DOI: 10.14750/ME.2020.008

UNIVERSITY OF MISKOLC

Faculty of Earth Science and Engineering

Department of Geophysics

INVERSION-BASED FOURIER TRANSFORMATION ALGORITHM USED IN PROCESSING NON-EQUIDISTANTLY MEASURED MAGNETIC DATA.

PhD THESIS

by

DANIEL O.B NUAMAH

Scientific supervisors:

Prof. Dr. Mihály Dobróka

Assoc. Prof. Dr. Péter Vass

MIKOVINY SÁMUEL DOCTORAL SCHOOL OF EARTH SCIENCES
Head of the Doctoral School: Prof. Dr. habil. Péter Szűcs

Miskolc, 2020

HUNGARY

DOI: 10.14750/ME.2020.008

CONTENT

DOI: 10.14750/ME.2020.008

Chapter 1

INTRODUCTION

Data processing is an essential discipline in the science and engineering fields of study. The ability to acquire quality information from interpretation largely depends on the efficacy of the data processing method applied. In geophysics applications where interpretations are made from data collected at the earth's surface to forecast subsurface features, the quality of the processing method is of great importance. In a broader perspective, this thesis focuses on the development of new methods in inversion-based Fourier transformation for geophysical applications in the area of regular and random data processing. The continual improvement in geophysical data acquisition over the years require more advanced data processing methods. Data translation from a time domain to the frequency domain is commonly practiced in geophysical data processing, which enhances interpretation, especially in signal processing. This change can be realized through the application of Fourier transformation. For individually sampled time-domain datasets, the Discrete Fourier Transformation (DFT) algorithm is usually applied to determine its Discrete Frequency Component (DFC). As measured data often contain noise, the noise sensitivity of the processing methods is an essential feature. The noise recorded in the time domain is directly transformed into the frequency domain. Hence, the traditional discrete variants of Fourier transformation, although very stable, are noise sensitive techniques that require improvement.

To reduce this problem, an inversion based 1D Fourier transformation algorithm with the capability of reducing geophysical data outliers was presented by Dobróka et al. (2015) known as the Steiner Iteratively Reweighted Least Square Fourier Transform (S-IRLS-FT) which proved to be an effective tool for noise reduction. The method was generalized to 2D, and an application was presented in solving a reduction to the pole of a magnetic data set (Dobróka et al., 2017). Geophysical data processing covering inverse problem theory has a collection of methods with outstanding noise rejection capabilities. This necessitated the proposition to handle 1D Fourier Transform as an overdetermined inverse problem (Dobroka et al. 2012). As established in the inverse problem theory, the simple least-squares give the best solution only when data noise follows Gaussian distribution. For outliers that are irregularly distributed large errors, the estimated model parameters may be highly inconclusive, which constitute a restrictive factor to the application of the least-squares method since geophysical measurements routinely contain outliers. To achieve statistical robustness, various methods have been developed over the years to deal with data outliers. A commonly applied

robust optimization method, the Least Absolute Deviation (LAD), minimizes the L_1-norm characterizing the misfit between the observed and predicted data, and can be numerically achieved by using linear programming (Scales et al. 1988) or applying the Iteratively Reweighted Least Squares method (IRLS). Although largely used, continual practice demonstrates that inversion with minimization of the L_1-norm gives more reliable estimates only when a smaller number of large errors contaminate the data. An alternative solution involves the use of the Cauchy criterion, which adopts a Cauchy-distributed data noise. The IRLS procedure, which iteratively recalculates the so-called Cauchy weights, results in a very efficient robust inversion method (Amundsen et al. 1991). The application of data weights in inversion is very crucial to guarantee each data contribute to the solution based on its error margin. Cauchy inversion is normally applied in geophysical inversion as a robust optimization method (Steiner F. 1997). The integration of the IRLS algorithm with Cauchy weights, though a useful procedure, problematic since the scale parameter of the weights has to be known prior to the inversion. Steiner (1988,1997) adequately solved this challenge by deriving the scale parameters from the real statistics of the data set in the framework of the Most Frequent Value (MFV) method. Dobróka et al. 1991 established globally that the MFV-weights calculated on the bases of Steiner's method result in a very efficient robust inversion method by inserting them into an IRLS procedure. A successful application of the MFV method in processing borehole geophysical data was reported by Szűcs et al. 2006. The Cauchy weights improved by Steiner (the so-called Cauchy-Steiner weights) were further applied in robust tomography algorithms by Szegedi H. and Dobróka M., 2014.

Relying on the above techniques, Dobróka et al. (2015) developed the inversion based 1D Fourier transformation method known as the S-IRLS-FT, which proved to be an effective tool for noise reduction. It was revealed that the noise sensitivity of the continuous Fourier transform (and its discrete variants DFT and FFT) was sufficiently reduced by using robust inversion. The 1D Fourier transform was handled as a robust inverse problem using the IRLS algorithm with Cauchy-Steiner weights. The Fourier spectrum was further discretized using series expansion as a discretization tool. Series expansion based inversion methods were successfully used in the processing of borehole geophysical data (Szabó 2004, Dobróka et al. 2010) as well as Induced polarization data (Turai et al. 2010). The S-IRLS-FT method was generalized to 2D, and an application presented in solving reduction to the pole of the magnetic data set (Dobróka et al., 2017). In this study, it is shown that the newly developed inversion-

DOI: 10.14750/ME.2020.008

based Fourier transformation algorithm can also be used in processing a non-equidistant (random) measurement geometry dataset.

Chapter 2

THE INVERSION-BASED FOURIER TRANSFORM

2.1 An Overview of Geophysical Inversion Methods

An important part of geophysical studies is to make inferences about the interior of the earth from data collected at or near the surface of the earth. The measured data is indirectly related to the properties of the earth that are of interest. An inverse problem can be solved to obtain estimates of the physical properties within the earth. The aim of a geophysical inverse problem is to find an earth model described by a set of physical parameters that is consistent with the observational data (Barhen et al. 2000). The process first involves the calculation of simulated data for an earth model from a forward problem. Thus, accurate synthetic data is generated for an arbitrary model. The inverse problem is then posed as an optimization problem where the function to be optimized generally called the objective, misfit, or fitness function, which is a measure of the difference between observed data and synthetic data extrapolated. Due to data inaccuracies occurring from field measurement procedures and data processing techniques, the objective function often incorporates some additional form of regularization or constraints. Data Inversion problems are not restricted to geophysics but can be found in a wide variety of disciplines where inferences must be made based on indirect measurements.

Inversion applications in geophysics require two special considerations when methods of solution are being generated. First, the observed data are usually incomplete in the sense that they do not contain enough information to resolve all features of the model. Solving a geophysical inverse problem normally involves finding an optimum solution and appraising the validity of that solution. The appraisal includes an analysis of resolution, which is a determination of what features of the solution are essential to explain the data. Invariably, the optimum solution is non-unique in the sense that some of its features could be changed without changing the fit to the data. Secondly, the observed data always has a noise component from two primary sources, a random component in the observed data and approximations or errors contained in the theory that connects the data and model. The presence of noise requires an analysis of uncertainty in the appraisal stage of the inverse problem, which is a determination of how much the optimum solution would change if a different realization of the noise were to

be used (Barhen et al. 2000). A seemingly possible problem to geophysical inverse solution is its non-uniqueness. Thus, there are many possible solutions to the problem. Hence, it requires a comprehensive explanation of the possible solutions in other to constrain the solution. Inversion methods development attempt this task by performing a general search of the model space including grid, random and pseudo-random searches. There are also methods that estimate a relative probability density for the model space. The common method of addressing the fundamental non-uniqueness of geophysical inverse problems is to impose additional constraints on the solution reducing the number of acceptable solutions (Parker, 1994; Oldenburg et al., 1998), and this process is known as regularization. Regularization is generally a measure of some property of the model that is deemed to be desirable. The constraints imposed on the model space try to retain certain properties that are thought to be necessary and are quite subjective, relying on information that is independent of the data. After parameterization of the data and model spaces, next is a determination of the constraint types to be placed upon the model space to specify a model or group of models that are compatible with a set of observations drawn from the data space. Several types of constraints are possible. A theoretical constraint involves mapping from the model space to the data space and allows a direct relationship to be established (Barhen et al. 2000).

Numerous inversion techniques have been developed by various researchers for optimum objective function determination. Mostly used are the linear optimization methods due to their very quick and effective procedures in cases of the suitable initial model but are not absolute minimum searching methods and generally assign the solution to a local optimum of the objective function. This problem can be avoided by using global optimization methods, for example, Simulated Annealing (Metropolis et al., 1953) or Genetic Algorithm (Holland J.H, 1975). Global optimization methods have high performance, great adaptability, and previously used in other fields such as well-logging interpretation (Zhou et al. 1992, Szucs and Civan 1996, Goswami et al. 2004 and Szabó 2004).

2.1.1 Linearized Inversion Procedures

For geophysical inverse problems where the relationship between the data and the model is linear, methods of solution are well developed and understood (Menke, 1989; Parker, 1994). Linear inversion methods are based on the solution of a set of linear equations, which are relatively fast procedures. These prevailing methods are used for several geophysical problems. The common starting point of these methods is the linearization of a function connection. In

formulating the discrete inverse problem, the column vector of *'M'* number of model parameters is introduced as

$$\vec{m} = \{m_{1,}m_{2,\ldots..,}m_M\}^T \tag{1}$$

where *'T'* denotes the matrix transpose. Similarly, the *'N'* number of data measured by geophysical surveys are collected into the data vector

$$\vec{d}_m = \left\{d_1^{(m)}, d_2^{(m)}, \ldots\ldots, d_N^{(m)}\right\}^T \tag{2}$$

Let the calculated theoretical data be sorted into the following *N*-dimensional vector

$$\vec{d}_c = \left\{d_1^{(c)}, d_2^{(c)}, \ldots\ldots, d_N^{(c)}\right\}^T \tag{3}$$

a connection between vectors $\vec{d}_c$ and $\vec{m}$ is given as

$$\vec{d}_c = \vec{g}(\vec{m}) \tag{4}$$

Now, considering $\vec{m}_0$ as a starting point in the model space where

$$\vec{m} = \vec{m}_0 + \delta\vec{m} \tag{5}$$

the model correction vector is given by $\delta\vec{m}$. Let the connection be approximated by its Taylor series truncated at the first-order additive term,

$$d_{\boldsymbol{e_k}} = \boldsymbol{g_k}(\vec{m}_0) + \sum_{j=1}^{M}\left(\frac{\partial_{g_k}}{\partial_{m_j}}\right)_{\vec{m}_o} \delta m_{j,} \quad \text{(k=1,2,….N)} \tag{6}$$

By introducing the Jacobi's matrix

$$G_{kj} = \left(\frac{\partial_{g_k}}{\partial_{m_j}}\right)_{\vec{m}_o} \text{ and } d_k^{(0)} = \boldsymbol{g_k}(\vec{\boldsymbol{m}}_{\boldsymbol{0}})$$

Equation (6) can be written as

$$d_{\boldsymbol{e_k}} = d_k^{(0)} + \sum_{j=1}^{M} G_{kj}\delta m_j \tag{7}$$

or in vector form

$$\vec{d}_e = \vec{d}^{(0)} + \underline{\underline{G}}\delta\vec{m} \tag{8}$$

By applying $\delta\vec{d} = \vec{d}_e - \vec{d}^{(0)}$, it can be seen that $\delta\vec{d} = \underline{\underline{G}}\delta\vec{m}$ is the linearized form of equation (8). Different optimization principles are available for model parameterizations that are either continuous or discrete. Solutions based on maximum likelihood are normally used where data noise is present, and its distribution is known. Measurement of resolution in both the data space and model space can also be calculated (Berryman, 2000), allowing quantitative estimates of the fitting of the data and the uniqueness of the model. The Gaussian least-square, which minimizes the *L2-norm* of the deviation vector, has proven to be a faster and an operative linear method. The objective function to be minimized is the squared *L2*-norm of the deviation vector characterizing the misfit between the calculated and observed data, given as

$$E = \vec{e}^T\vec{e} = \sum_{k=1}^{N} e_k^2 = \sum_{k=1}^{N}\left(d_k - \sum_{J=1}^{M} G_{kj}m_j\right)\left(d_k - \sum_{i=1}^{M} G_{ki}m_i\right) \quad (9)$$

which has optimum if the following set of equations are fulfilled for $l = 1, 2, \ldots, M$

where $\frac{\partial E}{\partial m_l} = 0$, and minimized to the normal equation

$$\sum_{i=1}^{M} m_i \sum_{k=1}^{N} G_{ki}\, G_{kl} = \sum_{k=1}^{N} G_{kl}\, d_k \quad (10)$$

with a vectorial form given as

$$\underline{\underline{G}}^T\underline{\underline{G}}\,\vec{m} = \underline{\underline{G}}^T\vec{d} \quad (11)$$

Here, the model parameters are obtained from

$$\vec{m} = (\underline{\underline{G}}^T\underline{\underline{G}})^{-1}\underline{\underline{G}}^T\vec{d} \quad (12)$$

A similarly applicable linearized procedure is the Weighted Least Squares Method. The Weighted Least Squares Method can be effectively used for solving overdetermined inverse problems (Menke, 1984) and efficiently suppresses data outliers. It is often encountered that the uncertainties of observed data are of different amounts, which requires that a datum contributes to the solution with a given weight proportional to its uncertainty. This is done by the application of a symmetric weighting matrix, which includes the weighs of the data in its main diagonal. The solution is developed by the minimization of the following objective function

$$E = \vec{e}^T\underline{\underline{W}}\vec{e} = \sum_{k=1}^{N}(d_k - \sum_{i=1}^{M} G_{ki}m_i)\sum_{r=1}^{N} W_{kr}\left(d_r - \sum_{j=1}^{M} G_{rj}m_j\right) \quad (13)$$

where $\frac{\partial E}{\partial m_l} = 0$ and minimize to the normal equation,

$$\frac{\partial E}{\partial m_l} = 2\sum_{i=1}^{M} m_i \sum_{k=1}^{N}\sum_{r=1}^{N} W_{kr}\, G_{ki}\, G_{rl} - 2\sum_{k=1}^{N} d_k \sum_{k=1}^{N} W_{kr}\, G_{rl} = 0 \qquad (14)$$

With a vectorial form,

$$\underline{\underline{G}}^T \underline{\underline{W}}\, \underline{\underline{G}}\, \vec{m} = \underline{\underline{G}}^T \underline{\underline{W}} \vec{d} \qquad (15)$$

Here, the model parameters are estimated from

$$\vec{m} = (\underline{\underline{G}}^T \underline{\underline{W}}\, \underline{\underline{G}})^{-1} \underline{\underline{G}}^T \underline{\underline{W}} \vec{d} \qquad (16)$$

The actual model is gradually refined until the best fitting between measured and calculated data is achieved in the inversion procedure. Although very useful, Linearized Inversion Procedures has a general possibility to map noise in the data directly into a measure of uncertainty in the model space and hence, requires more optimum handling of regularization. As discussed earlier, most geophysical inverse problems are not well-posed as originally formulated and usually involve the imposition of some form of regularization to alleviate the situation. Rarely but decisively, the degree of regularization is optimized as part of obtaining the solution by the introduction of an independent variable parameter. With a wider application that exists for solving linear inverse problems, it is possible to formulate problems so that linear methods can be used whenever possible. Problems that are not too strongly non-linear can be solved by the process of linearization. As much as the solution does not stray too far from a reference model, the problem can be solved with standard linear methods, including the standard linear estimates of resolution and uncertainty. In circumstances where the reference model is unknown, it is handled by an iterated linearization procedure in which a new reference model is produced, the entire linearization and solution process is then repeated. This type of linearized approach to the solution of an inverse problem is commonly used in the location of earthquakes, where it is known as Geiger's method (Lee and Stewart, 1981). Linearized methods do not guarantee to find the absolute minimum of the objective function as they tend to assign the solution to some local minimum. This problem requires to apply such methods that can seek the global minimum of the objective function. Global optimization methods such as Simulated Annealing and Genetic Algorithms can be used effectively to find a global optimum in geophysical applications.

2.1.2 Global Inversion Procedures

For geophysical inverse problems where both the objective function and the constraints are significantly nonlinear in the model space, Global Inversion Procedures are applied. This may

include a derivative or non-derivative based approach to solving the problem. Several developed solutions make use of derivatives, global and local convergence. For problems without constraints, Newton's method is applicable while those requiring both first and second derivatives, the Hessian matrix and Nonlinear Conjugate Gradient Methods (Paige and Saunders, 1982) have provided satisfactory results (Nolet, 1984, 1985, Newman and Alumbaugh, 1997). Some developed procedures solve a non-linear inverse problem by solving a series of linearized problems. Practically, this is not different from standard iterative methods developed for solving non-linear problems such as the line search and trust region methods (Dennis and Schnabel, 1996). It is advantageous to use these established non-linear methods as convergence proofs exist, and well-tested algorithms are accessible. For many geophysical problems, it is difficult to justify the linearization of a problem when efficient methods of solving the non-linear problem are available.

Methods without derivatives such as Genetic Algorithms, Simulated Annealing, and Pattern Search Algorithms have seen considerable development and preference over the years because they are less likely to converge to a nearby local optimum than derivative-based methods.

Multiple solutions are important for non-linear inverse problems, as most optimization methods only provide a local extremum, and separate procedures are used to find a more global extremum. Methods specifically designed to find global optima are the grid and stochastic search methods. Grid search methods are mostly applicable to smaller problems whilst larger problems require a Monte Carlo search, which has the advantage of being simple to implement and easy to check (Mosegaard and Tarantola, 1995; Mosegaard, 1998). However, for most geophysical inverse problems, the number of model parameters and the required accuracy are such that a complete Monte-Carlo search is unfeasible, simply because of the number of times the forward problem would have to be calculated to achieve sufficient sampling of the model space. While neither enumerative nor completely random searches of the model space have proven to be an effective method of solving larger geophysical inverse problems, there are some directed or pseudo-random search methods such as Simulated Annealing and Genetic Algorithms that have been more successful. Both approaches retain some aspects of a random statistical search of the model space but use the gradually accumulating information about acceptable models to direct the search and appear to be feasible for moderately sized problems where a full Monte Carlo approach would be prohibitive (Scales et al., 1992).

Simulated Annealing is based upon an analogy with a natural optimization process in thermodynamics and uses a directed stochastic search of the model space. It requires no derivative information. Its use in numerical optimization problems began with Kirkpatrick et

al. (1983) and was first use in geophysical problems by Rothman (1985, 1986). A review of the method and its application to geophysical problems can be found in Sen and Stoffa (1995). The Simulated Annealing (SA) approach covers a group of global optimization methods. The earliest form is called the Metropolis algorithm, which was further developed to improve mainly the speed of the optimum-seeking procedure, referred to as the fast and very fast Simulated Annealing methods. The Metropolis Simulated Annealing (MSA) algorithm employs the heating technology to search for the global optimum of an objective function and has been applied in several geophysical probes, for instance, in calculating seismic static corrections (Rothman, 1985; Rothman, 1986; Sen and Stoffa, 1997), global inversion of vertical electric sounding data collected above a parallel layered 1-D structure (Sen et al., 1993). The advantages of MSA are initial model independence, simple and clear-cut program coding, and exact mathematical treatment of the conditions of finding a global optimum. The method has a slow rate of convergence, which sets a limit to the reduction of control temperature.

Genetic algorithms are direct search methods based on the natural optimization processes found in the evolution of biological systems (Goldberg, 1989). It applies the operators of coding, selection, crossover, and mutation to a finite population of models and allows the principle of "survival of the fittest" to guide the population toward a composition that contains the optimum model (Barhen et, al. 2000). The first application for its use in solving optimization problems was proposed by John Holland (1975) and has been applied to several geophysical problems (Stoffa and Sen, 1991; Sen and Stoffa, 1992, 1995; Sambridge and Drijkoningen, 1992; Kennett and Sambridge, 1992; Everett and Schultz, 1993; Sambridge and Gallagher, 1993; Nolte and Frazer, 1994; Boschetti et aI., 1996; Parker, 1999). The Genetic Algorithm procedure improves a population of random models in an iteration process. In optimization problems, the model is considered as an individual of an artificial population. Each individual of a given generation has a fitness value, which represents its survival ability. The purpose of the Genetic algorithm procedure is to improve the subsequent populations by maximizing the average fitness of individuals. In application, the fitness function is connected to the distance between the observed data and theoretical data calculated. Normally, an initial population of models is generated from the search space randomly. In the forward modeling phase of the inversion procedure, theoretical data are calculated for each model and then compared to real measurements. The model population is improved through the use of some random genetic operations such as selection, crossover, mutation, and reproduction to reduce the misfit between the observation and prediction data. Instead of a one-point search, several models are analyzed

simultaneously to avoid the local optimum places in the model space. The Genetic Algorithm technique is also advantageous because it does not require the calculation of derivatives and require less prior information. It is practically independent of the initial model. Approaches that attempt some combination of stochastic and deterministic search methods would appear to hold considerable promise in optimization. This will enable a combination of the global search property of the stochastic methods with the efficiency of the deterministic methods.

