Geosciences and Engineering, Vol. 14, No. 1 (2026), pp. 23–33.
https://doi.org/10.33030/geosciences.2026.01.002

ROBUST MFV ANALYSIS OF GEOELECTRICAL DATA IN HETEROGENEOUS SUBSURFACE MEDIA

MARCELL SZILVÁSI
Institute of Geography and Geoinformatics, University of Miskolc
marcell.szilvasi@uni-miskolc.hu
https://orcid.org/0009-0000-4900-2654

Abstract: Geoelectrical datasets in heterogeneous environments often show non-Gaussian distributions and outliers, limiting conventional averaging. This study evaluates the Most Frequent Value (MFV) method for robust characterization of vertically structured geophysical data. Results show that the fixed *ε* parameter, the scale parameter of the weighting function, provides stable representations, while the conventional iterative *ε* method enhances locally dominant structures. Quantitative comparison confirms that method performance depends on data variability and sampling density. The method offers a flexible alternative for interpreting complex subsurface conditions.

Keywords: *most frequent value, robust statistics, electrical resistivity, induced polarization, geophysical data processing, environmental geophysics*

1. INTRODUCTION

Geoelectrical measurement datasets applied in geological and environmental investigations often result in highly heterogeneous, non-Gaussian data structures affected by outliers. The natural inhomogeneity of the subsurface medium, spatial variability of contaminant distribution, and measurement noise may produce distributions in which the application of the classical arithmetic mean may lead to physically misleading results (Hampel et al., 1986; Huber, 1981). The application of robust statistical methods is therefore of particular importance in geophysical data processing. Their aim is to determine a location parameter that represents the dominant behavior of a dataset while limiting the excessive influence of extreme values (Rousseeuw and Leroy, 1987). In geosciences, robust approaches are especially important because outliers are not necessarily measurement errors but are often indicators of real geological phenomena.

Ferenc Steiner developed the Most Frequent Value (MFV) method and formalized it for geoscientific applications (Steiner, 1988; Steiner, 1990; Steiner, 1992). The MFV is an iterative weighting procedure designed to determine the center of the densest region of a dataset. Conceptually, it applies distance-dependent weights to the sample elements, allowing the algorithm to converge toward the most characteristic value of the distribution. A key component of the procedure is the parameter *ε*, which acts as a bandwidth (scale) parameter controlling the sensitivity range of the weighting function; in Steiner's original terminology, this is referred to as the "dihesion" parameter. The mathematical formulation of the weighting scheme and the role of ε in the iterative process are detailed in the Materials and Methods section.

The advantage of the MFV method compared to the arithmetic mean is that it is robust against outliers and non-normal distributions, while, unlike the median, it takes amplitude information and density structure into account. However, in the case of small sample size, uniform distribution, or excessive reduction of ε, the method may become unstable or locally biased. The MFV method has been applied in geophysical practice by several authors for various data processing and clustering tasks. Szabó and co-authors (Szabó, 2012; Szabó and Balogh, 2018) applied MFV-based factor analysis to interpret multi-parameter geophysical log data. Robust procedures operating on principles similar to MFV have also appeared in multivariate reservoir identification research (Szabó et al. 2023). During research within the multi-year Pulse project, the MFV method was also applied to compare parameters of abandoned hydrocarbon wells (Szűcs et al. 2022). These studies demonstrated that MFV-based aggregation improves the separability of geological units and the stability of structural trends compared to traditional averaging procedures.

Recultivated landfill environments represent particularly complex near-surface systems in this respect. Artificial layering, heterogeneous backfill materials, waste-related inclusions, and variable moisture conditions produce strongly irregular physical parameter distributions. Multi-electrode electrical resistivity tomography (ERT) and induced polarization (IP) surveys carried out above such sites provide high-resolution spatial information; however, the resulting inversion-derived datasets are vertically stratified and spatially heterogeneous. In such environments, it is particularly important to examine to what extent robust MFV estimation can represent dominant physical behavior and how spatial patterns change depending on the treatment of ε. In the present study, the investigated dataset originates from multi-electrode geoelectrical measurements conducted above the recultivated landfill site of Nyékládháza, northeastern Hungary. The site provides an appropriate test environment for evaluating robust statistical aggregation techniques under realistic field conditions characterized by strong spatial variability and anthropogenic heterogeneity. By applying the MFV method to vertically arranged resistivity and IP data, representative physical parameter values can be determined for each spatial cell while limiting the influence of extreme or depth-restricted anomalies.

