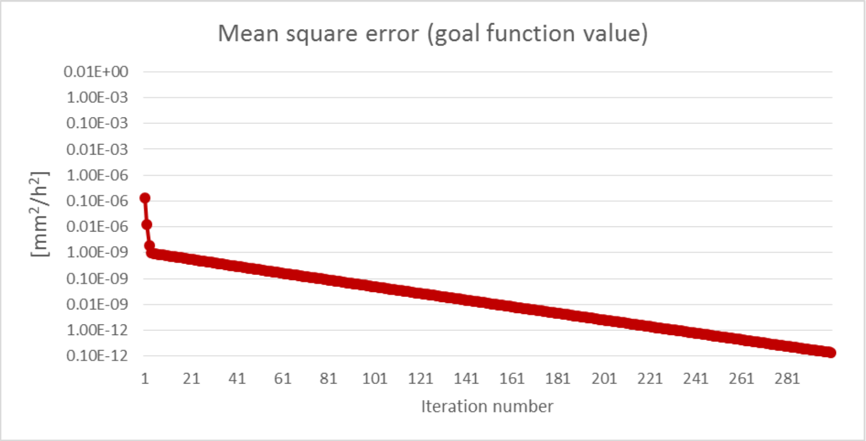

Fig. 5. Mean square error value in verification run (logarithmic scale).

4. Results

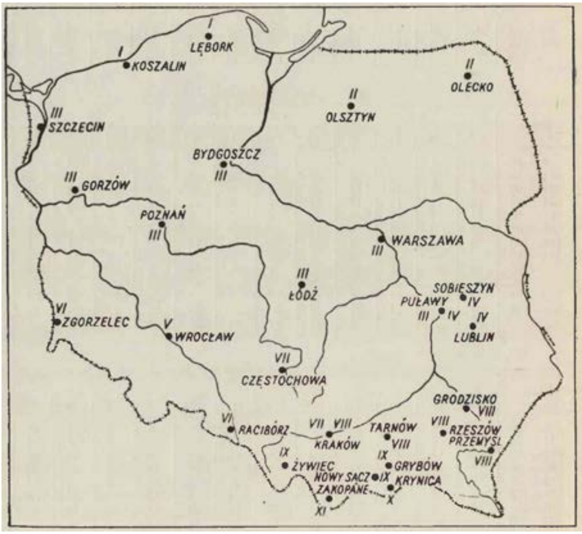

Fig. 6. Poland precipitation divisions [7].

In order to evaluate the usefulness of the proposed methodology, we performed three experiments with real datasets. For each case, calibration was performed for the same time interval - from 21.06.2015 to 23.09.2015, for the same starting parameters c = 1 mm km/h; CWP_0 = 18 g/m^2; α = 1.6, and with the same step calculation formula:

$$0.5(10^{-5}I + \nabla_p^2 E_{\underline{R}})^{-1}\nabla_p E_{\underline{R}} \quad (15)$$

The only difference between experiments was the spatial scope of the analysed data. The first real data experiment was performed on the area of entire Poland. Second experiment was constrained to a fragment of Poland, precipitation division no I. Third experiment was constrained to a fragment of Poland, precipitation division no III. Precipitation divisions are shown on Fig. 6.
For entire Poland, after "stop" condition was met, results were as follows:

- c = 0.0225682 mm km/h
- CWP_0 = 16.3789 g/m^2
- α = 0.557807
- Final mean square error: 2.55528 * 10^{-3} mm^2/h^2

The minimum of goal function was obtained after 102 iterations, and parameter values were:

- c = 0.103338 mm km/h
- CWP_0 = 24.0147 g/m^2
- α = 2.01764
- Mean square error: 2.49207 * 10^{-3} mm^2/h^2

Course of goal function was depicted on Fig. 7.

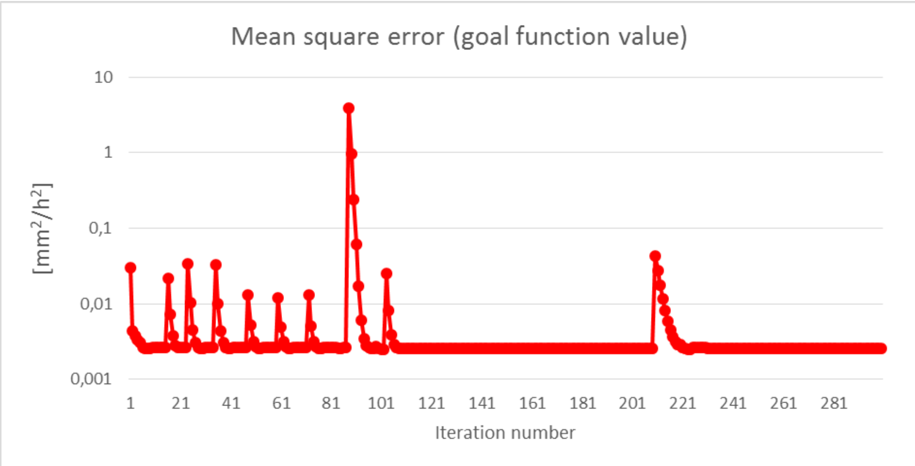

Fig. 7. Mean square error value in real data run for entire Poland (logarithmic scale).

For division I results were as follows:

- c = 0.662674 mm km/h
- CWP_0 = 64.5735 g/m^2
- α = 2.9977
- Final mean square error: 3.502 * 10^{-3} mm^2/h^2

Minimal error occurred in 140th iteration:

- c = 0.0522732 mm km/h
- CWP_0 = 14.4615 g/m^2
- α = 1.48839
- Mean square error: 3.102 * 10^{-3} mm^2/h^2

Plot of mean square error vs iteration no. was presented on Fig. 8.