2.2 The Series Expansion-Based Inversion Methods

Complex geological structures require the forward problem to be solved by approximate numerical methods such as finite difference (FDM) and finite element (FEM) methods. These methods enable the use of discretization for the adequate approximation of a spectrum. For instance, a space can be divided into properly sized blocks or an accurate number of cells. In this case, adequate calculations require some distinct number of cells in both horizontal and vertical directions. In the inverse problem solution, the physical parameters of the cells are assumed to be unknowns. At the Department of Geophysics, the University of Miskolc, the series expansion-based discretization scheme, which is based on the discretization of the model parameter is suggested and have proven to be useful in several cases. Considering a model parameter showing spatial dependency in the form of series expansion,

$$p(x,y,z) = \sum_{i=1}^{N_x} \sum_{j=1}^{N_y} \sum_{k=1}^{N_z} B_l \Psi_i\,(x) \Psi_j(y) \Psi_k(z) \tag{17}$$

Where $\Psi_i \ldots\ldots \Psi_N$ are the basis functions and N_x, N_y, N_z are the requisite numbers in the description of x,y,z dependencies. The basis functions constitute an orthonormal system of functions and are chosen carefully since they affect the stability of the entire inversion procedure. The unknowns of the inverse problem (model parameters) are the series expansion coefficients B_l, and their number is given as $M = N_x, N_y, N_z$. Based on the number of elements of the model vector and data available, the inverse problem may be underdetermined, overdetermined or mixed determined. If the number of data is more than that of the model parameters ($N>M$), the inverse problem is overdetermined. As explained earlier, an overdetermined problem can be solved adequately with the Gaussian Least Squares by minimizing the l_2-norm but can be weighted when the data has uncertainties. In the previous case, it is unnecessary to use additional constraints because the result is dependent only on the data. Conclusions from the study of the purely underdetermined inverse problem show that a unique solution could be obtained only by assuming additional conditions, which cannot be

joined to measurements. These conditions should formulate an obvious requirement on the solution such that they introduce some level of simplicity, smoothness, or have a significant effect on the magnitude of the derivatives. In some cases, the use of additional conditions can be advantageous, but in extremely underdetermined problems, the solution is problematic.

2.2.1 The Algorithm

In most appropriate situations in geophysical inversion, a priori information about the area of investigation is usually accessible for the interpretation. This knowledge is of great importance during the inversion procedure since the inversion algorithms have internal uncertainty (instability, ambiguity), which can be reduced by the use of a priori information. Series expansion-based inversion procedures have a similar situation as one can only assume or specify the number of expansion coefficients. With adequate information about the structure, one can make additional assumptions within the series expansion-based inversion method, which can facilitate a reduction in the number of unknowns of the inverse problem. In case of a 3-D Layer-wise homogeneous model, the q-th layer-boundary can be described as a function $\mathrm{z} = \mathrm{f_q(x, y)}$, which can be discretized by series expansion as,

$$z = f_q(x,y) = \sum_{i=1}^{N_x^{(q)}} \sum_{j=1}^{N_y^{(q)}} C_l^{(q)} \Psi_j(x) \Psi_j(y) \tag{18}$$

Where $C_l^{(q)}$ represents the expansion coefficients, $l = L_q + i + (j-1) * N_x$, L_q is the initial index required in the *q*-th layer. The number of unknowns for a given layer-boundary is $N_x^{(q)} N_y^{(q)}$, while that of the *P*-layered model assuming one physical parameter per layer is

$$\mathrm{M} = \sum_{q=1}^{p} N_x^{(q)} N_y^{(q)} + P + 1 \tag{19}$$

In practice, assuming a layer-wise homogeneous model is often not adequate. For a vertically inhomogeneous model, the physical parameter of the *q*-th layer can be written as

$$p_q(z) = \sum_{i=1}^{N_q^{(p)}} D_l^{(q)} \Psi(z) \tag{20}$$

where the number of unknowns including the layer-boundaries is also given by

$$\mathrm{M} = \sum_{q=1}^{p} (N_x^{(q)} N_y^{(q)} + N_q^{(p)}) + 1 \tag{21}$$

Also, $l = L_q + i$, L_q is the initial index in the *q*-th layer. Assuming lateral inhomogeneity in each layer, the series expansion-based discretization of the physical parameter is given as

$$p_q(x,y) = \sum_{i=1}^{N_{p,x}^{(q)}} \sum_{j=1}^{N_{p,y}^{(q)}} D_l^{(q)} \Psi_i(x) \Psi_j(y) \tag{22}$$

Where $D_l^{(q)}$ represent the expansion coefficients, $l = L_q + i + (j-1) * N_x$, L_q is the initial index required in the *q*-th layer. The number of unknowns for a given layer-boundary has been

broadened with $N_{p,x}^{(q)}N_{p,y}^{(q)}$ in comparison to the layer-wise homogeneous model, for the *P*-layered model,

$$\mathrm{M} = \sum_{q=1}^{p}(N_x^{(q)}\,N_y^{(q)} + N_{p,x}^{(q)}N_{p,y}^{(q)}) + 1 \qquad (23)$$

Model parameterization through series expansion increases the overdetermination ratio in geophysical inversion. Comparably, a four-layered structural boundary and its physical parameters approximated by fifth-degree polynomials can be defined by M=4*(5*5+5*5) +1=201 number of expansion coefficients while the number of unknowns of underdetermined problems is typically ~10^6. Thus, the choice of a discretization procedure has the potential to improve the results of inverse modeling. For vertically and laterally inhomogeneous model, a standard model which combines vertical and lateral inhomogeneity is considered. A discretization of the physical parameters can be given by

$$p_q(x,y,z) = \sum_{i=1}^{N_{p,x}^{(q)}}\sum_{i=1}^{N_{p,x}^{(q)}}\sum_{k=1}^{N_{p,z}^{(q)}} B_l^{(q)}\,\Psi_i(x)\Psi_j(y)\Psi_k(z) \qquad (24)$$

Where $l = L_q + i + (j-1) * N_{p,x}^{(q)} + (k-1) * N_{p,x}^{(q)} * N_{p,y}^{(q)}$, L_q is the initial index required in the q-th layer. The number of unknowns for a given layer-boundary has been broadened with $N_{p,x}^{(q)}N_{p,y}^{(q)}N_{p,z}^{(q)}$ in comparison to the layer-wise homogeneous model, thus, the *P*-layered model can be obtained from

$$\mathrm{M} = \sum_{q=1}^{p}(N_x^{(q)}\,N_y^{(q)} + N_{p,x}^{(q)}N_{p,y}^{(q)}N_{p,z}^{(q)}) + 1 \qquad (25)$$

The choice of discretization procedure is essential in inverse modeling. Inversion algorithms used for the discretization of 3- D structures are greatly overdetermined and does not include additional subjective conditions. The suggested algorithm does allow to integrate a priori information into the inversion method as well as keeping the computing procedures, and can be applied to the 3-D inversion of measurement data of any geophysical surveying method.

2.2.2 Some Applications In Near Surface Geophysics

Geophysical method development in robust inversion at the Department of Geophysics, University of Miskolc, largely depends on the processing and evaluation of data measured on complex (laterally and vertically inhomogeneous) geological structures. It involves using series expansion discretization where the expansion coefficients are defined in an inversion process. The main advantage of this method is that a suitable resolution can be realized by introducing a relatively small number of expansion coefficients so that the task leads to an overdetermined inverse problem. The concept of series expansion based inversion has been used in numerous fields of geophysics. A general solution of the method was illustrated by Turai and Dobróka

2001. An application of series expansion based inversion in borehole geophysics to solve a non-linear well-logging inverse problem by Simulated Annealing as a global optimization method was shown by Szabó 2004. Dobróka M. and Szabo N.P 2011 further processed borehole geophysical data using this method, where the depth-dependent physical parameters were written as series expansion and the series expansion coefficients defined within the framework of the inversion process. An original method was presented for the processing of induced polarization (IP) data using the series expansion inversion by Turai et al. (2010), known as the TAU transformation. A monotonously decreasing apparent polarizability curve observable in the time domain can be described by Fredholm type integral equation

$$\eta_a(t) = \int_0^\infty w(\tau) \exp(-t/\tau)\, d\tau \tag{26}$$

Applying series expansion, the time constant spectrum *w(τ)*, which is a continuous real-valued function, was estimated with accuracy from a finite number of measurement data through discretization. The time constant spectrum was written in the form of series expansion as

$$w(\tau) = \sum_{q=1}^{Q} B_q \Phi q(\tau) \tag{27}$$

where Φ_q is the q-th basis function and B_q is the q-th expansion coefficient. Since the basis functions are a priori given, the extraction of the time constant spectrum reduced to the determination of unknown expansion coefficients. Defining TAU transformation as an inverse problem, the vector of series expansion coefficients B_q became the unknown model vector, and the forward problem was solved by substituting the discretized spectrum (equation 27) into the response function (equation 26) to give a connection at measured time t_k as

$$\eta(t_k) = \eta_k^{calc} = \int_0^\infty \sum_{q=1}^{Q} B_q \Phi_q(\tau) \exp\left(-\frac{t_k}{\tau}\right) d\tau = \sum_{q=1}^{Q} B_q \int_0^\infty \Phi_q(\tau) \exp\left(-\frac{t_k}{\tau}\right) d\tau \tag{28}$$

By introducing the following notation

$$S_{kq} = \int_0^\infty \Phi_q(\tau) \exp\left(-\frac{t_k}{\tau}\right) d\tau \tag{29}$$

the calculated data were generated by the expression

$$\eta_k^{calc} = \sum_{q=1}^{Q} B_q\, S_{kq}\,, \tag{30}$$

which in matrix form is

$$\vec{\eta}^{calc} = \underline{\underline{S}}\vec{B} \tag{31}$$

The TAU transformation was successful in delineating municipal waste contaminates. Dobroka et al., 2013 proposed a similar approximate series expansion based inversion method for imaging Magneto Telluric (MT) data measured above 2-D geological structures. In discretizing the model parameters, a series expansion formula was used with interval-wise constant functions or Chebishev polynomials as base functions. The expansion coefficients served as the unknown parameters of the inverse problem, and the imaging algorithm was restricted to layer-wise homogeneous geological models with laterally changing boundaries. Writing the (*n*-th) thickness function in the form of a series expansion gave

$$h_n(x) = \sum_{q=1}^{Q} B_{nq}\,\phi_q \quad , \quad n = 1,..., N \;\; -1 \tag{32}$$

where $\phi_q(x)$ is the q-th base function and B_{nq} is the *q*-th expansion coefficient of the *n*-th layer, *x* denotes the lateral coordinate, N is the number of layers. Here, Q is a priori given number of the base functions taken into account in the truncated series expansion. This number depended on the variability of the model whilst the choice of the base functions depended on the nature of the geological model. The applied Chebishev polynomials used as the basis function for discretization was given as

$$\phi_q(x) = T_q(x) \tag{33}$$

Using Eötvös torsion and gravity measurements, deflections of the vertical and digital terrain model data by series expansion inversion based reconstruction of a three-dimensional gravity potential was presented by Dobróka and Völgyesi (2010). The Fourier transform was also handled as a series expansion based inverse problem by Dobróka and Vass (2006). In addition to the above, an efficient method for the series expansion based inversion of geoelectric data measured on two-dimensional geological structures was shown by Gyulai et al. 2010.

2.3 Fourier Transform as Series Expansion-Based Inversion

The application of series expansion based Inversion to Fourier data processing was proposed by Dobroka et al., 2012, by introducing the LSQ-FT method. This procedure involves series expansion based discretization using Hermite functions as basis functions. Taking advantage of the beneficial properties of Hermite-functions, that they are the eigenfunctions of the inverse Fourier transformation, the elements of the Jacobian matrix were calculated quickly and easily without integration. The series expansion coefficients are given by the solution of a linear inverse problem. In this Thesis, the Hermite functions based method will be abbreviated as H-LSQ-FT method. The entire process was also robustified using the IRLS method by the

application of Steiner weights, thereby enabling an internal iterative recalculation of the weights. This resulted in a very efficient, robust, and resistant inversion procedure with a higher noise reduction capability. The integration of the IRLS algorithm with Steiner weights is a very useful procedure since the scale parameter of the weights can be derived from the real statistics of the data set in the framework of the Most Frequent Value method (Steiner F. 1988, 1997). In the following this Hermite functions based robust method will be abbreviated as H-IRLS-FT method. The procedure was further improved for noise reduction by Dobroka et al., 2017, where it was successfully used to reduce magnetic data to the pole.

2.3.1 1D H-LSQ-FT method

Data conversion from the time domain to the frequency domain can be established using a Fourier transform. The connection enhances data interpretation since certain features are improved in one data format than the other. For the one-dimensional case, the Fourier transform is defined as

$$U(\omega)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}u(t)e^{-j\omega t}dt\,, \tag{34}$$

where t denotes the time, ω is the angular frequency and j is the imaginary unit. The frequency spectrum $U(\omega)$ is the Fourier transform of a real-valued time function $u(t)$, and it is generally a complex-valued continuous function. Thus, the Fourier transform provides the frequency domain representation of a phenomenon investigated by the measurement of some quantity in the time domain. The inverse Fourier transform ensures a return from the frequency domain to the time domain.

$$u(t)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}U(\omega)e^{j\omega t}d\omega \tag{35}$$

In defining the Fourier transform as an inverse problem, the frequency spectrum $U(\omega)$ should be described by a discrete parametric model. In order to satisfy this requirement, we assumed that $U(\omega)$ is approximated with sufficient accuracy by using a finite series expansion

$$U(\omega)=\sum_{i=1}^{M}B_i\Psi_i(\omega)\,, \tag{36}$$

Where the parameter B_i is a complex-valued expansion coefficient and Ψ_i is a member of an accordingly chosen set of real-valued basis functions. Using the terminology of (discrete)

inverse problem theory, the theoretical values of time-domain data (forward problem) can be given by the inverse Fourier transform

$$u^{theor}(t_k) = u_k^{theor} = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} U(\omega) e^{j\omega t_k} d\omega ,$$

where t_k is the k-th sampling time. Inserting the expression given in Eq. (1) one finds that

$$u_k^{theor} \cong \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \left(\sum_{i=1}^{M} B_i \Psi_i(\omega) \right) e^{j\omega t_k} d\omega = \sum_{i=1}^{M} B_i \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \Psi_i(\omega) e^{j\omega t_k} d\omega .$$

Introducing the notation

$$G_{k,i} = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \Psi_i(\omega) e^{j\omega t_k} d\omega , \tag{37}$$

where $G_{k,i}$ is an element of the Jacobian matrix of the size N-by-M. The Jacobian matrix is the inverse Fourier transform of the Ψ_i basis function. Parameterization of the model is achieved by introducing a special feature of the Hermite functions, thus, by making them the eigenfunctions of the forward Fourier transform as

$$\mathrm{F}\left\{H_n^{(0)}(t)\right\} = (-j)^n H_n^{(0)}(\omega) , \tag{38}$$

and respectively for the inverse Fourier transform

$$\mathrm{F}^{-1}\left\{H_n^{(0)}(\omega)\right\} = (j)^n H_n^{(0)}(t) , \tag{39}$$

The Hermite functions were modified by scaling because, in geophysical applications, the frequency covers wider ranges. The theoretical values can, therefore, be written in the linear form as

$$u_k^{theor} = \sum_{i=1}^{M} B_i G_{k,i} . \tag{40}$$

2.3.2 2D H-LSQ-FT method

The 2D Fourier transform of a function *u(x,y)* can be calculated by the integral

$$U(\omega_x, \omega_y) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} u(x, y) e^{-j(\omega_x x + \omega_y y)} dx\, dy , \tag{41}$$

its inverse is given by the formula

$$u(x, y) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} U(\omega_x, \omega_y) e^{j(\omega_x x + \omega_y y)} d\omega_x d\omega_y , \tag{42}$$

where *x*, *y* are the spatial coordinates, $U(\omega_x,\omega_y)$ is the 2D spatial-frequency spectrum and ω_x, ω_y indicate the spatial-angular frequencies. The discretization of the continuous spectrum can be done through series expansion,

$$U(\omega_x,\omega_y)=\sum_{n=1}^{N}\sum_{m=1}^{M}B_{n,m}\Psi_{n,m}(\omega_x,\omega_y), \tag{43}$$

where $\Psi_{n,m}(\omega_x,\omega_y)$ are frequency-dependent basis functions, $B_{n,m}$ are the expansion coefficients that represent the model parameters of the inverse problem. The basis function system should be square-integrable in the interval (-∞, ∞). The Hermite functions meet this criterion with an additional advantage. Dobróka et al. (2015) showed that the elements of the Jacobian matrix could be considered as the inverse Fourier transform of the basis function system. Therefore, they can be calculated more easily if the basis functions are chosen from the eigenfunctions of the inverse Fourier transformation. By introducing 'α' as a scale parameter, it can be shown, that the normed and scaled Hermite functions are given by

$$H_n(\omega_x,\alpha)=\frac{e^{-\frac{\alpha\omega_x^2}{2}}h_n(\omega_x,\alpha)}{\sqrt{\sqrt{\frac{\pi}{\alpha}}n!(2\alpha)^n}}, \tag{44}$$

$$H_m(\omega_y,\beta)=\frac{e^{-\frac{\beta\omega_y^2}{2}}h_m(\omega_y,\beta)}{\sqrt{\sqrt{\frac{\pi}{\beta}}m!(2\beta)^m}}, \tag{45}$$

and are the eigenfunctions of the inverse Fourier transformation. The Jacobian matrix of the inverse problem can be written as

$$G_{k,l}^{n,m}=\frac{(j)^{n+m}}{\sqrt[4]{\alpha\beta}}H_n^{(0)}\left(\frac{x_k}{\sqrt{\alpha}}\right)H_m^{(0)}\left(\frac{y_l}{\sqrt{\beta}}\right). \tag{46}$$

Here $H_n^{(0)},H_m^{(0)}$ denote the non-scaled Hermite functions and provides a fast solution to the forward problem.

$$u(x_k,y_l)=\sum_{n=1}^{N}\sum_{m=1}^{M}B_{n,m}G_{k,l}^{n,m}. \tag{47}$$

2.3.3 The robust Inversion algorithm used in H-IRLS-FT

The Gaussian Least Squares Method (LSQ), which minimizes the L_2-norm of the deviation vector between the observed and calculated data is normally applied when the data noise

follows the regular distribution. Unfortunately, most geophysical data contains irregular noise with randomly occurring outliers making the least-squares method (LSQ) less effective for processing. Dobroka et al. 2012 emphasized the possibilities of obtaining a good result in an inverse problem solution when the data is weighted. To develop a robust algorithm, the weighted norm of the deviation vector was minimized using Cauchy weights, which were further modified to Cauchy-Steiner weights. The minimized weighted norm is given as

$$E_w = \sum_{k=1}^{N} w_k e_k^2 \tag{48}$$

Where $'w_k'$ is the Cauchy weights, given by

$$w_k = \frac{\sigma^2}{\sigma^2 + e_k^2}$$

Applying Steiner's Most Frequent Value method (MFV), the scale parameter σ^2 was determined from the data set in an internal iteration loop. By experience, a stop criterion was defined from a fixed number of iterations. After this, the Cauchy weights were calculated using the Steiner's scale parameter. The so-called Cauchy-Steiner weights at the last step of the internal iterations are given by

$$w_k = \frac{\varepsilon^2}{\varepsilon^2 + e_k^2}, \tag{49}$$

where ε_{j+1}^2 the Steiner's scale factor called dihesion is determined iteratively.