The novelty of this study lies not in the MFV methodology itself, but in its application within a spatially controlled measurement system, enabling quasi-3D interpretation of a complex landfill environment. The objective of this study is to evaluate the applicability of the Steiner-type MFV method to vertically structured, multi-parameter geoelectrical datasets and to compare the behavior of fixed ε and iteratively recalculated ε approaches.

2. MATERIALS AND METHODS

2.1. MFV Calculation Procedure

The Most Frequent Value (MFV) method is a robust location estimator based on an iterative weighted averaging scheme. Its purpose is to determine a representative sample characteristic that reflects the densest region of the data distribution rather than being influenced by extreme values. Let a symmetric weighting function *φ(x)* be defined over the deviations of the data points from the current estimate of the location parameter. The MFV estimate is obtained as a weighted average, where the weights assigned to each observation depend on their distance from the current estimate *M*

$$M = \frac{\sum_{i=1}^{n} x_i \varphi_i}{\sum_{i=1}^{n} \varphi_i} \quad \text{where} \quad \varphi_i = \frac{\varepsilon^2}{\varepsilon^2 + (x_i - M)^2} \qquad (1)$$

The fundamental principle of the procedure is that observations located far from the main concentration of the dataset receive small weights, whereas observations situated within high-density regions are assigned to larger weights. Therefore, the resulting estimate converges toward the center of the densest subset of the sample rather than toward the arithmetic mean of all observations. The location parameter *M* corresponds to the maximum of the symmetric weighting function. Because the weighting scheme is distance-dependent, the estimation procedure is inherently nonlinear and requires iteration

$$M_{j+1} = \frac{\sum_{i=1}^{n} \frac{\varepsilon_{j+1}^2}{\varepsilon_{j+1}^2 + (x_i - M_j)^2} x_i}{\sum_{i=1}^{n} \frac{\varepsilon_{j+1}^2}{\varepsilon_{j+1}^2 + (x_i - M_j)^2}} \qquad (2)$$

A key parameter controlling the behavior of the weighting function is *ε*, which defines the effective width of the weighting function. The weights can be formally expressed as

$$\varepsilon_{j+1}^2 = \frac{3 \sum_{i=1}^{n} \frac{(x_i - M_j)^2}{[\varepsilon_j^2 + (x_i - M_j)^2]^2}}{\sum_{i=1}^{n} \frac{1}{[\varepsilon_j^2 + (x_i - M_j)^2]^2}} \qquad (3)$$

Parameter *ε* determines the effective width of the weighting function and thus defines the sensitivity range of the estimator. If *ε* is too large, the weighting function assigns nearly equal weights to all observations. In this case, the procedure approaches the arithmetic mean, and extreme values may significantly distort the estimate. If *ε* is too small, the weighting function becomes overly concentrated around a narrow subset of observations. In such cases, even data points close to the central cluster may receive insufficient weight, potentially leading to unstable or locally biased estimates. Therefore, the choice of ε strongly influences the shape of the weighting function and the final estimated value. The ε can be interpreted as an inverse measure of data compactness: in highly concentrated datasets, smaller ε values may be sufficient, whereas more dispersed datasets require larger sensitivity ranges. In this work two approaches were applied with the Steiner-type MFV procedure. The standard iteratively recalculated ε, which is updated at each iteration step together with the location parameter M, and the fixed ε, which remains constant throughout the iteration process. These two strategies lead to different convergence characteristics and spatial sensitivity behaviors, which are examined in the Results section.