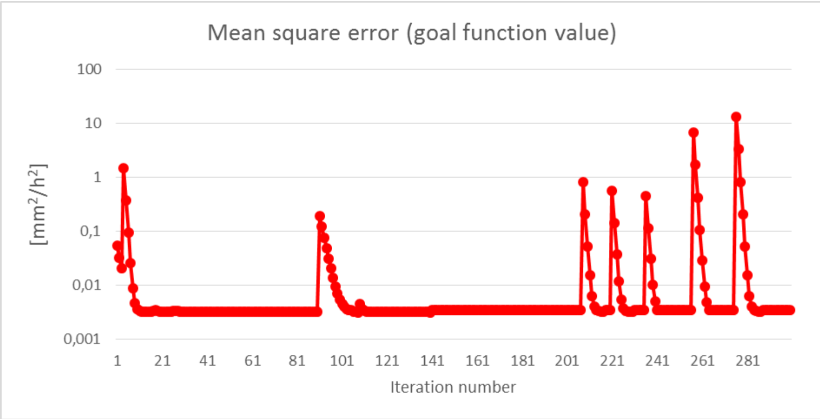

Fig. 8. Mean square error value in real data run for division I (logarithmic scale).

For division III after 300 iterations results were as follows:

- c = 1.21074 mm km/h
- CWP_0 = 100.546 g/m^2
- α = 3.47495
- Final mean square error: 4.3153 * 10^{-3} mm^2/h^2

Minimal error occurred in 169th iteration:

- c = 0.0712125 mm km/h
- CWP_0 = 10.8029 g/m^2
- α = 1.29595
- Mean square error: 3.6828 * 10^{-3} mm^2/h^2

Course of goal function was depicted on Fig. 9.

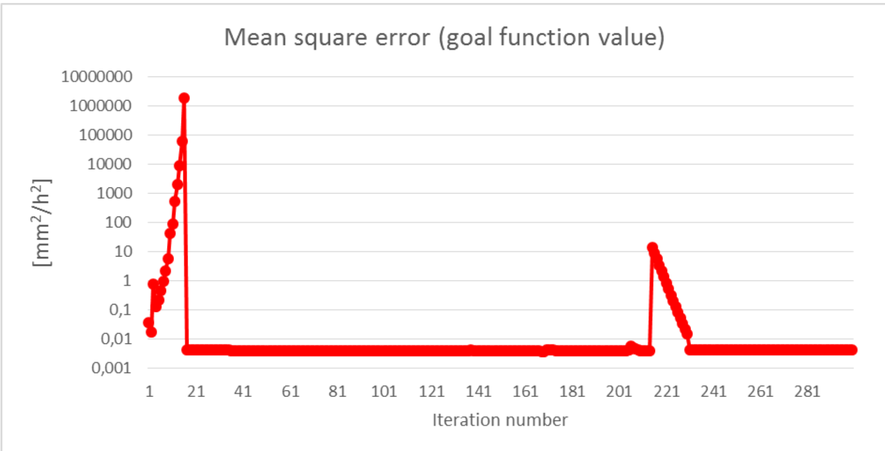

Fig. 9. Mean square error value in real data run for division III (logarithmic scale).

Analysis of results gave the following observations:

- in the case of real data, the calibration process tends to be unstable,
- CWP_0 and α seems to impact in an additive way to each other, which is not always positive,
- sometimes very different parameter sets give similar error values,
- parameters for minimal error have close values in the cases of division I and III.

5. Conclusions

In the paper, the method and technological framework for remote precipitation estimation was presented. The method utilizes a gradient method algorithm augmented by application of a hessian matrix, and regularisation parameter. Authors show that the proposed method is convergent, and finds an optimized set of parameters for particular sets of equations based on the reference big data validation set. However, the calibration process requires further improvements. First is the stability of the iterative method. This can be achieved by assigning a maximal step length in the domain of every parameter.

References

[1] H. Krawczyk, M. Nykiel, J. Proficz. "Tryton supercomputer capabilities for analysis of massive data streams", Polish Maritime Research 22.3 (2015), 99-104.

[2] R. A. Roebeling, I. Holleman; SEVIRI rainfall retrieval and validation using weather radar observations; JORNUAL OF GEOPHYSICAL RESEARCH Volume: 114/2009.

[3] R. A. Roebeling, H. M. Deneke, A. J. Feijt; Validation of Cloud Liquid Water Path Retrievals from SEVIRI Using One Year of CloudNET Observations; JORNUAL OF APPLIED METEOROLOGY AND CLIMATOLOGY Volume: 47/2009.

[4] D. M. Aminou, H. J. Luhmann , C. Hanson, P. Pili, B. Jacquet, S. Bianchi, F. Faure (2003, September). Meteosat Second Generation: A comparison of on-ground and on-flight imaging and radiometric performances of SEVIRI on MSG-1. In Proceedings of the 2003 EUMETSAT Meteorological Satellite Conference, Weimar, Germany (Vol. 29).

[5] S. Moszkowicz, I. Tuszyńska, Meteorologia radarowa: podręcznik użytkownika informacji radarowej IMGW. Instytut Meteorologii i Gospodarki Wodnej, 2003 (in Polish).

[6] D. W. Marquardt, An algorithm for least-squares estimation of nonlinear parameters. Journal of the society for Industrial and Applied Mathematics, 11(2), 431-441, 1963.

[7] K. Zofia, Opady w Polsce w przekroju wieloletnim, pages 55 – 57, Wydawnictwa Geologiczne, 1962 (in Polish).