In practice, the misfit function is non-quadratic in the case of Cauchy-Steiner weights (because e_k contains the unknown expansion coefficients) and so the inverse problem is nonlinear which can be solved again by applying the method of the Iteratively Reweighted Least Squares (Scales, 1988). In the framework of this algorithm, a *0*-th order solution $\vec{B}^{(0)}$ is derived by using the non-weighted LSQ method and the weights are calculated as

$$w_k^{(0)} = \frac{\varepsilon^2}{\varepsilon^2 + (e_k^{(0)})^2}$$

with $e_k^{(0)} = u_k^{measured} - u_k^{(0)}$, where $u_k^{(0)} = \sum_{i=1}^{M} B_i^{(0)} G_{ki}$ and the expansion coefficients are given by the LSQ method. In the first iteration, the misfit function

$$E_w^{(1)} = \sum_{k=1}^{N} w_k^{(0)} e_k^{(1)2}$$

is minimized resulting in the linear set of normal equations

$$\boldsymbol{G}^T\boldsymbol{W}^{(0)}\boldsymbol{G}\vec{B}^{(1)} = \boldsymbol{G}^T\boldsymbol{W}^{(0)}\vec{u}^{measured}$$

The minimization of the new misfit function

$$E_w^{(2)} = \sum_{k=1}^{N} w_k^{(1)} e_k^{(2)^2}$$

gives $\vec{B}^{(2)}$ which serves again for the calculation of $w_k^{(2)}$. This procedure is repeated giving the typical *j*-th iteration step

$$\boldsymbol{G}^T\boldsymbol{W}^{(j-1)}\boldsymbol{G}\,\vec{B}^{(j)} = \boldsymbol{G}^T\boldsymbol{W}^{(j-1)}\vec{u}^{measured} \tag{50}$$

with the $\boldsymbol{W}^{(j-1)}$ weighting matrix

$$W_{kk}^{(j-1)} = w_k^{(j-1)} \tag{51}$$

Each step of these iterations contains an internal loop for the determination of the Steiner's scale parameter which is repeated until a proper stop criterion is met.

2.4 Some Features and Problems of Inversion-Based Fourier Transform

The basic concept of the H-IRLS-FT method can be summarized into four distinct steps which are the formulation of Fourier transformation as an over-determined inverse problem:

- discretization by series expansion using Hermite functions as basis functions,
- calculation of the Jacobi matrix using Hermite functions as the Eigenfunctions of the Fourier transform and
- robustification of the entire process by IRLS Method using Steiner weights.

The use of Hermite function as a basis function of discretization is important for the method development because they are orthonormal and square-integrable between the interval -∞ to ∞. In geophysical applications, the frequencies cover wider ranges. Hence, the Hermite functions had to be modified by scaling. This necessitated the introduction of a scale parameters 'α' and 'β' into equation (44) and (45) above. Unfortunately, the value of the scale parameter is inserted into the algorithm from practical experience, which is problematic, making it difficult to assume. There is a real need to exclude this problem either

a.) by defining a new method (with different discretization) or
b.) improve the H-LSQ-FT or H-IRLS-FT procedure giving the optimal values of the scale parameters.

DOI: 10.14750/ME.2020.008

For instance, other useful functions with previous successful applications in series expansion based discretization such as power functions or Legendre polynomials may be considered for further development. Legendre polynomials have been used in interval inversion of well log data (Dobroka et al., 2016, Szabó et al, 2018) to give accurate estimates to the series expansion coefficients. It is well known that the choice of a better basis function affects the stability of the inversion procedure; hence, other alternatives can be tested for the inversion based FT method.

An iteratively derived scale parameter has the potential to improve the efficiency of the algorithm and the entire output of the H-IRLS-FT method.

In spite of the successes achieved by the H-IRLS-FT algorithm in equidistant geophysical data processing, specifically noise reduction and outlier suppression,

c.) the theory and algorithm can further be improved for processing non-equidistant (randomly measured) data.

Recent developments in random walk field data acquisition in geophysics have increased the need for robust processing methods like the H-IRLS-FT. The improvement in geophysical data acquisition tools coupled with higher digitization and reduction in tool sizes enable easy navigation in the field of survey. Also, the development of advanced survey equipment which incorporates a global positioning system (GPS) facilitates random-walk data acquisition in recent times. Traditional survey designs employ equidistant measurement on a regular grid. Unfortunately, measurements are sometimes taken out of the grid due to several obstacles encountered in the field of survey. Inaccessible sample locations are caused by natural (such as caves) or man-made (buildings) reasons which distorts already planned regular survey designs. This has necessitated the development of methods for the effective processing of datasets taken in a non-equidistant grid (random geometry). The above a.), b.) and c.) subsections denote the main directions of the research work presented in this Thesis.

DOI: 10.14750/ME.2020.008

Chapter 3

NEW LEGENDRE POLYNOMIAL-BASED FT METHODS: L-LSQ-FT, L-IRLS-FT

3.1 Legendre polynomials as basis functions

Legendre polynomials are a system of complete orthogonal polynomials with numerous applications in science and engineering fields of study. Of interest to this study is its physical and numerical application in geophysics. They are defined as orthogonal thus, if $P_n(x)$ is a polynomial of degree 'n', then

$$\int_{-1}^{1} P_m(x)P_n(x)dx = 0 \qquad \text{if } n \neq m \tag{52}$$

Another distinguished property of Legendre polynomial is its definite parity, in that, they are symmetric or asymmetric given that

$$P_n(-x) = (-1)^n P_n(x) \tag{53}$$

These properties make it convenient when Legendre polynomials are used in series expansion to approximate a function in the interval (-1,1). Also, the Legendre differential equation and the orthogonality property are independent of scaling. The Legendre differential equation is given as

$$(1 - x^2)\frac{d^2y}{dx^2} - \frac{2xdy}{dx} + n(n+1)\, y=0 \tag{54}$$

where n>0, |x|<1, or equivalently

$$\frac{d}{dx}\left[(1 - x^2)\frac{dy}{dx}\right] + n(n + 1)y = 0 \tag{55}$$

The solution of the above equations gives the Legendre functions of order 'n' with a general solution expressed as

$$Y = AP_n(x) + BQ_n(x) \tag{56}$$

Where $P_n(x)$ and $Q_n(x)$ are Legendre functions of the first and second kind of order 'n'. For n=1, 2, 3,……, $P_n(x)$ functions are referred to as Legendre polynomials and are given by the Rodrigue's formula

$$P_n(x) = \frac{1}{2^n\, n!}\frac{d^n}{dx^n}(x^2 - 1)^n \tag{57}$$

DOI: 10.14750/ME.2020.008

Below is a table showing Legendre functions of the first kind $P_n(x)$ for n=0, 1, 2, 3…., using Eq. (57).

Table 1, Generated Legendre Polynomials of Order n=0 to 5.

n	Legendre polynomial
0	$P_0(x) = 1$
1	$P_1(x) = x$
2	$P_2(x) = \frac{1}{2}(3x^2 - 1)$
3	$P_3(x) = \frac{1}{2}(5x^3 - 3x)$
4	$P_4(x) = \frac{1}{8}(35x^4 - 30x^2+3)$
5	$P_5(x) = \frac{1}{8}(63x^5 - 70x^3 - 15x)$

Higher order Legendre polynomials can be obtained by the recursive formula below

$$P'_{n+1}(x) - P'_{n-1}(x) = (2n + 1)P_n(x). \quad (58)$$

For n=1,2,3…., where $P_n(1) = 1$ and $P_n(-1) = (-1)^n$. The graphical plot of these polynomials up to n=5 is shown in Figure 1 below

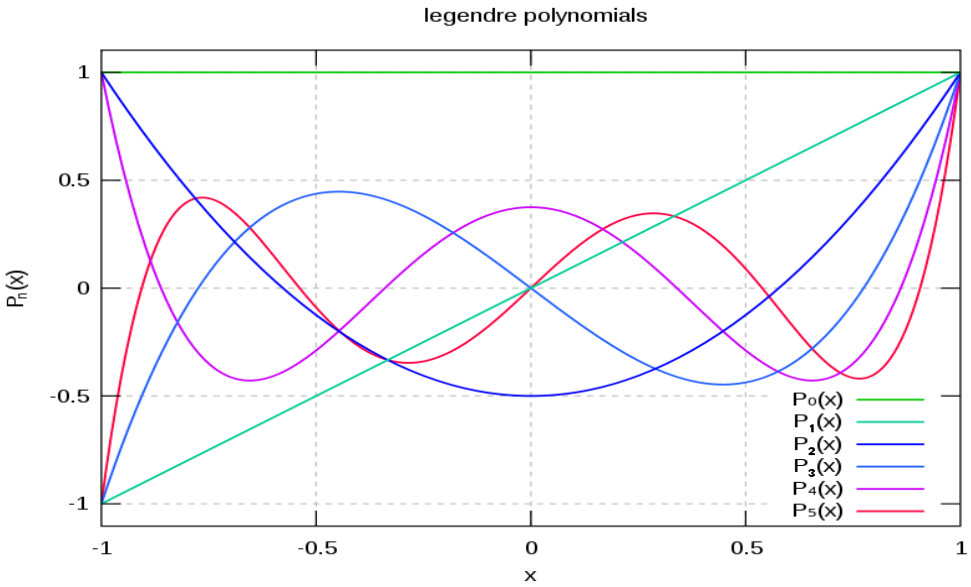

Figure 1; Graphical Plot of n=1,…,5 Legendre polynomials

Legendre polynomials have been used in interval inversion of well log data (Dobroka et al., 2016, Szabó N.P and Dobróka M., 2019) to give accurate estimates to the series expansion coefficients. It is well acknowledged that the choice of a better basis function affects the stability of the inversion procedure hence, in the following steps, Legendre polynomials will be tested for the inversion based FT method.

3.2 The L-LSQ-FT and L-IRLS-FT algorithm in 1D

As measured geophysical data always contain noise, the noise sensitivity of the processing methods is an important feature. In this chapter a new 1D robust inversion based Fourier transformation algorithm is introduced: the Legendre-Polynomials Least-Squares Fourier Transformation (L-LSQ-FT) and the Legendre-Polynomials Iteratively Reweighted Least-Squares Fourier Transformation (L-IRLS-FT). Noise in Geophysical data has varied sources, which may be regular or non-regular in nature. The interference of regular noise in geophysical data has long been a nuisance problem for geophysicists. These noises commonly originate from power-line radiations, global lightning, transmitters, oscillating sources and inadequate data processing (Butler and Russell, 1993; Jeng et al., 2007; Bagaini, 2010). Various methods have been proposed to suppress both systematic and non-systematic noise in geophysical records which include Subtracting an estimate of the noise from the recorded data (Nyman and Gaiser, 1983; Butler and Russell, 1993; Jeffryes, 2002; Meunier and Bianchi, 2002; Butler and Russell, 2003; Saucier et al., 2006). These methods are derived under the assumption that each sinusoidal contaminant is stationary, thus, constant in amplitude, phase, and frequency over the length of the record (Butler and Russell, 2003). Unfortunately, this assumption is impractical because the attributes of systematic noise always drift with time for many reasons. Other effective methods are by using inversion techniques or implementing filters with the pattern-based scheme (Guitton and Symes, 2003; Guitton, 2005; Haines et al., 2007). Filters employing pattern models are effective but they are time-consuming, and adequate pattern models are necessary for filter estimation (Haines et al., 2007).

The inversion technique-based methods require a sufficient number of regularization and are more applicable if data quality is good. In the field of inverse problem theory, a variety of numerous procedures are available for noise rejection, hence formulating the Fourier transformation as an inverse problem enables the use of sufficient tools to reduce noise sensitivity. Following the theory of Dobróka et al, 2012, the discretization of the continuous Fourier spectra is given in this thesis by a series expansion with Legendre polynomials as a square-integrable set of basis functions. By using Legendre polynomials as basis function of

discretization, the Fourier spectrum was adequately approximated and the expansion coefficients are determined by solving an overdetermined inverse problem. As deduced earlier, equation (37) above shows the general form of the Jacobi matrix in the case of a one-dimensional series expansion based inverse Fourier transform. Using the general Jacobian matrix

$$G_{k,n} = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \Psi_n(\omega) e^{j\omega t_k} d\omega ,$$

where $G_{k,n}$ is an element of the Jacobian matrix of the size N-by-M. The Jacobian matrix is the inverse Fourier transform of the Ψ_n basis function. Parameterization of the model is achieved by introducing the Legendre polynomials (equation 57) as basis function to give,

$$G_{k,n} = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} P_n(\omega) e^{j\omega t_k} d\omega \tag{59}$$

or in a more formal notation

$$G_{k,n} = \mathrm{F}_k^{-1} \left\{ P_n(\omega) \right\}. \tag{60}$$

The basic idea of introducing a new inversion-based Fourier Transformation method is to calculate the inverse FT of eq. (59) by using a common inverse DFT procedure:

$$G_{k,n} = IDFT_k \left\{ P_n(\omega) \right\}. \tag{61}$$

For the sake of simplicity, the sampling should be regular in time and frequency. Note, that the values of the $P_n(\omega)$ functions are accurate (noise-free), so the application of IDFT (or IFFT) is independent of the noise problem (of the data set), mentioned above. By using this procedure the $G_{k,n}$ elements of the Jacobi matrix can numerically be generated. At this point, the inversion method is to be defined. The theoretical value of the signal at a time point t_k is

$$u^{theor}(t_k) = u_k^{theor} = \sum_{i=1}^{M} B_n G_{k,n}$$

and the k-th element of the data deviation vector is written as

$$e_k = u_k^{(meas)} - u_k^{(theor)} = u_k^{(meas} - \sum_{n=1}^{M} B_n \, G_{kn} .$$

DOI: 10.14750/ME.2020.008

Using L_2-norm, the misfit function is given as

$$E_2 = \sum_{k=1}^{N} e_k^2 = \sum_{k=1}^{N} (u_k^{(meas)} - u_k^{(theor)})^2 = \sum_{k=1}^{N} (u_k^{(meas)} - \sum_{n=1}^{M} B_n G_{kn})^2 \quad .$$

The minimization of this function gives the normal equation of the Gaussian Least Squares method

$$\underline{\underline{\mathbf{G}}}^{\mathbf{T}} \underline{\underline{\mathbf{G}}} \vec{B} = \underline{\underline{\mathbf{G}}}^{\mathbf{T}} \vec{u}^{(meas)}$$

resulting in the solution

$$\vec{B} = (\underline{\underline{\mathbf{G}}}^{\mathbf{T}} \underline{\underline{\mathbf{G}}})^{-1} \underline{\underline{\mathbf{G}}}^{\mathbf{T}} \vec{u}^{(meas)} \quad .$$

In the knowledge of the expansion coefficients, the estimated spectrum is given as

$$U^{estimated}(\omega) = \sum_{n=1}^{M} B_n \; P_n(\omega) \qquad (62)$$

at any frequency in the relevant $(-\omega_{\max}, \omega_{\max})$ interval. The inversion-based Fourier Transformation procedure described above is referred to as Legendre Polynomial based Least Square FT method, abbreviated as L-LSQ-FT. As explained above, the L-LSQ-FT inversion algorithm development initially minimizes the L_2-norm of the deviation vector between the observed and calculated data through the Gaussian Least Squares Method (LSQ) Method, which is normal for data noise following regular distribution. Unfortunately, most geophysical data contains irregular noise with randomly occurring outliers making the Least-Squares Method (LSQ) less effective for processing. An outlier is a data point that is different from the remaining data (Barnett and Lewis 1994). Outliers are also referred to as abnormalities, discordant, deviants and anomalies (Aggarwal, 2013). Whereas data noises are measurements that are not related to conditions within the subsurface. An outlier is a broader concept that includes not only errors but also discordant data that may arise from the natural variation within a population or process. As such, outliers often contain interesting and useful information about the underlying system. The consequences of not screening the data for outliers can be catastrophic for geophysical interpretations. The negative effects of outliers can be summarized into three: (1) increase in error variance and reduction in statistical power of data (2) decrease in normality for the cases where outliers are non-randomly distributed (3) model bias by corrupting the true relationship between exposure and outcome (Osborne and Overbay, 2004). Hence, the need to weight the data by a robust approach for a

better result. To develop a robust algorithm (the L-IRLS-FT), the weighted norm of the deviation vector was minimized using Cauchy-Steiner weights while the discretization of the Fourier spectrum uses Legendre polynomials as basis functions. Applying the general Jacobian matrix derived from the inverse Fourier transform in 1D we find as above

$$G_{k,n} = \mathrm{F}_k^{-1}\left\{P_n(\omega)\right\}.$$

By defining the inversion method, the theoretical value of the signal at a time point t_k is

$$u^{theor}(t_k) = u_k^{theor} = \sum_{i=1}^{M} B_n G_{k,n}$$

and the k-th element of the data deviation vector is written as

$$e_k = u_k^{(meas)} - u_k^{(theor)} = u_k^{(meas} - \sum_{n=1}^{M} B_n\, G_{kn}\,.$$

The IRLS inversion procedure applied follows Dobróka et al, 2012 as discussed earlier where the minimized weighted norm is given as

$$E_w = \sum_{k=1}^{N} w_k e_k^2$$

Where w_k are the Cauchy-Steiner weights given by

$$w_k = \frac{\varepsilon^2}{\varepsilon^2 + e_k^2},$$

where ε_{j+1}^2 the Steiner's scale factor is determined iteratively. From earlier discussions, the misfit function is non-quadratic in the case of Cauchy-Steiner weights making the inverse problem nonlinear which can be solved by applying the method of the Iteratively Reweighted Least Squares (Scales, 1988). In the first iteration, the misfit function

$$E_w^{(0)} = \sum_{k=1}^{N} e_k^2$$

is minimized (Gaussian Least Squares) resulting in the linear set of normal equations

$$\mathbf{G}^T\mathbf{G}\vec{B}^{(0)} = \mathbf{G}^T\vec{u}^{measured} \quad \text{giving} \quad \vec{B}^{(0)} = (\mathbf{G}^T\mathbf{G})^{-1}\mathbf{G}^T\vec{u}^{measured}\,.$$

The data deviation is

DOI: 10.14750/ME.2020.008

$$e_k^{(0)} = u_k^{(meas} - \sum_{n=1}^{M} B_n^{(0)} G_{kn} .$$

resulting in the weights

$$w_k^{(0)} = \frac{\varepsilon^2}{\varepsilon^2 + (e_k^{(0)})^2}$$

and the new misfit function

$$E_w^{(1)} = \sum_{k=1}^{N} w_k^{(0)} (e_k^{(1)})^2$$

where

$$e_k^{(1)} = u_k^{(meas} - \sum_{n=1}^{M} B_n^{(1)} G_{kn} .$$

The minimization of $E_w^{(1)}$ results in a weighted least squares problem with the linear set of the normal equation

$$\boldsymbol{G}^T \boldsymbol{W}^{(0)} \boldsymbol{G} \vec{B}^{(1)} = \boldsymbol{G}^T \boldsymbol{W}^{(0)} \vec{u}^{measured}$$

where the $\mathbf{W}^{(0)}$ weighting matrix (independent of $\vec{B}^{(1)}$) is of the diagonal form $W_{kk}^{(0)} = w_k^{(0)}$. Solving the normal equation one finds

$$\vec{B}^{(1)} = (\mathbf{G}^T \mathbf{W}^{(0)} \mathbf{G})^{-1} \mathbf{G}^T \mathbf{W}^{(0)} \vec{u}^{measured}$$

with

$$u_k^{(1)} = \sum_{i=1}^{M} B_i^{(1)} G_{ki} , \qquad e_k^{(1)} = u_k^{measured} - u_k^{(1)} , \qquad w_k^{(1)} = \frac{\varepsilon^2}{\varepsilon^2 + (e_k^{(1)})^2} .$$

and so on, till the proper stop criterion is met. The described inversion-based Fourier Transformation procedure above is called Legendre Polynomial based Iteratively-Reweighted Least Square FT method, abbreviated as L-IRLS-FT.

3.2.1 Numerical testing in 1D

A time-domain signal (Figure 2) was created to test the noise reduction capability of the newly developed method, L-LSQ-FT and the traditional DFT in one dimension. The noiseless time function of the test data can be described by the formula below

$$u(t) = \kappa t^{\eta} e^{-\lambda t} \sin(\omega t + \varphi) \qquad (64)$$

DOI: 10.14750/ME.2020.008

where the Greek letters represent the parameters of the signal. Specified fixed values for the signal parameters as follows: $\kappa \approx 738.91$, $\eta = 2$, $\lambda = 20$, $\omega = 40\pi$, $\varphi = \pi / 4$.