2.2. Numerical Implementation

The calculations were implemented in a spreadsheet-based computational framework. Although the formulas appear lengthy due to repeated weighting steps, they are based on straightforward mathematical relationships. Each iteration step was calculated in a new column referencing the previous one, allowing a flexible number of iterations. The initial value of the location parameter M_0 was chosen as the sample mean. Tests using the sample mode as the initial value showed no significant difference in the final converged solution. The initial ε was defined as a $\frac{\sqrt{3}}{2}$ multiple of the sample range. In the fixed ε approach, this value remained unchanged during iteration. In the iterative approach, ε was recalculated at each step. Convergence was evaluated by monitoring the relative change between successive estimates

$$\frac{|M_i - M_{i-1}|}{M_{i-1}} \tag{4}$$

In highly heterogeneous datasets potentially containing multiple density concentrations, 20 iteration steps were performed to ensure stability. In the fixed ε case, convergence was typically achieved within 4–5 iterations (Figure 1), whereas the iterative ε approach required more steps (~10) due to progressive narrowing of the weighting range.

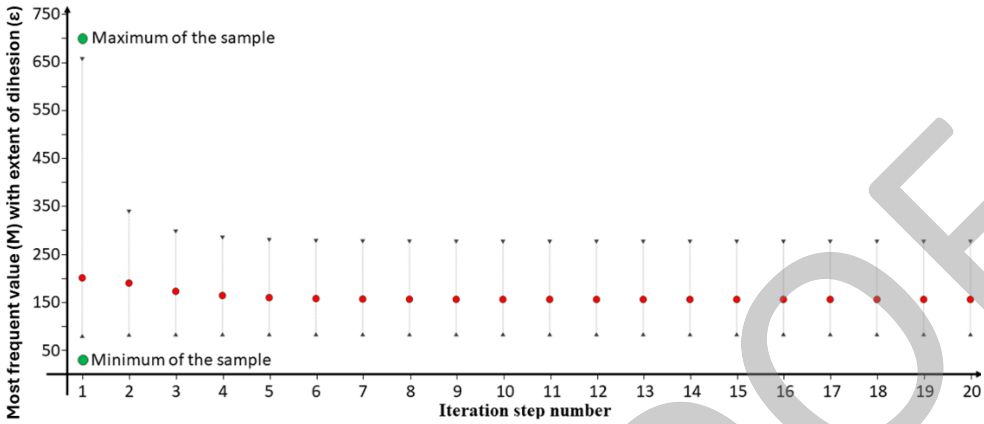

Figure 1

The most frequent value (red circles) and the change in the value of ε during the iteration steps

The dataset consists of 9,862 geoelectrical measurement points from the Nyékládháza recultivated landfill area. The objective was to generate plan-view maps displaying the most characteristic physical parameter of the given zone. The horizontal plane was divided into equal-sized square cells (20×20 m, 10×10 m, 5×5 m). All points within the full column of each cell were assigned to the cell center. The points within a cell were considered independent samples on which the MFV method was performed for both resistivity and IP values. The average sample sizes per column were 106 (20×20 m), 49 (10×10 m) and 23 (5×5 m) in each case. Samples with fewer than six elements may lead to unstable estimates; therefore, resolution reduction may be necessary, although this leads to information loss.

3. Results

3.1. Fixed ε Based MFV Method

In the case of fixed ε, the MFV formulation is effectively equivalent to a purely Cauchy-type weighting scheme, where the ε remains constant and no adaptive scaling is applied during the estimation process. The use of fixed ε requires less computational capacity and is less sensitive to highly variable datasets; therefore, it provides more stable results but may suppress characteristic patterns. The result is less noisy when creating 2D maps from independent samples; structural trends remain continuous, and local anomalies do not significantly distort surrounding areas. The method behaves as a quasi-robust estimator of the mean that well represents dominant resistivity behavior within a zone without giving excessive emphasis to extreme values (Figure 2). The maps gradually become more detailed as cell size decreases (Figure 2/A-C). In the case of coarser sampling, the overall effect of the landfill located in the northern part of the map becomes more dominant, while sensitivity to sudden changes decreases.