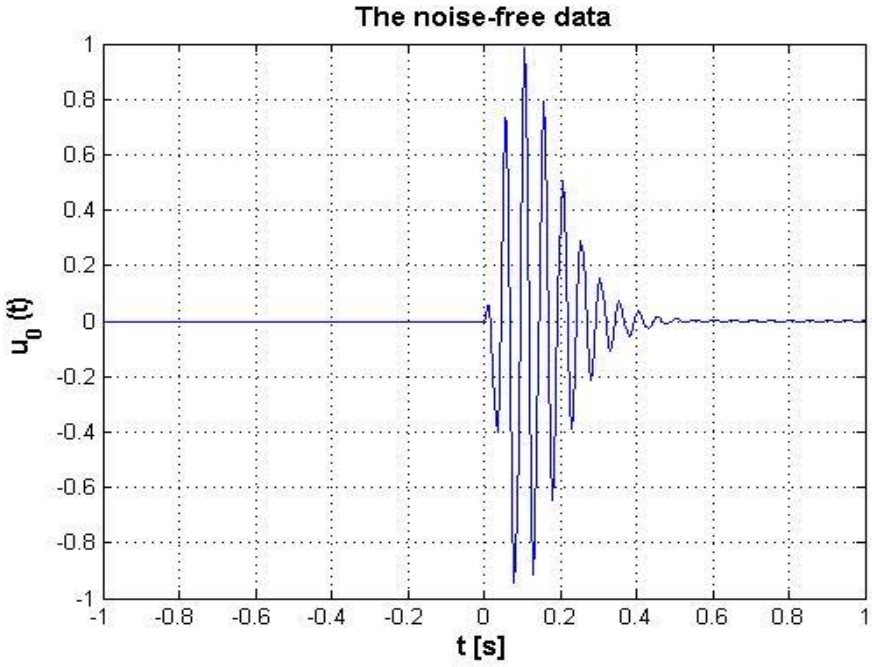

Figure 2; Calculated noise-free waveform

The noise-free waveform was sampled at regular intervals of 0.005 (sec) measurement points ranging over the time interval of [-1, 1] and processed using the traditional DFT method to give both the real and imaginary parts of the noise-free Fourier spectrum (Figure 3). The same noiseless waveform was also processed using the L-LSQ-FT method. The resultant processed signal is shown in Figure 4. The L-LSQ-FT spectrum was calculated using Legendre polynomials of the (maximal) order of M=300. For numeric reasons, the calculated Fourier spectra were made on the data set transformed to [-1,1] in both x and y coordinates resulting in an appropriate scale in the wavenumber domain. Both the traditional DFT and the L-LSQ-FT gave similar real and imaginary parts for the Fourier transformed spectrum. This demonstrates the effectiveness of both methods in processing noise-free data.

Following the successful application of both methods to the noise-free signal, Gaussian and Cauchy noise were introduced into the noise-free signal (Figure 2) for processing. Gaussian noise is a statistical noise having a probability distribution function equal to that of the normal distribution, which is also known as the Gaussian distribution. In geophysical applications, this type of noise distribution is occasionally encountered in the data processing. Its distribution is

DOI: 10.14750/ME.2020.008

symmetric and completely characterized by the Mean and Variance of the data. The Gaussian noisy signal with 0 mean and 0.01 variance is given in Figure 5.

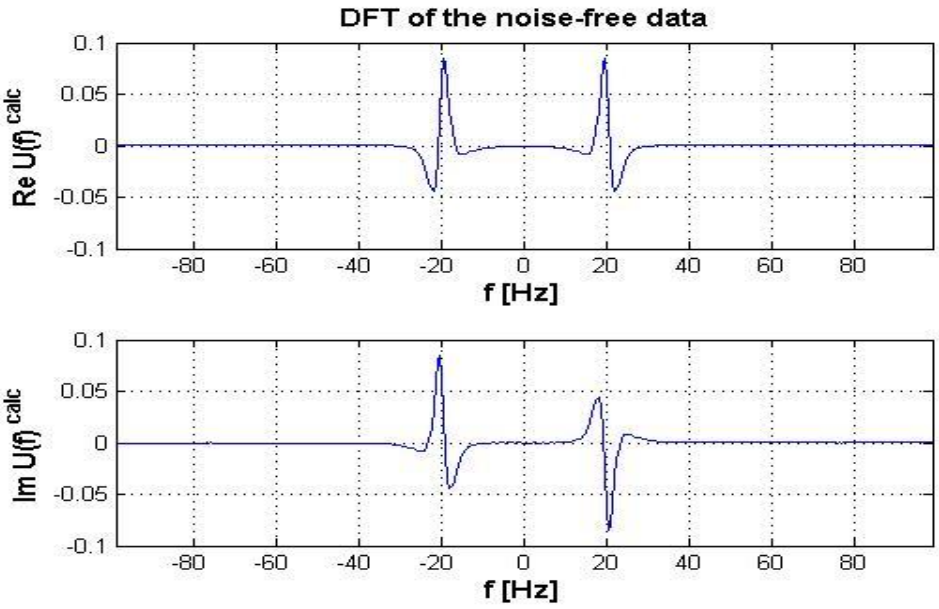

Figure 3; Processed DFT spectrum of the noise-free Morlet waveform

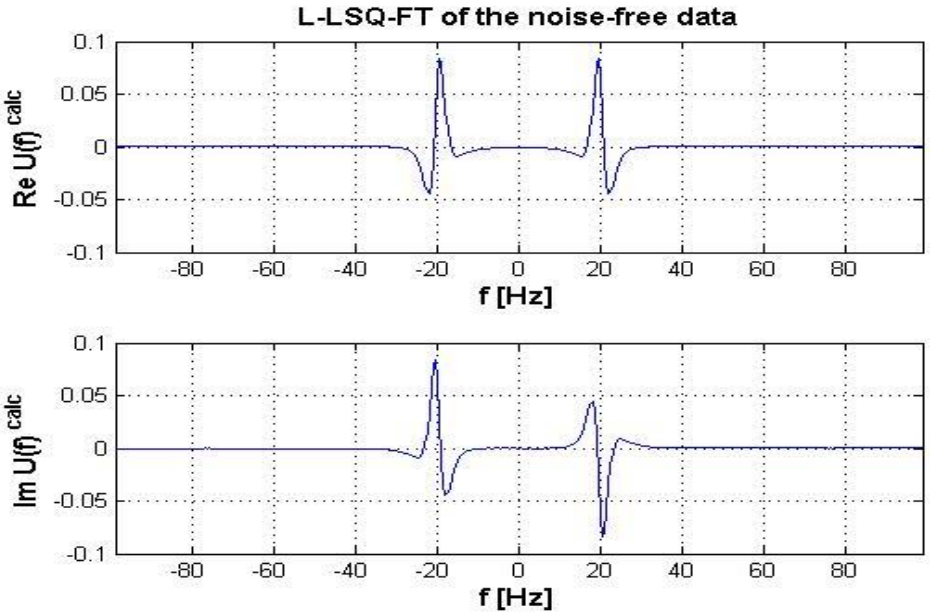

Figure 4; Processed L-LSQ-FT spectrum of the noise-free Morlet waveform

Random noise, on the other hand, is noise distributions in data that do not follow a regular distribution across a survey area. This type of noise is mostly introduced into survey data from external sources such as data acquisition or survey designs and equipment limitations.

DOI: 10.14750/ME.2020.008

They are inherent in geophysical data and are not related to the subsurface body of interest. Random noise reduction is a critical step to improve the signal to noise ratio in geophysical applications with several methods developed over the years to achieve this purpose (Liu et, al. 2006, Al-Dossary and Marfurt 2007, Liu, Liu and Wang 2009). This includes the development of filters using various forms of transforms such as Wavelet Transform (Deighan and Watts 1997), S-Transform (Askari and Siahkooli 2008) and Fourier Transform (Dobroka et, al. 2012). Failure to adequately suppress random noise affects the quality of processed data and interpretation.

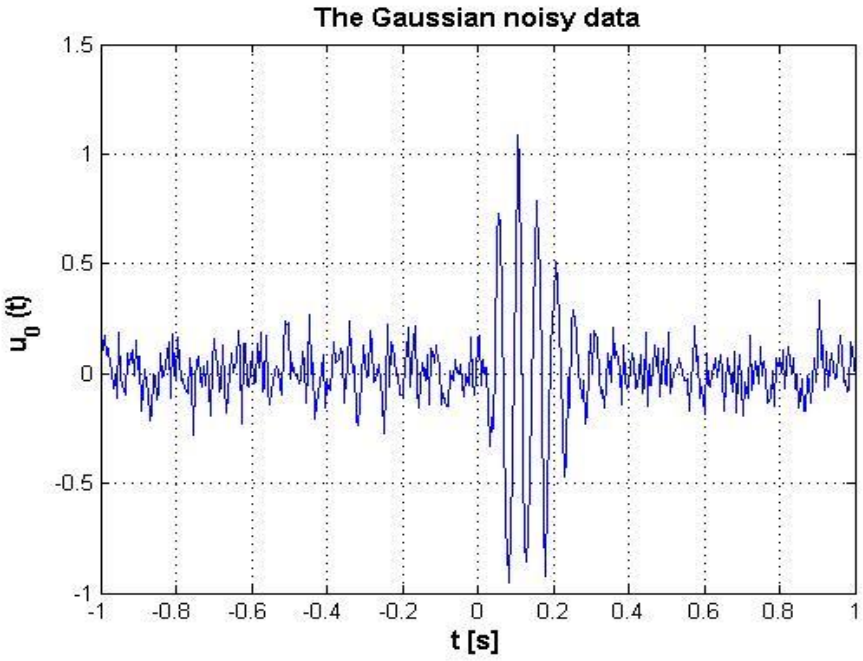

Figure 5; The generated noisy signal with Gaussian noise

Random noise following Cauchy distribution was added to the Morlet waveform to produce a noisy signal (Figure 6) for processing. To demonstrate the noise reduction capability of the two methods, the Gaussian noisy signal (Figure 5) was processed with the traditional DFT and the L-LSQ-FT methods. The resultant transformed spectra in the real and imaginary form are shown in Figures 7 and 8 for the DFT and L-LSQ-FT respectively. We further processed the Cauchy noisy signal (Figure 6) with both methods to give the resultant transformed spectra for DFT and the L-LSQ-FT methods in Figures 9 and 10 respectively. The output signals show a considerable suppression of Gaussian and Cauchy noise by the L-LSQ-FT method compared to the traditional DFT method. For the processed Cauchy noisy signal, a comparison between the real and imaginary spectrum as produced from the traditional DFT (Figure 9) and the L-LSQ-FT (Figure 10) shows not much improvement in output Fourier

spectra in both methods. Although the L-LSQ-FT algorithm was able to reject a substantial amount of the Cauchy noise, it still has some amount of noise at its extreme ends.

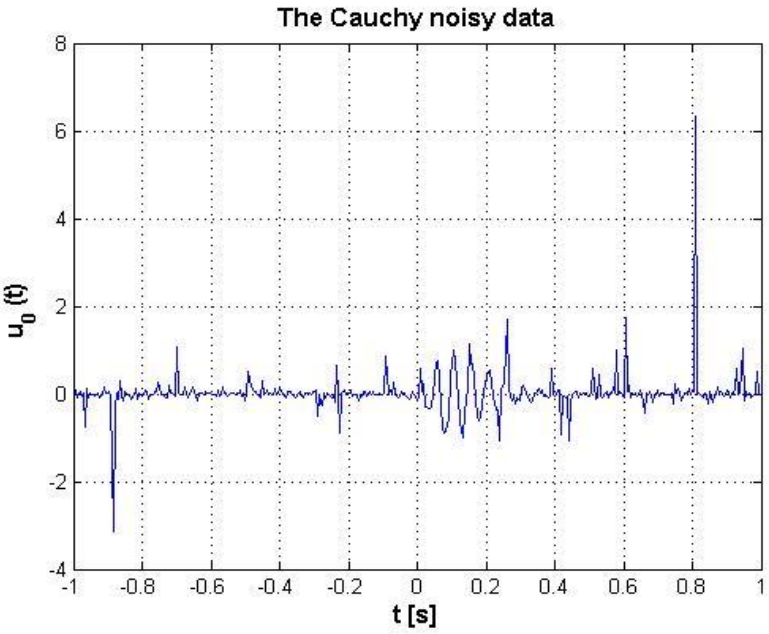

Figure 6; The generated noisy signal with Cauchy noise

For quantitative characterization of the results, we introduce the RMS distance between (a) and (b) data sets (for example noisy and noiseless) in the time domain (data distance)

$$d_{RMS} = \sqrt{\frac{1}{N}\sum_{k=1}^{N}\left(u^{(a)}(t_k) - u^{(b)}(t_k)\right)^2},$$

as well as the frequency domain (model or spectra distance)

$$D_{RMS} = \sqrt{\frac{1}{N}\sum_{k=1}^{N}\left\{\left(\mathrm{Re}\left[U^{(a)}(f_k) - U^{(b)}(f_k)\right]\right)^2 + \left(\mathrm{Im}\left[U^{(a)}(f_k) - U^{(b)}(f_k)\right]\right)^2\right\}}.$$

In the case of the Gaussian noise, the distance between the noisy and noiseless data sets, d = 0.1032. The model or spectra distance between the DFT spectrum (Figure 7) of the noisy (contaminated with Gaussian noise) and the noiseless data sets gave D = $1.03*10^{-2}$. Figure 8 represents sufficient improvement characterized by the spectra distance between the noiseless and the noisy (given by L-LSQ-FT) spectra: D = $8.2*10^{-3}$. Similarly, the DFT gave a spectra distance D=$4.16*10^{-2}$ for spectrum produced from the noisy Cauchy signal whilst the L-LSQ-FT gave a spectra distance: D=$2.43*10^{-2}$. From the above analyses, a higher noise reduction capability was exhibited by the L-LSQ-FT method compared to the traditional DFT method.

DOI: 10.14750/ME.2020.008

The results demonstrate the outlier and random noise sensitivity of the DFT and to some extent, the Least Squares Methods, hence the need to define a more robust method for outliers and random noise suppression. We, therefore, introduce the L-IRLS-FT Method.

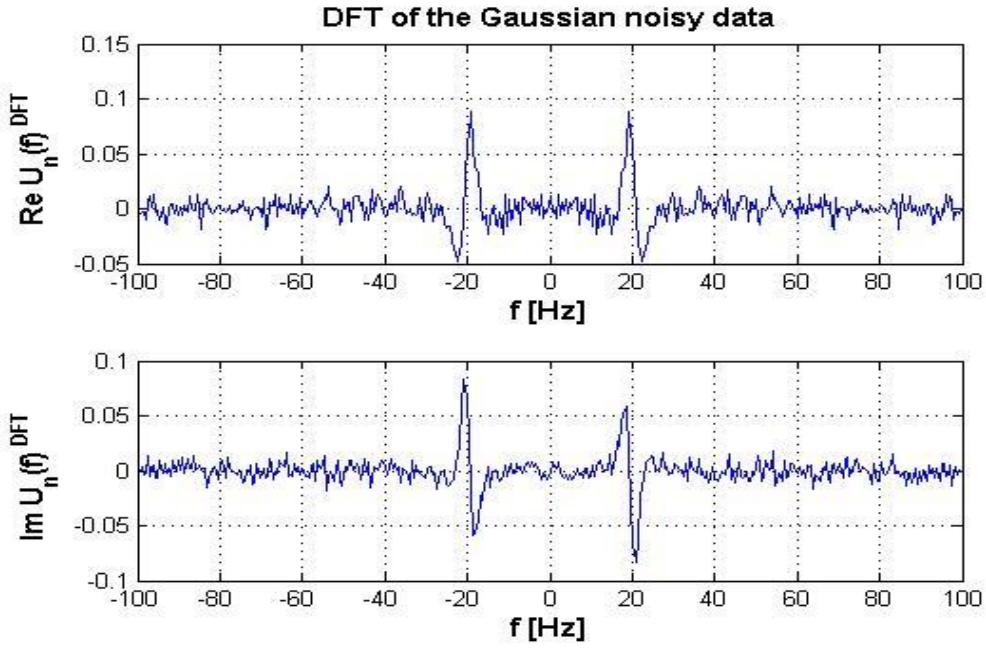

Figure 7; Processed DFT spectrum of the Gaussian noisy signal (D=1.03*10^{-2})

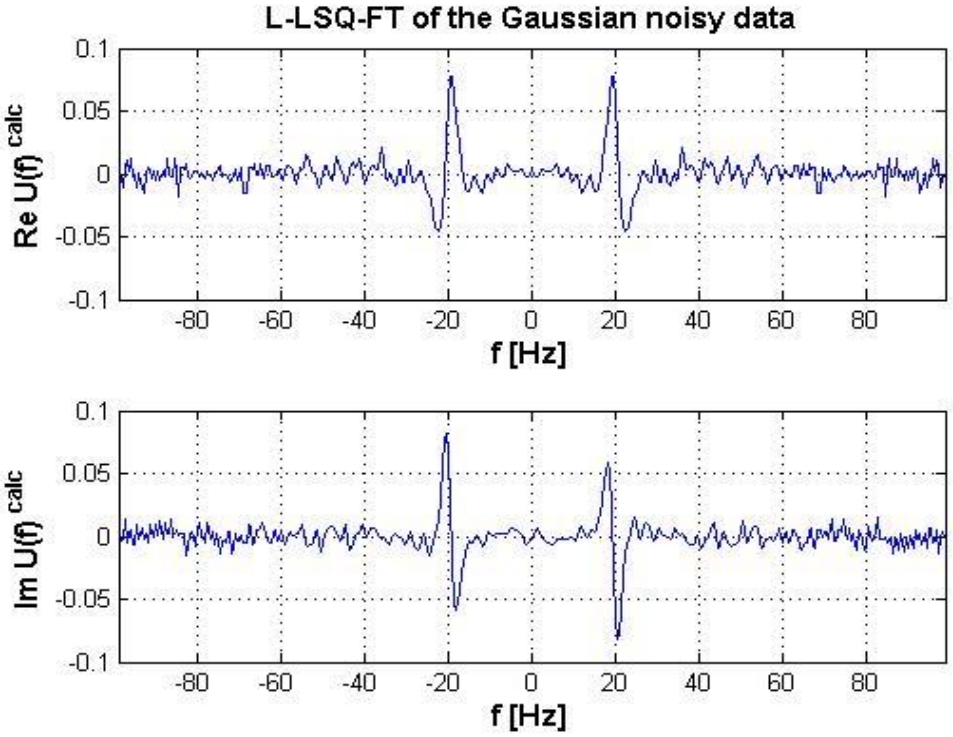

Figure 8; Processed L-LSQ-FT spectrum of the Gaussian noisy signal (D=8.2*10^{-2})

DOI: 10.14750/ME.2020.008

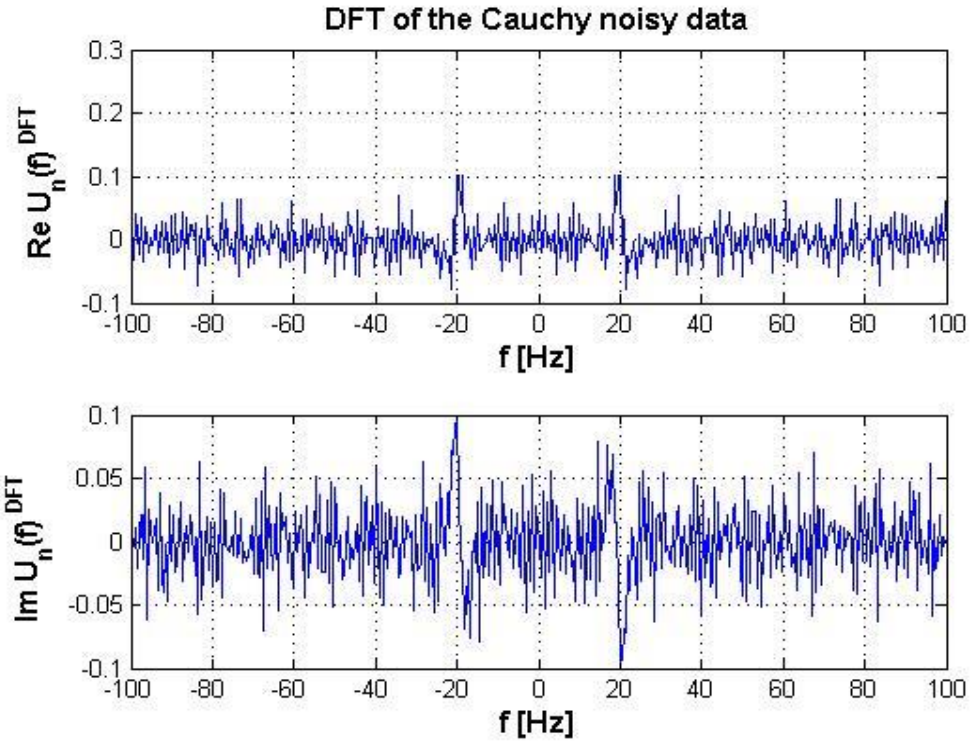

Figure 9, Processed DFT spectrum of the Cauchy noisy signal (D=4.16*10^{-2})

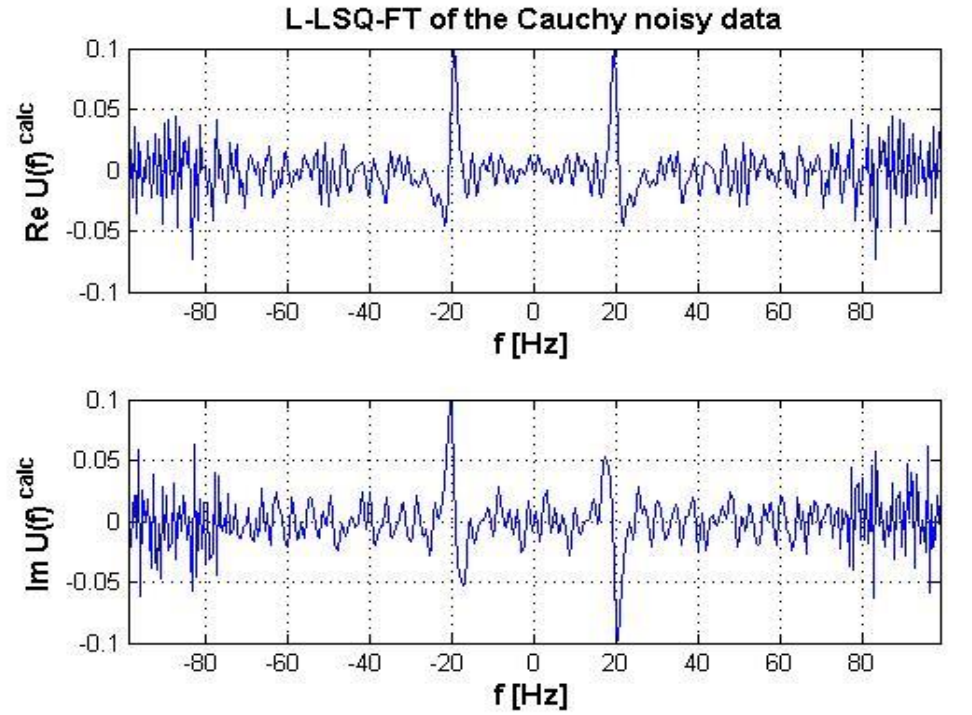

Figure 10, Processed L-LSQ-FT spectrum of the Cauchy noisy signal (D=2.43*10^{-2})

DOI: 10.14750/ME.2020.008

The same noiseless waveform as shown in Figure 2 above was processed using the L-IRLS-FT method. The resultant processed signal is shown in Figure 11. The L-IRLS-FT spectrum was calculated using Legendre polynomials of the (maximal) order of M=300. For numeric reasons, the calculated Fourier spectra were made on the data set transformed to [-1,1] in both x and y coordinates (as in the case of the L-LSQ-FT) resulting in an appropriate scale in the wavenumber domain. A comparison of the real and imaginary spectrum of the L-IRLS-FT processed noise-free signal (Figure 11) to the output signals from the traditional DFT and L-LSQ-FT (Figures 3 and 4 above) shows a very good similarity, indicating that the L-IRLS-FT algorithm was efficient in processing the noise-free signal. To test the noise reduction capability of the L-IRLS-FT, the Gaussian and Cauchy noisy signals (Figures 5 and 6) were this time processed with the L-IRLS-FT algorithm. For the Gaussian noisy signal, the output processed Fourier spectrum for DFT and L-IRLS-FT are shown in Figures 12 and 13 respectively. Also, the output processed Fourier spectrum for DFT and L-IRLS-FT for the Cauchy noisy signal are shown by Figures 14 and 15 below.

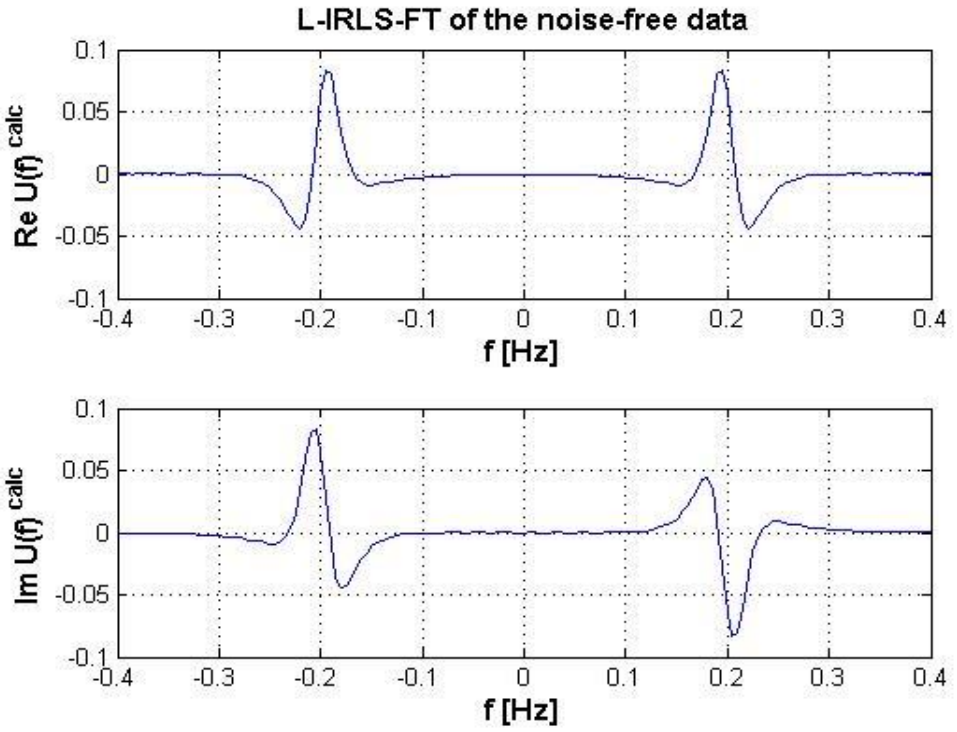

Figure 11; Processed L-IRLS-FT spectrum of the noise-free Morlet waveform

DOI: 10.14750/ME.2020.008

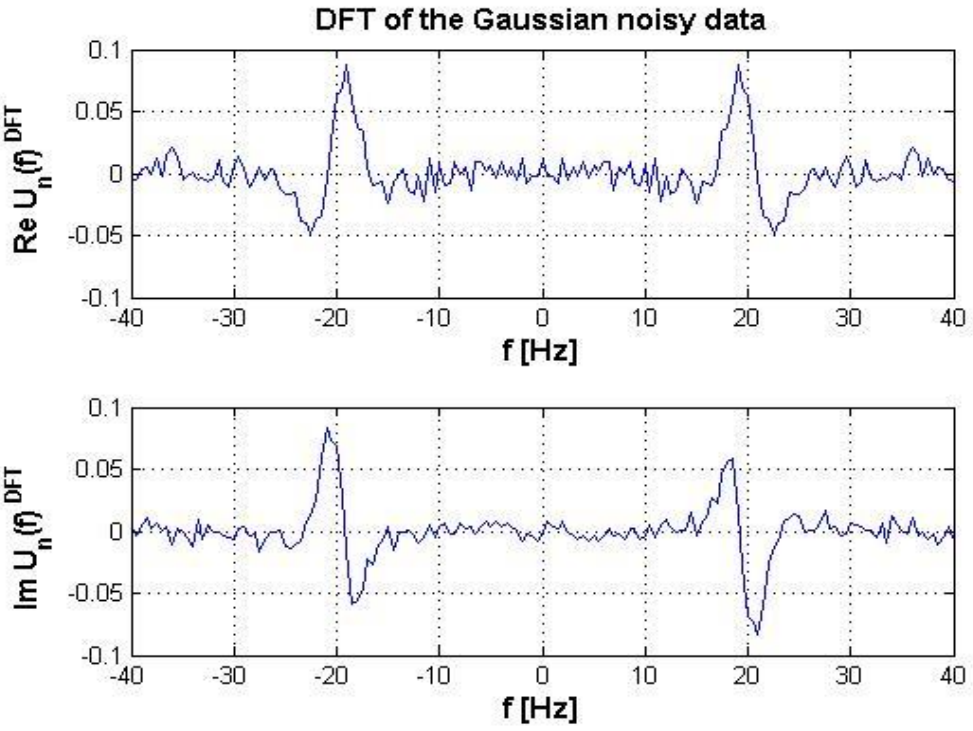

Figure 12; Processed DFT spectrum of the Gaussian noisy signal (D=4.1*10^{-3})

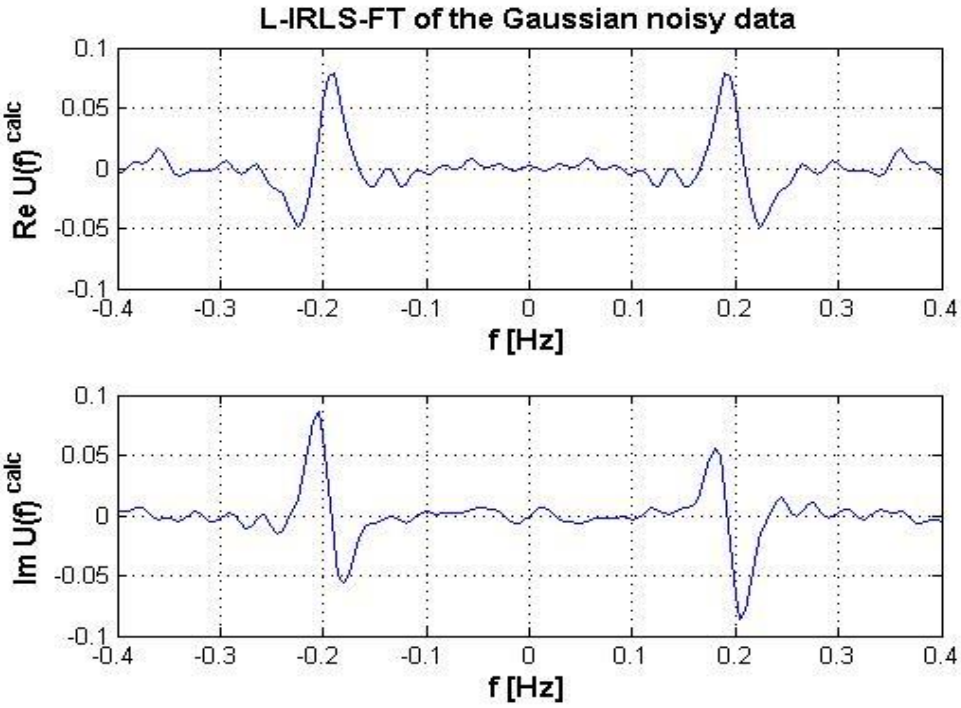

Figure 13; Processed L-IRLS-FT spectrum of the Gaussian noisy signal (D=2.6*10^{-3})

DOI: 10.14750/ME.2020.008

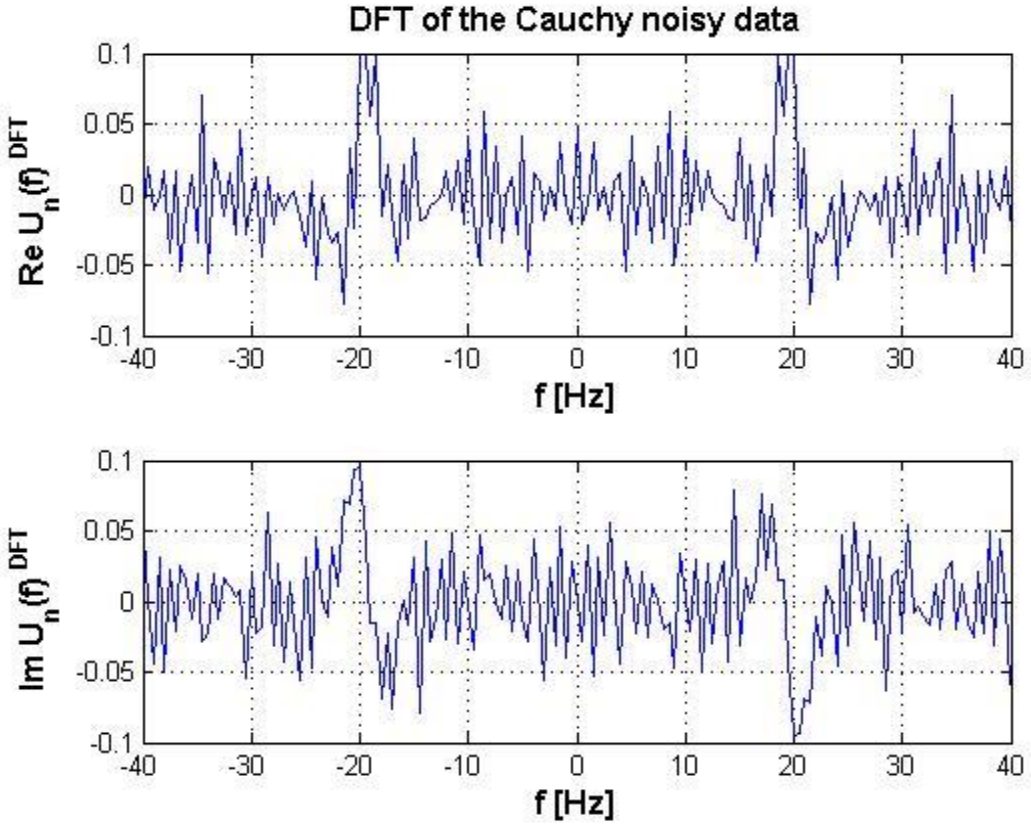

Figure 14; Processed DFT spectrum of the Cauchy noisy signal (D=4.16*10^{-2})

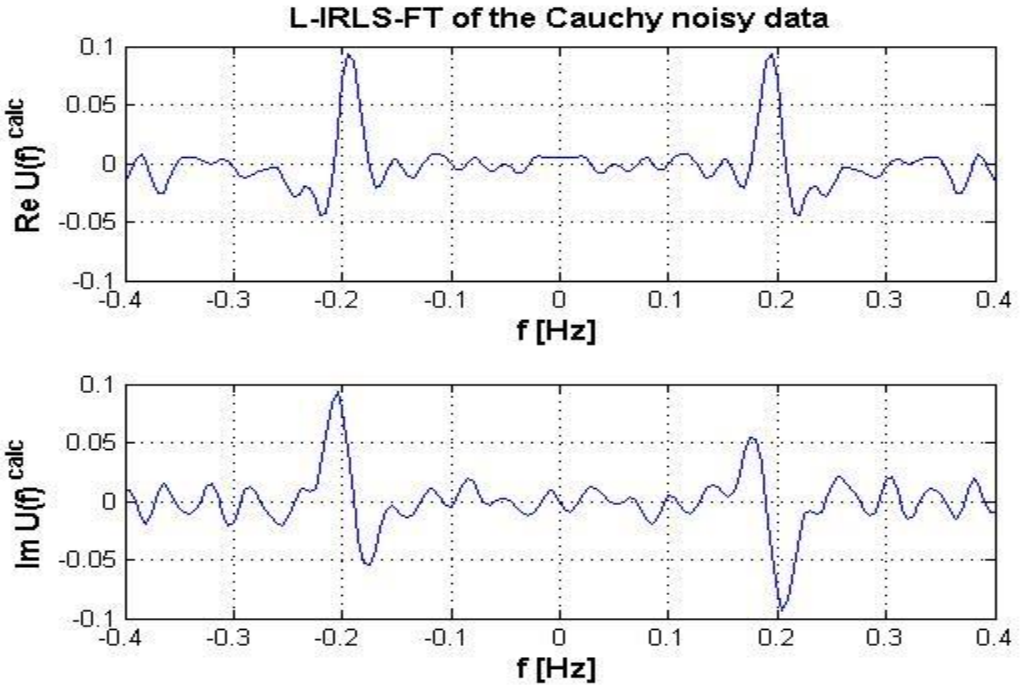

Figure 15; Processed L-IRLS-FT spectrum of the Cauchy noisy signal (D=1.32*10^{-2})

DOI: 10.14750/ME.2020.008

From the above output signals, the newly developed L-IRLS-FT algorithm was more effective in reducing the Gaussian and Cauchy noise component of the noisy signal compared to the traditional DFT. In the case of Cauchy noise, the DFT real and imaginary parts of the spectrum (Figure 14) were noisier with a lot of spikes. This goes a long way to emphasize the limitation of the traditional DFT in eliminating randomly occurring outliers and recursive random noise from a signal. To analytically characterize the results, we applied the RMS distance between two data sets (for example noisy and noiseless) in the frequency domain as well as the model or spectra distance. For processed Gaussian noisy dataset, the model or spectra distance between the DFT spectrum (Figure 12) of the noisy (contaminated with Gaussian noise) and the noiseless data sets is D = $4.1*10^{-3}$. Figure 13 represents sufficient improvement characterized by the spectra distance between the noiseless and the noisy (given by L-IRLS-FT) spectra: D = $2.6*10^{-3}$. Likewise, the DFT gave a spectra distance D=$4.16*10^{-2}$ for spectrum produced from the noisy Cauchy signal whilst the L-IRLS-FT gave a spectra distance of D=$1.32*10^{-2}$. From the above analyses, the L-IRLS-FT method compared to the traditional DFT method showed a higher noise reduction capability when both regular and irregular noise was added to the Morlet waveform for processing. The results fully demonstrate the outlier and random noise sensitivity of the traditional DFT method. Hence, we propose a new method, the L-IRLS-FT which is robust and resistant enough to suppress randomly occurring data noise. Based on the successful application of the L-LSQ-FT and the L-IRLS-FT, it was necessary to compare the results to that of the original H-LSQ-FT and H-IRLS-FT which forms the basis of the inverse Fourier transform method development. To do that, we first processed the same noise-free Morlet waveform (Figure 2) with the H-LSQ-FT and H-IRLS-FT. The real and imaginary parts of the processed spectrum are given in Figures 16 and 17 respectively. Equally, a comparison to the DFT processed spectrum (Figure 3) shows that the H-LSQ-FT and H-IRLS-FT were efficient in processing the noise-free signal.

We further processed the Gaussian and Cauchy noisy signals (Figures 5 and 6) with the H-LSQ-FT and the H-IRLS-FT to give the resultant spectrum for processed Gaussian and Cauchy noisy signal in Figures 18 and 19 respectively for the H-LSQ-FT. Also, the output processed Fourier spectrum for H-IRLS-FT for the Gaussian and Cauchy noisy signals are shown in Figures 20 and 21 below.