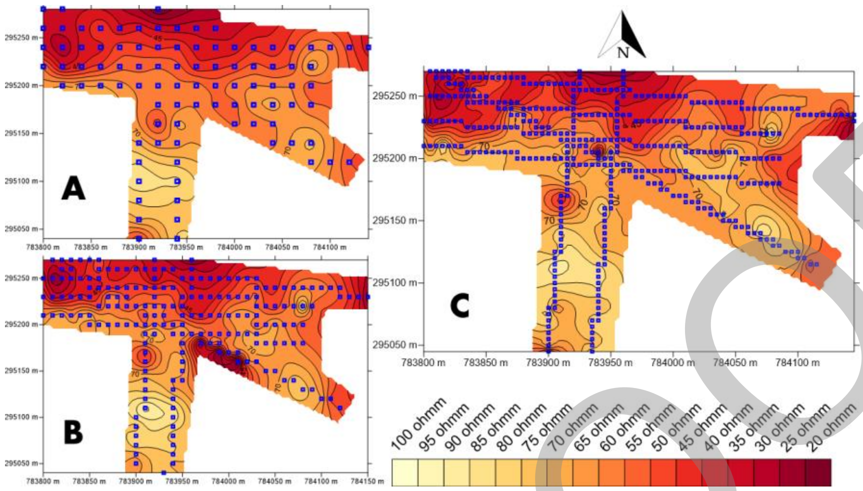

Figure 2

Apparent resistivity maps using the fixed ε MFV method, where blue squares indicate reference points. A: 20×20 m zones; B: 10×10 m zones; C: 5×5 m zones

At 5 m × 5 m resolution, the maps become highly detailed; however, in highly variable datasets, the MFV method may highlight values that do not fit the trend, appearing as small spikes (Figure 2/C). Therefore, the smoothing effect of coarser cell sampling may be beneficial despite reduced geometric resolution. The common color scale was set based on the smallest and largest occurring values; the redder the color, more it represents physical properties characteristic of waste, similarly to the IP maps.

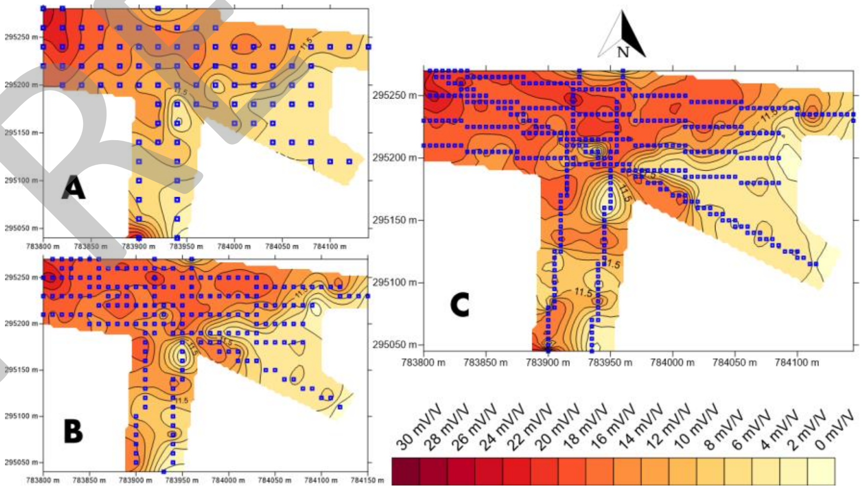

Figure 3

Apparent polarizability maps using the fixed ε MFV method, where blue squares indicate reference points. A: 20×20 m zones; B: 10×10 m zones; C: 5×5 m zones

In the case of IP data, similar behavior can be observed; however, it is interesting that in Figure 3/A the method highlights higher values only in the northwestern part of the map, while in Figures 3/B–C the cells with parameters characteristic of waste appear more scattered but overall dominant.

3.2. Iteratively recalculated ε MFV method

In the next investigation, ε was recalculated in each iteration step in addition to M, resulting in characteristic differences compared to the fixed ε method. Compared to the previous robust averaging, the algorithm now converges more strongly toward the most characteristic value in the sample and is less damped by the other values. As iteration progresses, the method assigns weights over an increasingly narrow range, resulting in median-like behavior that may excessively exclude elements outside the subset. This can be influenced by introducing a lower ε limit, which was not examined in this study. Due to stronger contrasts, the color scale was not evenly distributed, especially in resistivity maps to better observe different characteristic effects. The method resembles cluster analysis; however, here dominant clusters within zones appear rather than the relative proportions of different clusters.

In Figure 4, compared to the 100 Ωm maximum shown in Figure 2, a 400 Ωm maximum appears at the same points, and minimum values approach 0 Ωm instead of 20 Ωm. After manual data review, these values provide a more realistic picture of the dominant value of the given cell column compared to the fixed ε method. The resolution differences in Figures 4/A/B/C do not show as significant variation each other; at 10 m × 10 m resolution, dominant spatial effects can be observed without local outlier-calculated values.

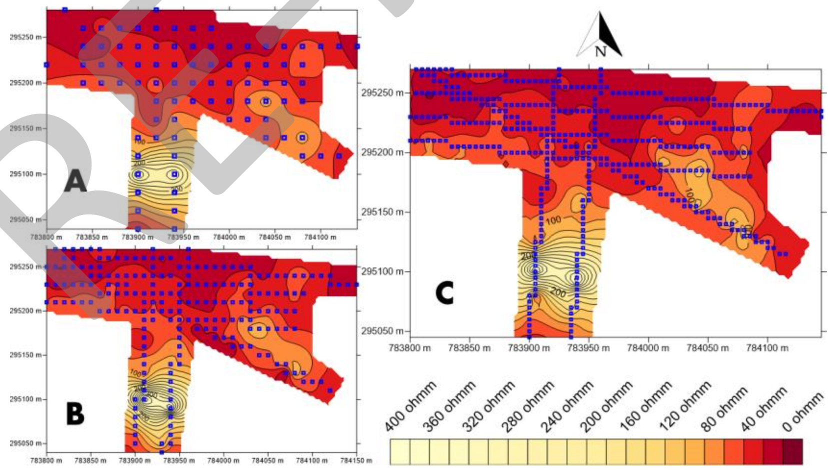

Figure 4

Apparent resistivity maps using the iteratively recalculated ε MFV method, where blue squares indicate reference points. A: 20×20 m zones; B: 10×10 m zones; C: 5×5 m zones

It is important to note that these outlier-like values often contain valuable information and should not always be smoothed or filtered out.

In Figure 5, IP data also show more pronounced characteristic zone values compared to Figure 3. In Figures 5/B-C, an interesting phenomenon can be observed: dominant data from four east–west oriented profiles do not converge with the two north–south oriented profiles crossing them. A single cell column may include multiple depth zones with different physical properties, forming non-Gaussian data distributions with multiple density maxima; thus, the MFV method may converge toward one or another depending on which is slightly more dominant. In such cases, the fixed ε MFV method may be more reliable. In Figure 5/A, sufficiently large cell sizes mix the data enough to allow high chargeability values to become dominant everywhere; here the method performs well.