DOI: 10.14750/ME.2020.008

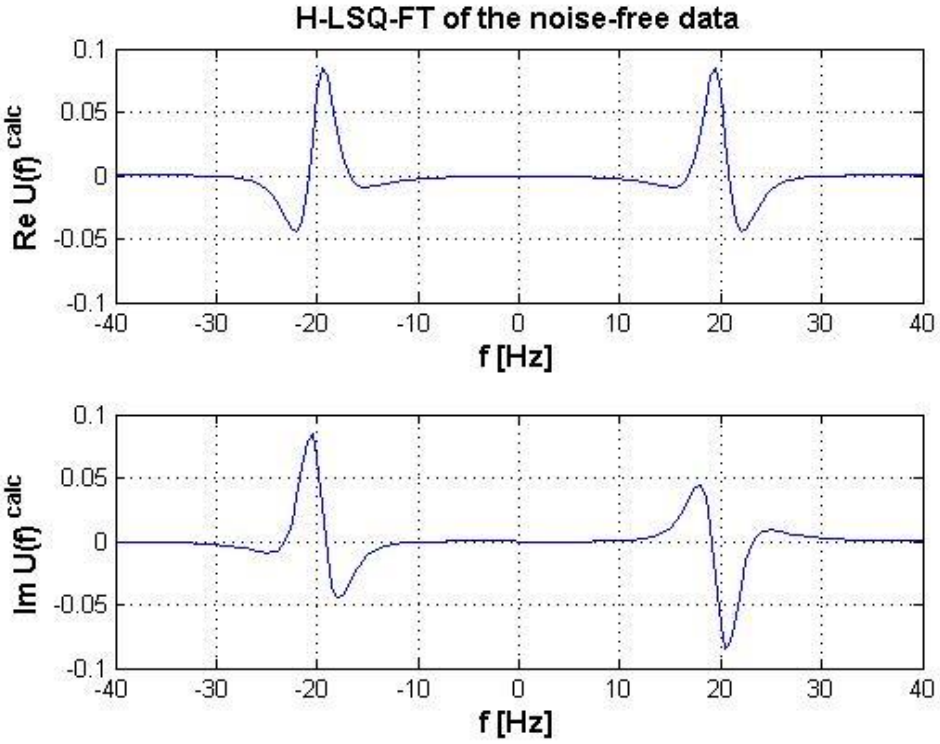

Figure 16; Processed H-LSQ-FT spectrum of the noise-free Morlet waveform

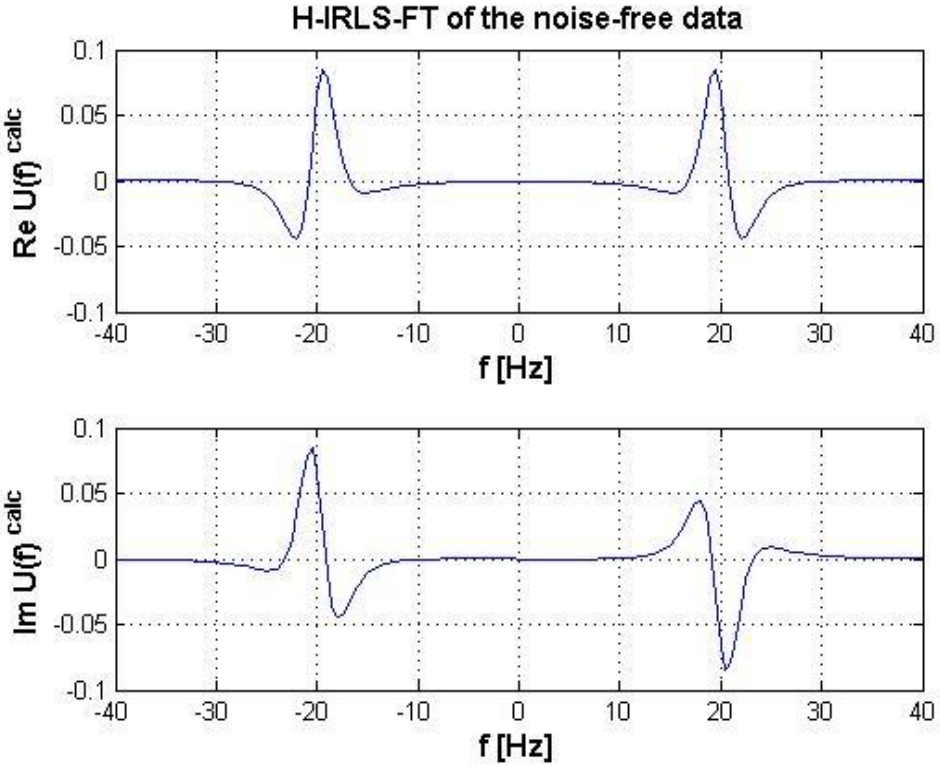

Figure 17; Processed H-IRLS-FT spectrum of the noise-free Morlet waveform

DOI: 10.14750/ME.2020.008

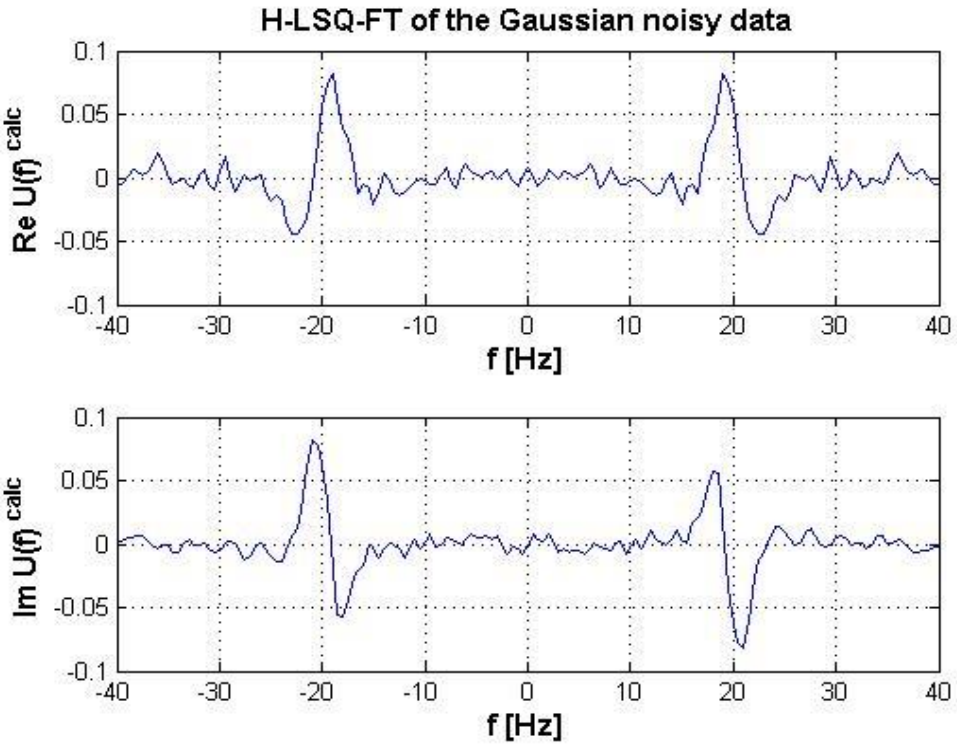

Figure 18; Processed H-LSQ-FT spectrum of the Gaussian noisy signal (D=6.2*10^{-3})

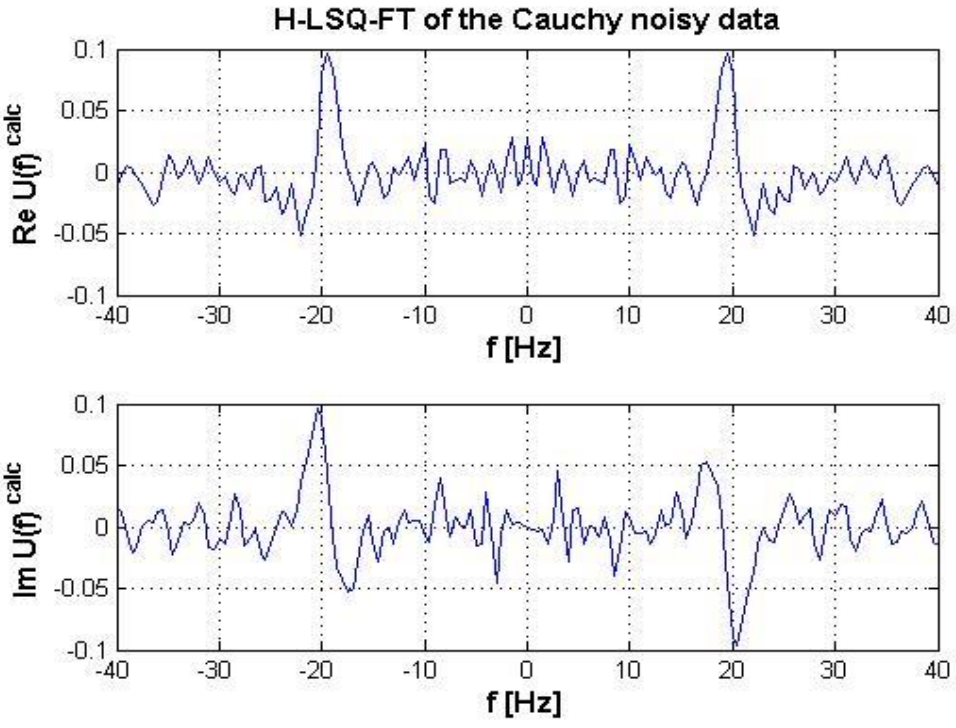

Figure 19; Processed H-LSQ-FT spectrum of the Cauchy noisy signal (D=1.23*10^{-2})

DOI: 10.14750/ME.2020.008

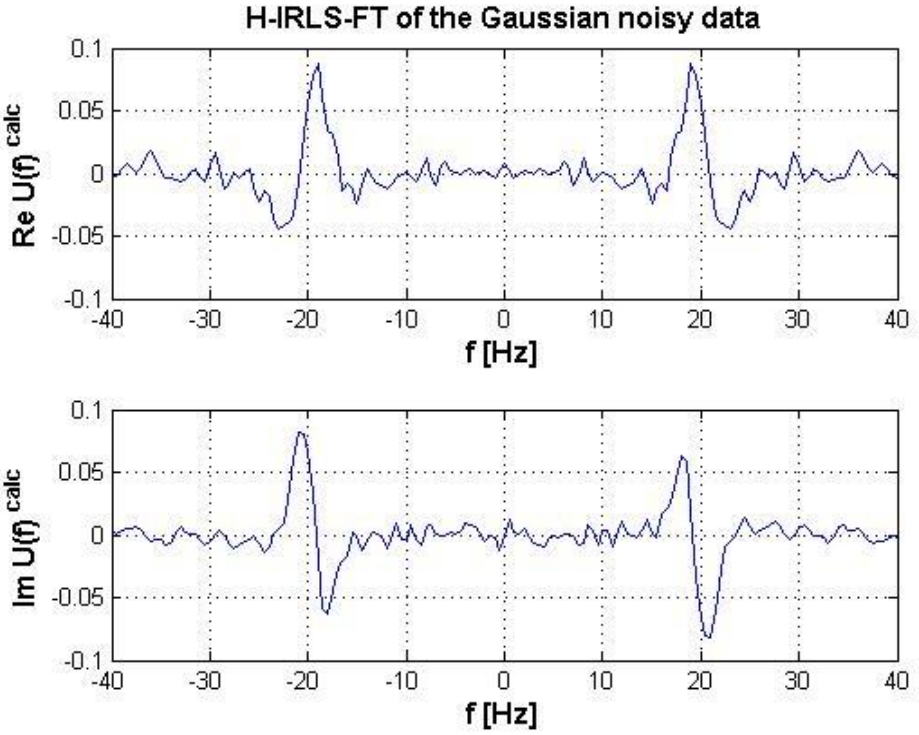

Figure 20; Processed H-IRLS-FT spectrum of the Gaussian noisy signal (D=6.5*10^{-3})

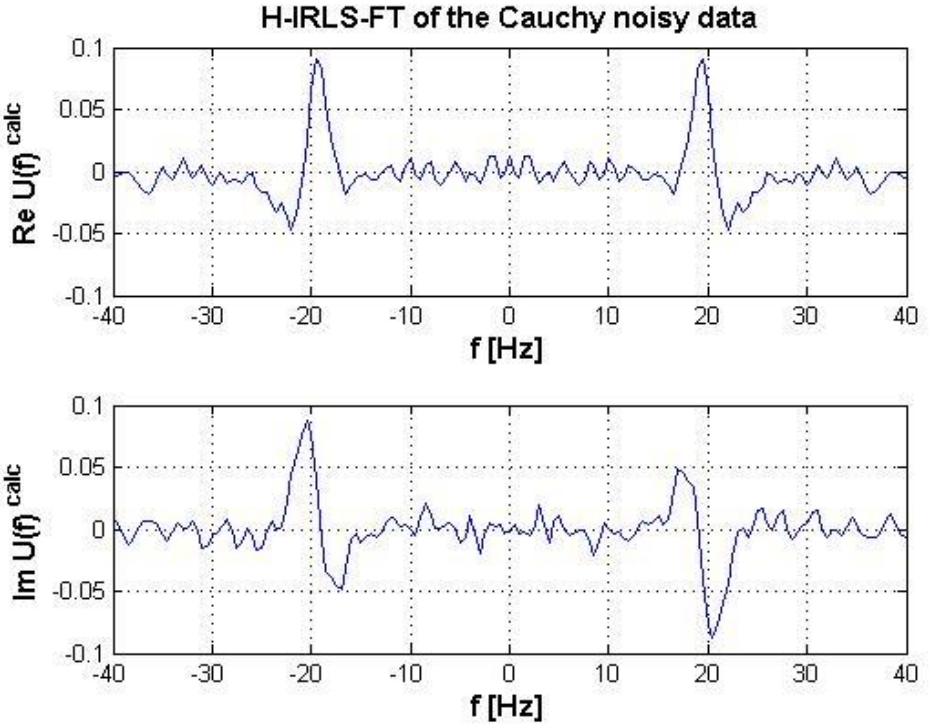

Figure 21; Processed H-IRLS-FT spectrum of the Cauchy noisy signal (D=6.8*10^{-3})

DOI: 10.14750/ME.2020.008

To analytically characterize and compare the results of the newly developed Legendre polynomial bases Fourier transformation and the existing Hermite-based Fourier transformation, we applied the RMS distance between two data sets in the frequency domain as well as the model or spectra distance. Table 2 shows the calculated spectra distances after various noise was introduced into the noise-free Morlet waveform and processed. It was observed that the Hermite based inverse Fourier transformation was more efficient in its Least Square and iteratively reweighted form than the Legendre polynomial bases Fourier transformation, especially in the suppression of Cauchy noise. The H-IRLS-FT gave a spectra distance of $6.8*10^{-3}$ when random Cauchy noise was added to the data while the L-IRLS-FT quantified the spectra distance as $1.32*10^{-2}$. This is so because the applied Hermite functions are the Eigenfunctions of the inverse Fourier transform while the Legendre polynomials are not. Comparatively, the newly developed L-LSQ-FT and L-IRLS-FT can be considered as a better alternative to the traditional DFT as demonstrated above but complementary support to the original H-LSQ-FT and H-IRLS-FT.

Table 2. Calculated Spectra Distance for L-LSQ-FT, H-LSQ-FT, L-IRLS-FT, and H-IRLS-FT Methods

	CALCULATED SPECTRA DISTANCE	
APPLIED METHOD	**GAUSSIAN NOISE**	**CAUCHY NOISE**
L-LSQ-FT	$8.2*10^{-3}$	$2.43*10^{-2}$
H-LSQ-FT	$6.2*10^{-3}$	$1.23*10^{-2}$
L-IRLS-FT	$2.6*10^{-3}$	$1.32*10^{-2}$
H-IRLS-FT	$6.5*10^{-3}$	$6.8*10^{-3}$

DOI: 10.14750/ME.2020.008

3.3 The L-LSQ-FT and L-IRLS-FT algorithm in 2D

To develop the L-LSQ-FT and L-IRLS-FT algorithm in 2D, the same inversion procedures in the one-dimensional case was followed. The general form of the Jacobi matrix in the case of two-dimensional series expansion based inverse Fourier transform is given as

$$G_{nm}^{kl} = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \psi_n(\omega_x) e^{j\omega_x x_k} d\omega_x \, . \, \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \psi_m(\omega_y) e^{j\omega_y y_l} d\omega_y$$

where G_{nm}^{kl} is an element of the Jacobian matrix of the size N-by-M. The Jacobian matrix is the inverse Fourier transform of the two basis functions ψ_n and ψ_m . Parameterization of the model is achieved by introducing the Legendre polynomials (equation 57) as basis functions to give,

$$G_{nm}^{kl} = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} P_n(\omega_x) e^{j\omega_x x_k} d\omega_x \, . \, \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} P_m(\omega_y) e^{j\omega_y y_l} d\omega_y \tag{67}$$

or in a more formal notation

$$G_{n,m}^{kl} = \mathcal{F}_k^{-1}\{P_n(\omega)\}.\mathcal{F}_l^{-1}\{P_m(\omega)\} \tag{68}$$

We then introduce a new Legendre polynomial 2D inversion-based Fourier Transformation method to calculate the inverse FT of eq. (67) by using the frequently used 2D inverse DFT procedure:

$$G_{n,m}^{kl} = IDFT\{P_n(\omega)\}.IDFT\{P_m(\omega)\} \tag{69}$$

By using this procedure the $G_{n,m}^{kl}$ elements of the Jacobi matrix can numerically be generated. At this point, a new inversion method is to be defined. The theoretical value of the data at point (x_k, y_l) is given as

$$\mathrm{u}^{theor}(x_k, y_l) = u_s^{theor} = \sum_{n=1}^{N} \sum_{m=1}^{M} B_{n,m} \, G_{n,m}^{kl}$$

Which can be simplified further to

$$u_s^{theor} = \sum_{i=1}^{I} B_i G_{s,i}$$

(where $i = n + (m-1)N, s = k + (l-1)k$)

The general element of the deviation vector can be given in the form

$$e_s = u_s^{(meas)} - u_s^{(theor)} = u_s^{(meas)} - \sum_{i=1}^{I} B_i G_{s,i}$$

By means of L_2-norm, the misfit function is given as

$$E_2 = \sum_{k=1}^{N} e_k^2 = \sum_{k=1}^{N} (u_k^{(meas)} - u_k^{(theor)})^2 = \sum_{k=1}^{N} (u_k^{(meas)} - \sum_{n=1}^{N} \sum_{m=1}^{M} B_{n,m} G_{nm}^{kl})^2.$$

The minimization of this function gives the normal equation of the Gaussian Least Squares method

$$\underline{\underline{\mathbf{G}}}^{\mathbf{T}} \underline{\underline{\mathbf{G}}} \vec{B} = \underline{\underline{\mathbf{G}}}^{\mathbf{T}} \vec{u}^{(meas)}$$

resulting in the solution

$$\vec{B} = (\underline{\underline{\mathbf{G}}}^{\mathbf{T}} \underline{\underline{\mathbf{G}}})^{-1} \underline{\underline{\mathbf{G}}}^{\mathbf{T}} \vec{u}^{(meas)} .$$

In the knowledge of the expansion coefficients, the estimated spectrum is given as

$$U^{estimsted}(\omega) = \sum_{n=1}^{N} \sum_{m=1}^{M} B_{n,m} P_n(\omega) P_m(\omega). \quad (70)$$

The inversion-based Fourier Transformation procedure described above is referred to as 2D Legendre Polynomial based Least Square FT method, abbreviated as 2D L-LSQ-FT. To develop a robust algorithm for outlier suppression, the IRLS method was this time applied and the 2D Fourier frequency spectrum discretized using Legendre Polynomials. This resulted in a new Legendre Polynomial based 2D Iteratively Reweighted Least Square FT method, abbreviated as 2D L-IRLS-FT. Robust inversion techniques such as the IRLS yield good performance when data is drawn from a wide range of probability distributions that are largely unaffected by outliers or small departures from model assumptions in a given dataset. These methods have two distinctive aspects, in that they can effectively ignore erroneous measurements during inversion and are very easy to implement. As explained in the 1D case, the methods work by applying a diagonal weighted matrix to data residuals based on their statistics. The estimation of weights can be completely automatic or iterative and relies on the assumption that the noisy measurements are statistically insignificant. As explained earlier, the N-by-M Jacobian matrix for two-dimensional series expansion based inverse Fourier transform is given as

$$G_{nm}^{kl} = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \psi_n(\omega_x) e^{j\omega_x x_k} d\omega_x \, . \, \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \psi_m(\omega_y) e^{j\omega_y y_l} d\omega_y$$

DOI: 10.14750/ME.2020.008

Parameterization of the model is achieved by introducing the Legendre polynomials (equation 57) as basis function to give,

$$G_{nm}^{kl} = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} P_n(\omega_x) e^{j\omega_x x_k} d\omega_x \,.\; \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} P_m(\omega_y) e^{j\omega_y y_l} d\omega_y$$

or in a more formal notation

$$G_{n,m}^{kl} = \mathcal{F}_k^{-1}\{P_n(\omega)\}.\mathcal{F}_l^{-1}\{P_m(\omega)\}$$