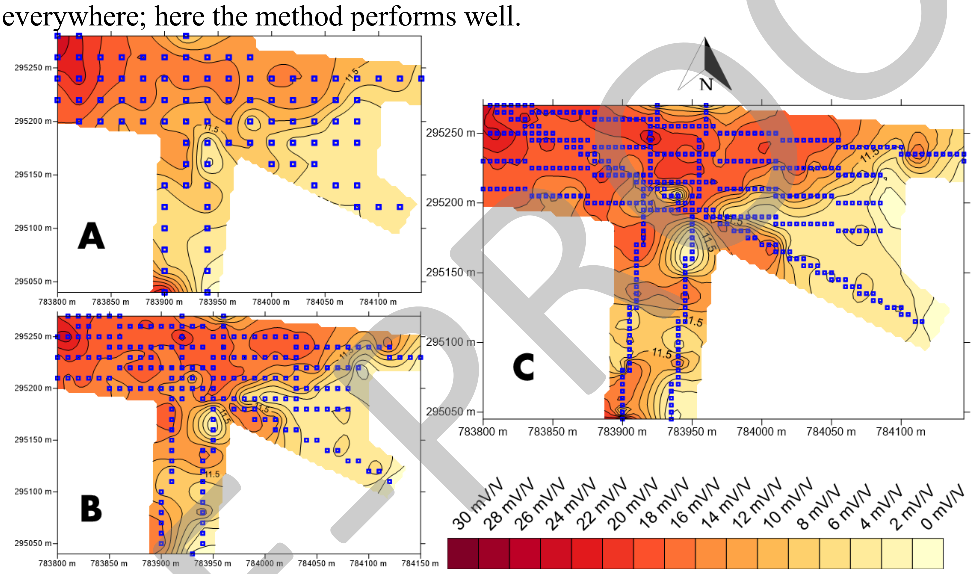

Figure 5

Apparent polarizability maps using the iteratively recalculated ε MFV method, where blue squares indicate reference points. A: 20×20 m zones; B: 10×10 m zones; C: 5×5 m zones

3.3. Quantitative comparison of MFV approaches

To provide an objective validation of the MFV approaches, a quantitative analysis was performed using statistical metrics. A quantitative comparison of the fixed ε and iteratively recalculated ε MFV methods was performed (Table 1) using relative deviation and its standard deviation as evaluation metrics across different spatial resolutions (5×5 m, 10×10 m, 20×20 m). Relative deviation expresses the difference between the MFV estimate and the arithmetic mean, interpreted here not as an error but as the degree of deviation from mean-based aggregation, while the standard deviation reflects the stability of the method.

Table 1

Quantitative comparison of MFV methods based on relative deviation and standard deviation

		5 x 5 m		10 x 10 m		20 x 20 m	
		fixed ε	iter. ε	fixed ε	iter. ε	fixed ε	iter. ε
Resistivity	Mean relative deviation [%]	20	21	18	21	22	25
	Standard deviation of relative deviation [%]	22	16	24	16	25	16
IP	Mean relative deviation [%]	43	19	39	19	21	37
	Standard deviation of relative deviation [%]	46	17	38	17	17	30

For resistivity data, the iterative ε approach consistently yielded slightly higher relative deviations (21–25%) and lower standard deviations (~16%) compared to the fixed ε solution (18–22% deviation, 22–25% standard deviation). This indicates that in highly heterogeneous datasets the iterative method more effectively emphasizes dominant values while maintaining stable convergence. In contrast, IP data exhibited resolution-dependent behavior. At higher sampling densities (5×5 m and 10×10 m), the fixed ε approach resulted in more balanced and stable outputs, whereas at lower sampling density (20×20 m) the iterative ε method performed more favorably. This suggests that for datasets with lower relative variability, stable MFV estimation requires increased spatial aggregation.

To characterize dataset heterogeneity, the coefficient of variation (CV = standard deviation / mean) was calculated. Resistivity data showed significantly higher variability (CV ≈ 1.34) compared to IP (CV ≈ 0.73), explaining the stronger and more consistent influence of the MFV method in resistivity datasets.

The optimal application of the MFV method depends not only on the choice of algorithm (fixed vs. iterative ε) but also on dataset variability and sampling density, which must be considered jointly for reliable interpretation. These findings highlight that MFV-based aggregation is not a universally optimal procedure, but a data-dependent method whose behavior is controlled by the interplay between spatial resolution and statistical variability.

4. Discussion

MFV-based aggregation provides a robust alternative to conventional averaging methods in heterogeneous geophysical datasets. The presence of non-Gaussian distributions and outliers significantly affects arithmetic mean-based representations, whereas MFV preserves the dominant structural characteristics of the data. The comparison of fixed and iterative ε formulations reveals that the two approaches emphasize different aspects of the data. The fixed ε method behaves as a stabilized estimator, producing spatially coherent and less noisy representations, which is particularly advantageous for mapping large-scale structural trends. In contrast, the iterative ε method progressively focuses on local data density maxima, enhancing dominant values but increasing sensitivity to local heterogeneities and multimodal distributions.