We then introduce a new Legendre polynomial Iteratively Reweighted Least Square inversion-based Fourier Transformation method to calculate the 2D inverse FT of the general Jacobian matrix by using the frequently used 2D inverse DFT procedure:

$$G_{n,m}^{kl} = 2D\ IDFT\{P_n(\omega)\}.2D\ IDFT\{P_m(\omega)\} \tag{71}$$

The $G_{n,m}^{kl}$ elements of the Jacobi matrix can numerically be generated. We then define a new inversion method with a theoretical value of the data at point (x_k,y_l) given as

$$\mathrm{u}^{theor}(x_k, y_l) = u_s^{theor} = \sum_{n=1}^{N} \sum_{m=1}^{M} B_{n,m}\, G_{n,m}^{kl}$$

The general element of the deviation vector can be given in the form

$$e_s = u_s^{(meas)} - u_s^{(theor)} = u_s^{(meas)} - \sum_{n=1}^{N} \sum_{m=1}^{M} B_{n,m}\, G_{n,m}^{kl}$$

At this stage, we introduce the IRLS algorithm which is characterized by a weight change from iteration to iteration and requires the minimization of the L-p norm of the deviation vector for which solutions of p<2 are of practical importance in geophysical applications. The applied weights are inversely proportional to the deviation vector between the calculated and observed data. As a consequence, the noisier data contribute to the solution to a less degree. The method generates a nonlinear set of equations which are linearized. This assures that a linear set of equations is solved in each step of the iterative procedure. In the first step, the weighting matrix is chosen as an identity matrix and the model vector estimated as demonstrated in 1D. The minimized weighted norm is given as

$$E_w = \sum_{k=1}^{N} w_k e_k^2$$

DOI: 10.14750/ME.2020.008

Where $'w_k'$ is the Cauchy-Steiner weights given by

$$w_k = \frac{\varepsilon^2}{\varepsilon^2 + e_k^2},$$

And determined iteratively from the data. In the first iteration, the misfit function

$$E_w^{(1)} = \sum_{k=1}^{N} w_k^{(0)} e_k^{(1)2}$$

is minimized resulting in the linear set of normal equations

$$\boldsymbol{G}^T \boldsymbol{W}^{(0)} \boldsymbol{G} \vec{B}^{(1)} = \boldsymbol{G}^T \boldsymbol{W}^{(0)} \vec{u}^{measured}$$

The minimization of the new misfit function

$$E_w^{(2)} = \sum_{k=1}^{N} w_k^{(1)} e_k^{(2)2}$$

gives $\vec{B}^{(2)}$ which serves again for the calculation of $w_k^{(2)}$. This procedure is repeated giving the typical *j*-th iteration step

$$\boldsymbol{G}^T \boldsymbol{W}^{(j-1)} \boldsymbol{G} \vec{B}^{(j)} = \boldsymbol{G}^T \boldsymbol{W}^{(j-1)} \vec{u}^{measured}$$

The Legendre Polynomial based 2D Iteratively Reweighted Least Square FT method, abbreviated as 2D L-IRLS-FT described above, iteratively recalculates Cauchy weights and results in a very efficient robust inversion method (Amundsen L. 1991). The purpose of robustification was to reduce the influence of data outliers.

3.3.1 Numerical testing in 2D

In order to test the 2D L-LSQ-FT and 2D L-IRLS –FT inversion-based Fourier transform, a noise-free 2D data set in a rectangular test area of the size [−1,1] units in both x and y directions was created (Figure 22). A regular noise-free homogeneous background was generated with a rectangular anomaly (u = 0.7) in the center of size [−0.2, 0.2] units in both directions. The sampling intervals were dx = dy = 0.02 units so the number of data is N = 45*45. The 2D Fourier spectrum of the (noise-free) discrete data set was calculated by means of 2D DFT and 2D L-LSQ-FT algorithms. Figures 23 and 24 show the absolute value (amplitude spectrum) produced by the traditional 2D DFT and the 2D L-LSQ-FT respectively. The 2D L-LSQ-FT spectrum was calculated using Legendre polynomials of the (maximal) order of M=45. For numeric reasons, the calculated Fourier spectra were made on the data set transformed to [-1,1]

DOI: 10.14750/ME.2020.008

in both x and y coordinates. Visual comparison of the two spectra shows approximately the same output, indicating the effectiveness of both methods in processing a noise-free dataset.

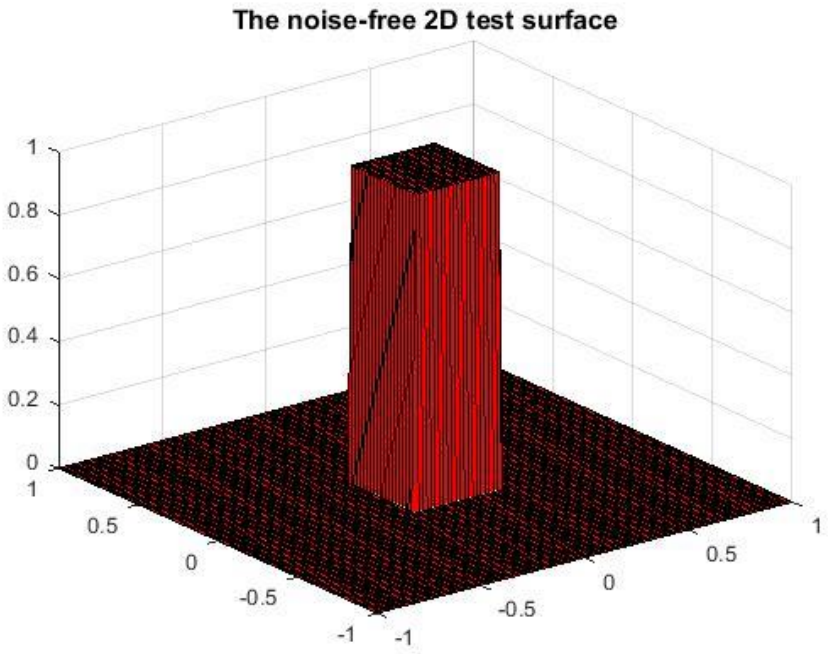

Figure 22; Generated noise-free 2D test surface

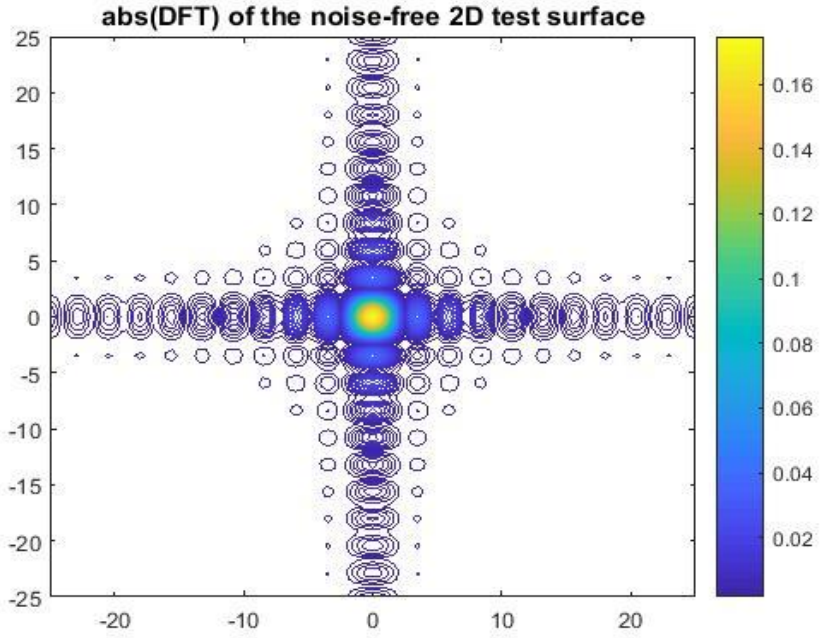

Figure 23; Processed 2D-DFT amplitude spectrum of the noise-free data

DOI: 10.14750/ME.2020.008

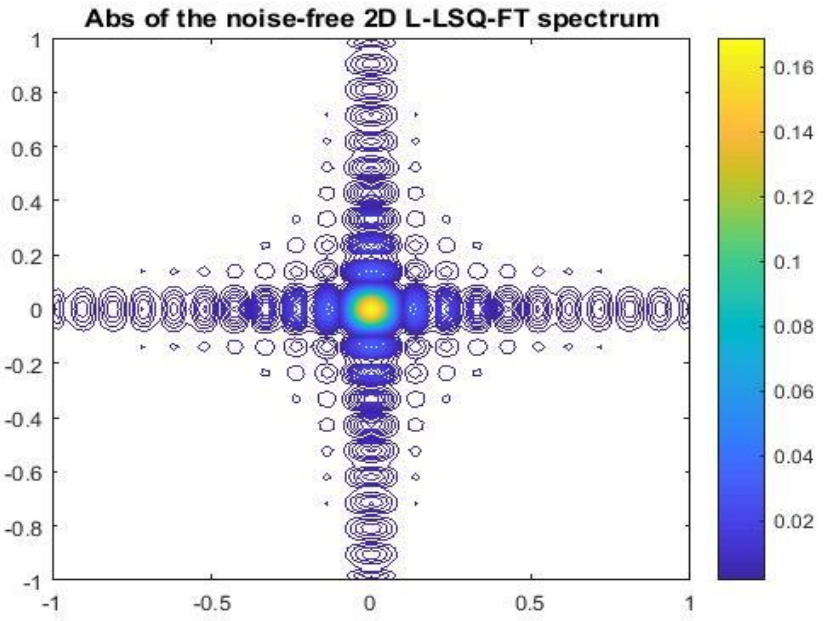

Figure 24; Processed 2D L-LSQ-FT amplitude spectrum of the noise-free data

After successfully applying both methods to the noise-free test surface, random noise following Gaussian- and Cauchy distribution were introduced into the noise-free test surface for processing. The Gaussian noisy data is shown in Figure 25 with a much rougher surface area. Random noise following Cauchy distribution was added to the test surface to produce outlier data with spikes (Figure 26). To demonstrate the noise reduction capability of the two methods, the Gaussian noisy data (Figure 25) was first processed with the traditional 2D DFT and the 2D L-LSQ-FT methods. The resultant Fourier transformed absolute value spectrums are shown in Figures 27 and 28 for the 2D DFT and 2D L-LSQ-FT respectively. We further processed the Cauchy noisy data (Figure 26) with both methods to give the resultant transformed Fourier spectra for 2D DFT and the 2D L-LSQ-FT methods in Figures 29 and 30 respectively. The output spectrums undoubtedly show a considerable suppression of Gaussian and Cauchy noise by the 2D L-LSQ-FT method compared to the traditional 2D DFT method. The 2D DFT could not eliminate the introduced Gaussian and Cauchy noise as demonstrated by the spread of data noise in its output Fourier spectrums (Figure 27 and 29). The processed spectrums from the 2D L-LSQ-FT (Figures 28 and 30) show minor noise across the test area occurring mostly at the edges but comparatively produced much better spectrums with lesser noise. Thus, the 2D L-LSQ-FT algorithm was able to better suppress the Gaussian and Cauchy noise compared to the traditional 2D DFT.

DOI: 10.14750/ME.2020.008

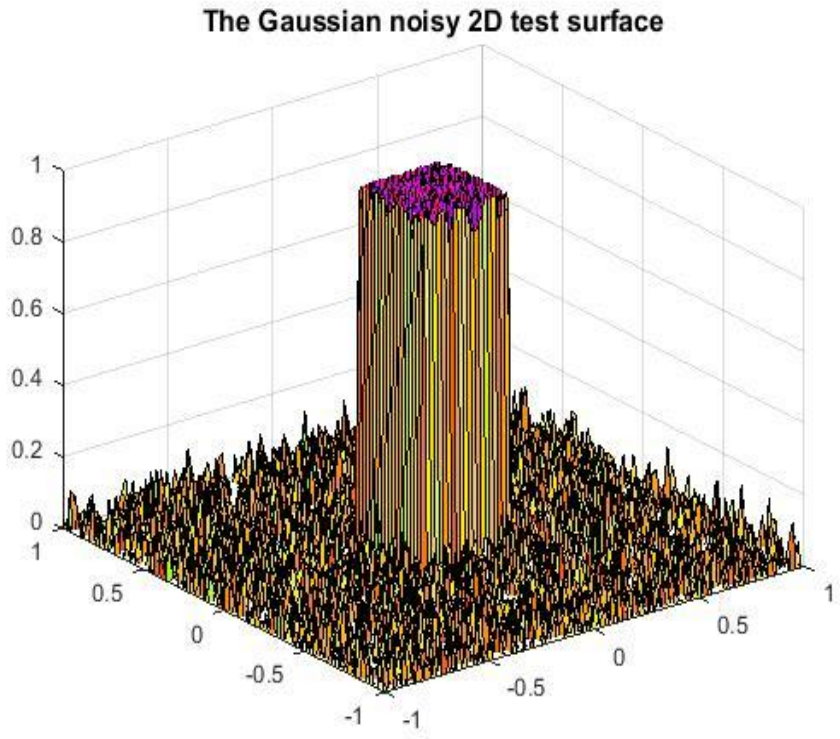

Figure 25; Generated 2D noisy signal with Gaussian noise

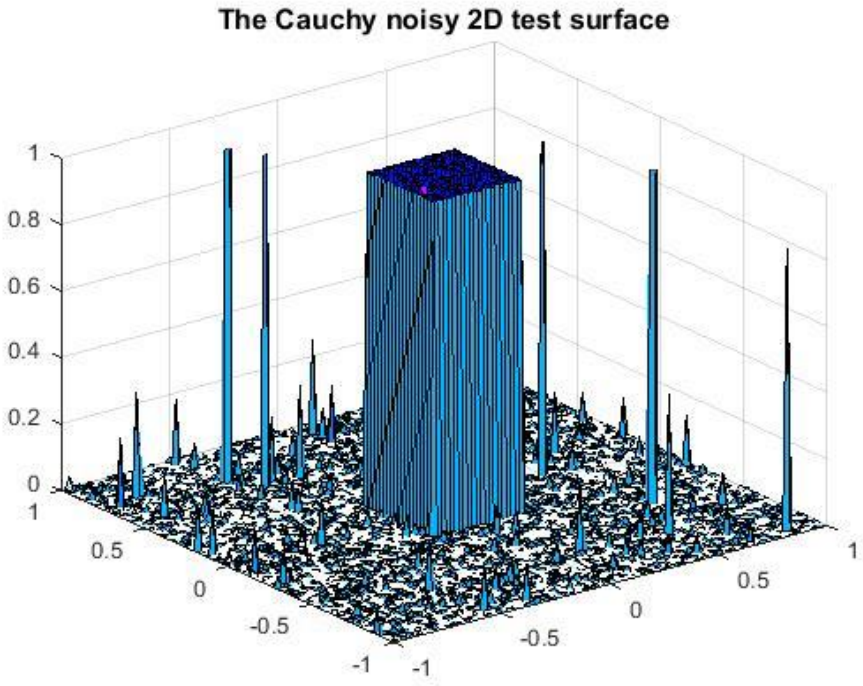

Figure 26; Generated 2D noisy signal with Cauchy noise

DOI: 10.14750/ME.2020.008

For quantitative characterization of the results, we introduce the RMS distance between (a) and (b) data in the space domain

$$d_{RMS} = \sqrt{\frac{1}{N}\sum_{i=1}^{N_x}\sum_{j=1}^{N_y}\left[u^{(a)}(x_i, y_j) - u^{(b)}(x_i, y_j)\right]^2}$$

in the space domain and the model distance

$$D_{RMS} = \left[\frac{1}{N}\sum_{i=1}^{N_x}\sum_{j=1}^{N_y}\left(\mathrm{Re}\left[U^{(a)}(\omega_{xi}, \omega_{yj})\right] - \mathrm{Re}\left[U^{(b)}(\omega_{xi}, \omega_{yj})\right]\right)^2 + \frac{1}{N}\sum_{i=1}^{N_x}\sum_{j=1}^{N_y}\left(\mathrm{Im}\left[U^{(a)}(\omega_{xi}, \omega_{yj})\right] - \mathrm{Im}\left[U^{(b)}(\omega_{xi}, \omega_{yj})\right]\right)^2\right]^{\frac{1}{2}},$$

(N_x, N_y and $N = N_x N_y$ are relevant numbers of data or space-frequency points in the 2D test area).

In the case of the Gaussian noise, the distance between the noisy and noiseless data sets, d = 0.0216. The model or spectra distance between the 2D DFT spectrum (Figure 27) of the noisy (contaminated with Gaussian noise) and the noiseless data sets gave D = $1.703*10^{-3}$. Figure 28 represents sufficient improvement characterized by the spectra distance between the noiseless and the noisy (given by 2D L-LSQ-FT) spectra: D = $1.037*10^{-3}$. Similarly, the 2D DFT gave a spectra distance D=$3.414*10^{-3}$ for spectrum produced from the noisy Cauchy signal whilst the 2D L-LSQ-FT gave a spectra distance: D=$1.336*10^{-3}$. From the above analyses, a higher noise reduction capability was exhibited by the 2D L-LSQ-FT method compared to the traditional 2D DFT method. The results demonstrate the outlier and random noise sensitivity of the 2D DFT and to some extent, the Least Squares Methods, hence the need to define a more robust method for outliers and random noise suppression. We, therefore, introduce the 2D L-IRLS-FT Method.

The same noiseless test surface as shown in Figure 22 above was processed using the 2D DFT and the 2D L-IRLS-FT method. The resultant processed spectrum is shown in Figures 31 and 32 respectively for the 2D DFT and the 2D L-IRLS-FT methods. The 2D L-IRLS-FT spectrum was calculated using Legendre polynomials of the (maximal) order of M=45 as in the case of the 2D LSQ-FT. For numeric reasons, the calculated Fourier spectra were made on the data set transformed to [-1,1] in both x and y coordinates (as in the case of the 2D L-LSQ-FT) resulting in an appropriate scale in the wavenumber domain.