The observed differences are strongly controlled by dataset variability and sampling density. In highly heterogeneous datasets, such as resistivity, the iterative approach more consistently emphasizes dominant structures, while in less variable datasets, such as IP, stable results require either fixed ε or increased spatial aggregation. This highlights that MFV is not a universally optimal method, but a data-dependent approach. An important implication is that the choice of MFV parameterization must consider both statistical properties (e.g., variability) and spatial sampling characteristics. The results also show that smaller sample sizes increase the influence of outliers and may reduce the reliability of MFV estimation, indicating a practical lower limit for cell-based aggregation. Overall, the study demonstrates that MFV-based processing can significantly improve the interpretability of complex geophysical datasets, if method selection is adapted to the underlying data structure.

5. Conclusions

This study evaluated the applicability of the Most Frequent Value (MFV) method for the robust aggregation of vertically structured, multi-parameter geophysical datasets in a reclaimed landfill environment. The results demonstrate that MFV effectively reduces the influence of outliers while preserving dominant physical characteristics of the subsurface. The comparison of fixed and iterative ε formulations shows that the fixed ε approach provides stable and reproducible estimation, whereas the iterative ε method enhances local dominant structures but is more sensitive to dataset heterogeneity. Quantitative evaluation confirms that the performance of MFV methods is strongly dependent on dataset variability and sampling density. Consequently, the selection of MFV parameters must be adapted to the statistical and spatial characteristics of the data. The findings highlight that MFV is a flexible and robust tool for geophysical data interpretation, particularly in heterogeneous environments, and can provide significant advantages over conventional averaging methods when appropriately applied. The quantitative comparison indicates that MFV performance is not method-dependent alone but controlled by the interaction between data variability and sampling density.

References

Hampel, F. R., Ronchetti, E. M., Rousseeuw, P. J., Stahel, W. A. (1986). *Robust Statistics: The Approach Based on Influence Functions*. Wiley. ISBN: 9780471735779 DOI: 10.1002/9781118186435

Huber, P. J. (1981). *Robust Statistics*. Wiley. ISBN: 9780471418054 DOI: 10.1002/0471725250

Rousseeuw, P. J., Leroy, A. M. (1987). *Robust Regression and Outlier Detection*. Wiley. ISBN: 9780471852339 DOI: 10.1002/0471725382

Steiner, F. (1988). Most frequent value procedures (a short monograph). *Geophysical Transactions*, 34, pp. 139–260.

Steiner, F. (1990). *A geostatisztika alapjai*. Budapest, Tankönyvkiadó. ISBN: 963-18-2819-0

Steiner, F. (1992). The Most Frequent Value: Introduction to a Modern Conception of Statistics. Budapest, Akadémiai Kiadó. ISBN: 963-05-5687-1

Szabó, N. P. (2012). Bevezetés a geostatisztikába. Miskolci Egyetem.

Szabó, N. P., Balogh, G. P., Stickel, J. (2018). Most frequent value-based factor analysis of direct-push logging data. Geophysical Prospecting, 66, pp. 530–548. DOI: 10.1111/13652478.12573

Szabó, N. P., Kilik, R., Dobróka, M. (2023). Robust reservoir identification by multi-well cluster analysis of wireline logging data. Heliyon, 9, pp. 1-17. DOI: 10.1016/j.heliyon.2023.e15957

Szűcs, P., Turai, E., Mádai, V., Vass, P., Miklós, R., Zákányi, B., Ilyés, Cs., Fekete, Zs., Kilik, R., Móricz, F., Nyíri, G., Szilvási, M., Szabó, N. P. (2022): Innovation in assessment of the geothermal energy potential of abandoned hydrocarbon wells in the southern and southeastern foreground of the Bükk Mountains, northeast Hungary. Hydrogeology Journal, 30, pp. 2267–2284. DOI: 10.1007/s10040-022-02560-y