DOI: 10.14750/ME.2020.008

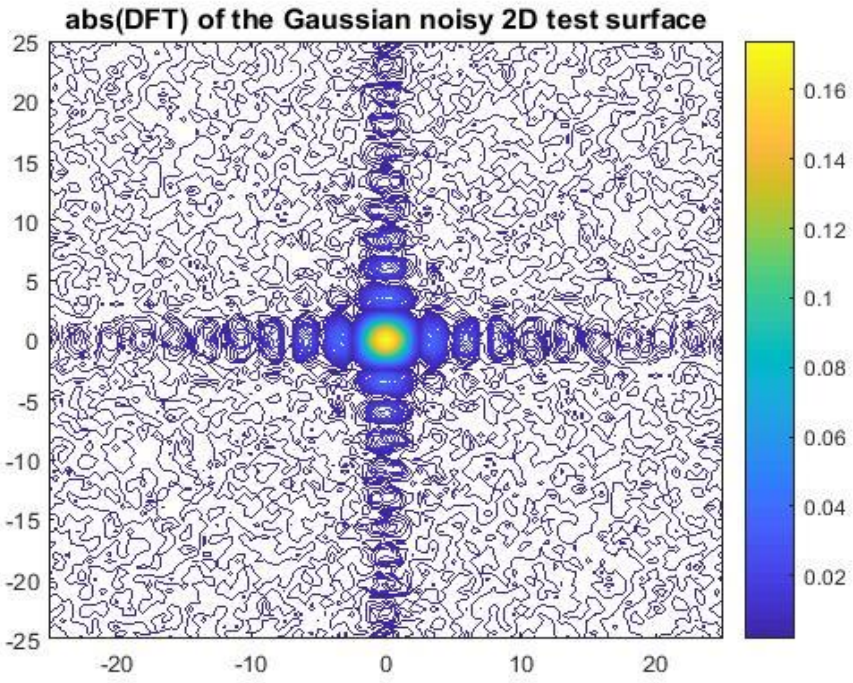

Figure 27; Processed 2D DFT spectrum of the Gaussian noisy signal (D=1.703*10^{-3})

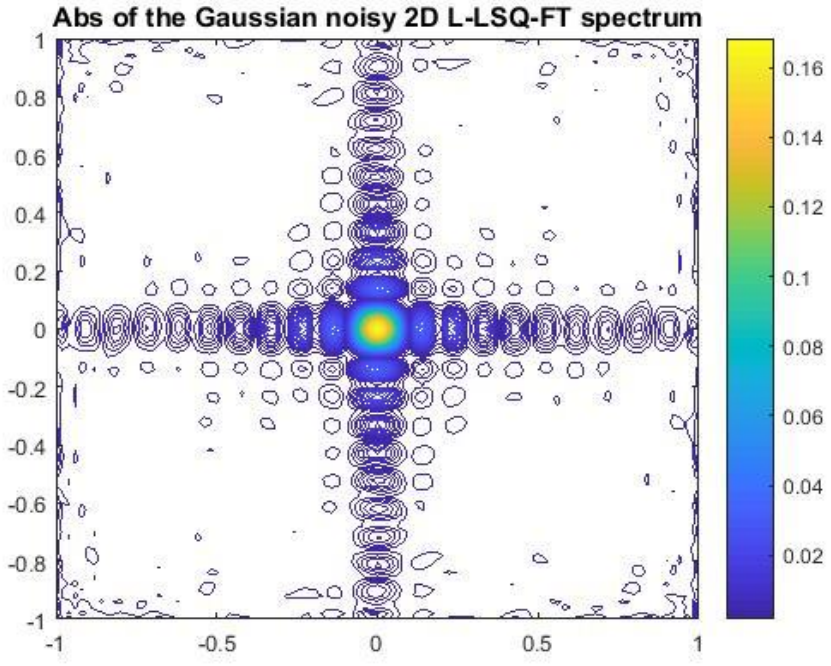

Figure 28; Processed 2D L-LSQ-FT spectrum of the Gaussian noisy signal (D=1.037*10^{-3})

DOI: 10.14750/ME.2020.008

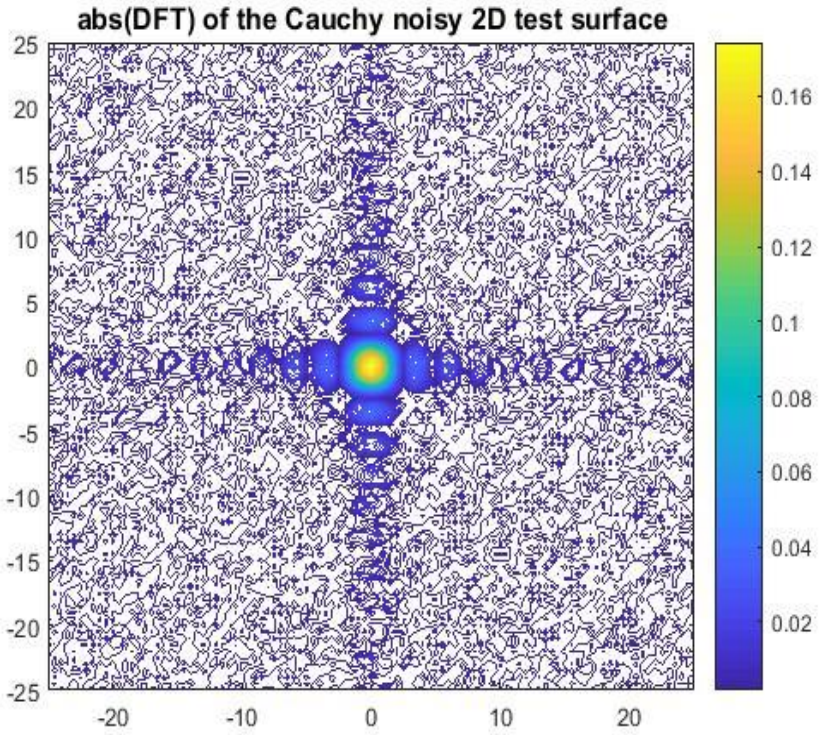

Figure 29; Processed 2D DFT spectrum of the Cauchy noisy signal (D=3.414*10^{-3})

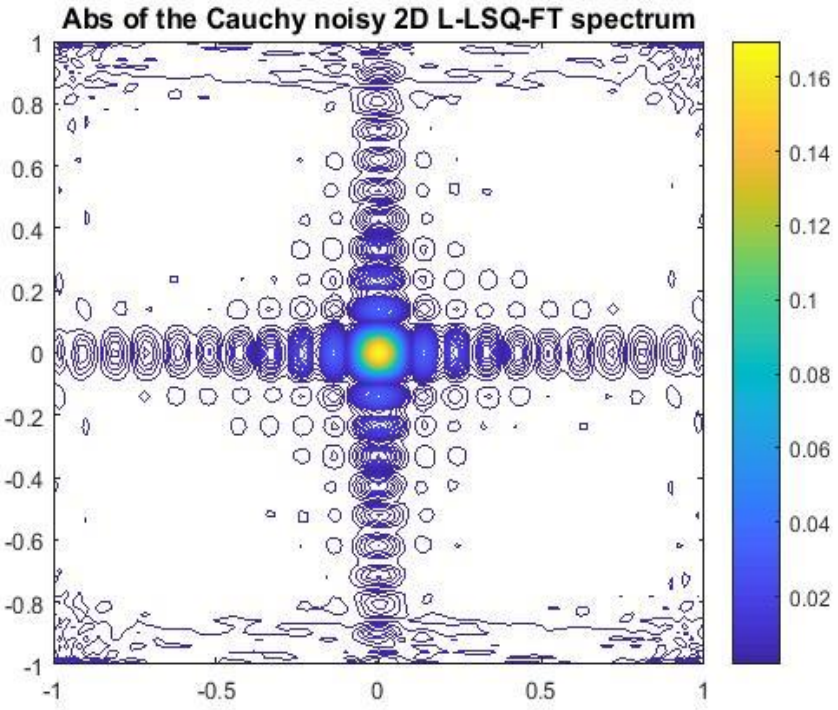

Figure 30; Processed 2D L-LSQ-FT spectrum of the Cauchy noisy signal (D=1.336*10^{-3})

A comparison of the spectrum from the 2D L-IRLS-FT processed noise-free signal (Figure 31) to the output signals from the traditional 2D DFT (Figures 32) shows a very good similarity, indicating that the 2D L-IRLS-FT algorithm was efficient in processing the noise-free signal. To test the noise reduction capability of the 2D DFT and the 2D L-IRLS-FT, the Gaussian and Cauchy noisy data sets (Figures 25 and 26 above) was this time processed with the 2D DFT and the 2D L-IRLS-FT algorithms. For the Gaussian noisy data, the output processed Fourier spectrum for the 2D DFT and 2D L-IRLS-FT are shown in Figures 33 and 34 respectively. Also, the output processed Fourier spectrum for the 2D DFT and the 2D L-IRLS-FT for the Cauchy noisy data are shown by Figures 35 and 36 below.

From the above output spectrums, the newly developed 2D L-IRLS-FT algorithm was more effective in reducing the Gaussian and Cauchy noise component of the noisy data sets compared to the traditional 2D DFT. In the case of Gaussian noise, the 2D DFT spectrum (Figure 33) was noisier compared to the 2D L-IRLS-FT spectrum (Figure 34), which gave an almost noise-free spectrum. The same was exhibited in the case of the Cauchy noisy data as the 2D L-IRLS-FT demonstrated higher noise reduction capability from its processed Fourier spectrum (Figure 36) compared to that of the 2D DFT (Figure 35). This highlighted the limitation of the traditional 2D DFT in rejecting randomly occurring outliers. To analytically characterize the results, we applied the RMS distance between two data sets in the space domain as well as the model or spectra distance. For the processed Gaussian noisy dataset, the model or spectra distance between the DFT spectrum (Figure 33) of the noisy (contaminated with Gaussian noise) and the noiseless DFT processed data sets is D =$1.700*10^{-3}$. Figure 34 represents sufficient improvement characterized by the spectra distance between the noiseless and the noisy (given by 2D L-IRLS-FT) spectra: D =$1.231*10^{-3}$. Likewise, the 2D DFT gave a spectra distance D=$3.013*10^{-3}$ for spectrum produced from the noisy Cauchy data whilst the 2D L-IRLS-FT gave a spectra distance of D=$6.759*10^{-4}$.

From the above analyses, the 2D L-IRLS-FT method compared to the traditional 2D DFT method, showed a much higher noise reduction capability when both Gaussian and Cauchy noise were added to the test surface for processing. The results wholly validate the outlier and random noise sensitivity of the traditional 2D DFT method. Hence, we recommend the new method, the 2D L-IRLS-FT which is tough, robust and resistant enough to subdue randomly occurring data noise.

DOI: 10.14750/ME.2020.008

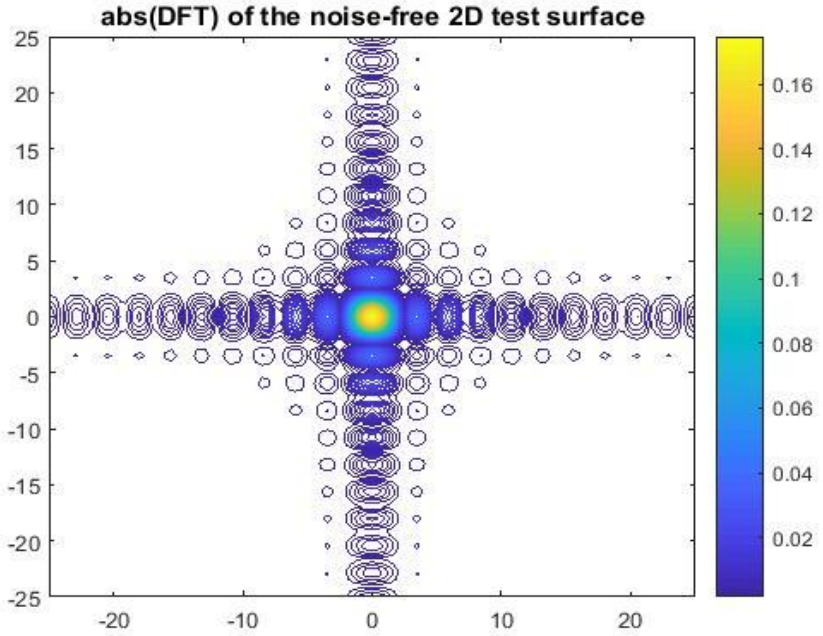

Figure 31; Processed 2D DFT spectrum of the noise-free test surface

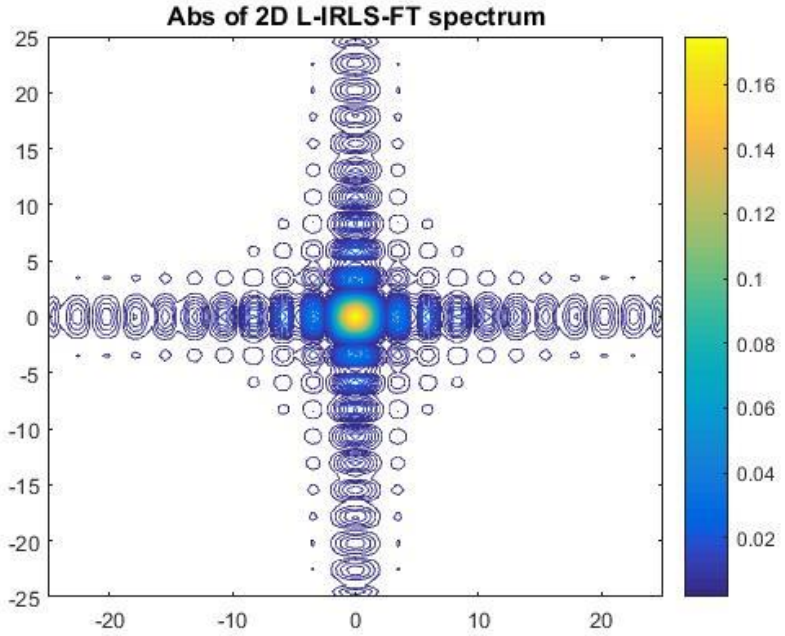

Figure 32; Processed 2D L-IRLS-FT spectrum of the noise-free test surface

DOI: 10.14750/ME.2020.008

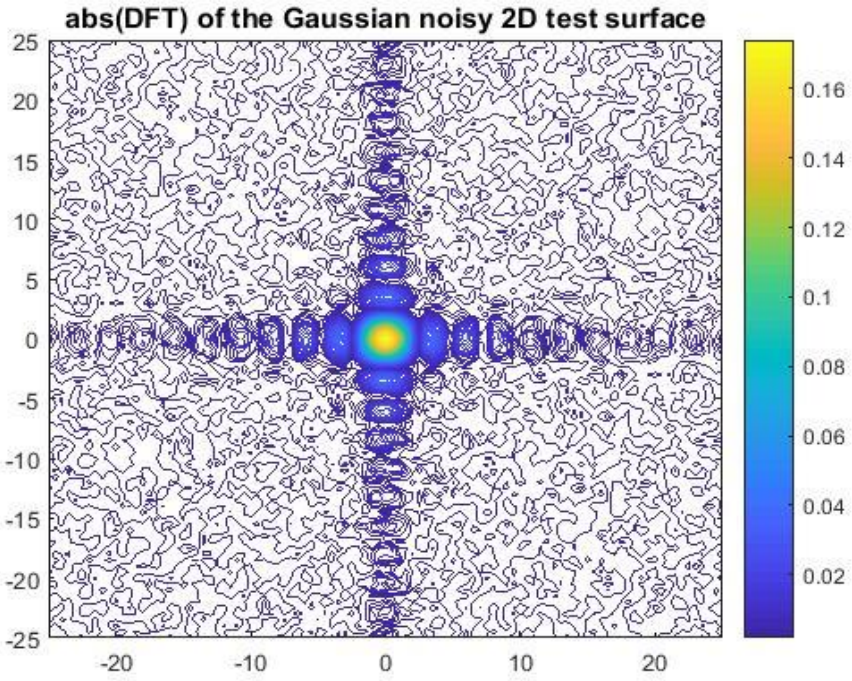

Figure 33; Processed 2D DFT spectrum of the Gaussian noisy Data (D=1.700*10^{-3})

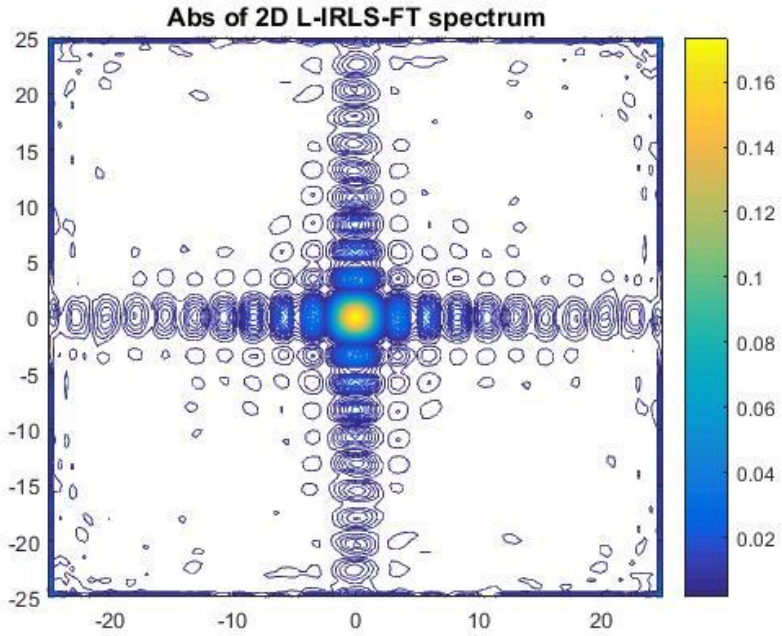

Figure 34; processed 2D L-IRLS-FT spectrum of the Gaussian noisy Data (D=1.231*10^{-3})

DOI: 10.14750/ME.2020.008

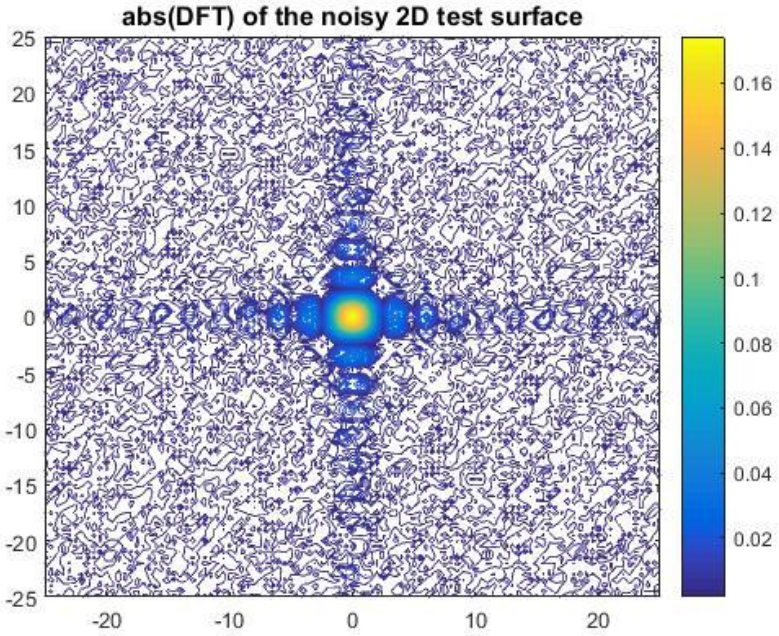

Figure 35; Processed 2D DFT spectrum of the Cauchy noisy Data (D=3.413*10^{-3})

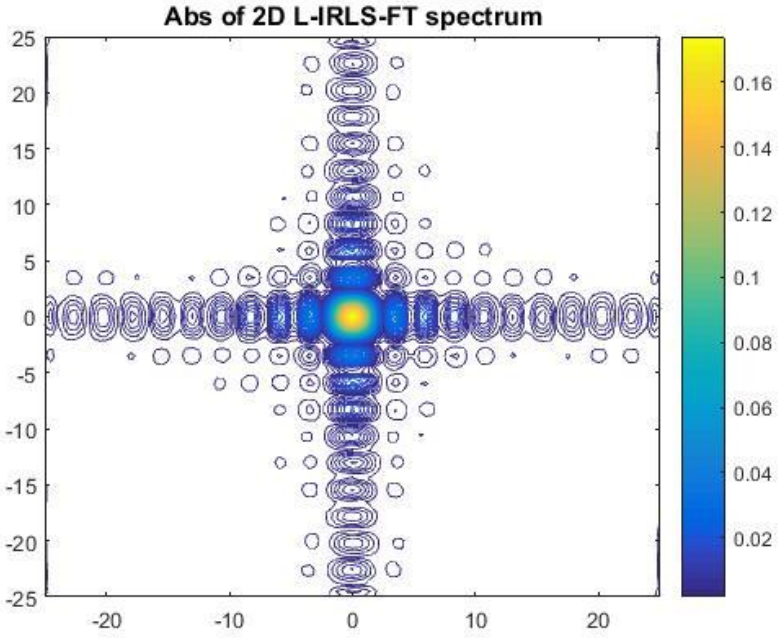

Figure 36; Processed 2D L-IRLS-FT spectrum of the Cauchy noisy Data (D=6.759*10^{-4})

DOI: 10.14750/ME.2020.008

Based on the successful application of the 2D L-LSQ-FT and the 2D L-IRLS-FT, it is required to relate the results to the original 2D H-LSQ-FT and the 2D H-IRLS-FT which forms the basis of the 2D inverse Fourier transformation method development. To do that, we first processed the same noise-free test surface (Figure 22) with the 2D H-LSQ-FT and the 2D H-IRLS-FT. The processed spectrums are given in Figures 37 and 38 respectively for the 2D H-LSQ-FT and the 2D H-IRLS-FT respectively. A comparison to the 2D DFT processed noise-free spectrum (Figure 23 for the LSQ and figure 31 for the IRLS) shows that the 2D H-LSQ-FT and 2D H-IRLS-FT were efficient in processing the noise-free data. We further processed the Gaussian and Cauchy noisy data sets (Figures 25 and 26) with the 2D H-LSQ-FT and the 2D H-IRLS-FT to give the resultant spectrums for processed Gaussian and Cauchy noisy signal in Figures 39 and 40 respectively for the 2D H-LSQ-FT. Also, the output processed Fourier spectrum for 2D H-IRLS-FT for the Gaussian and Cauchy noisy data are shown in Figures 41 and 42 below.

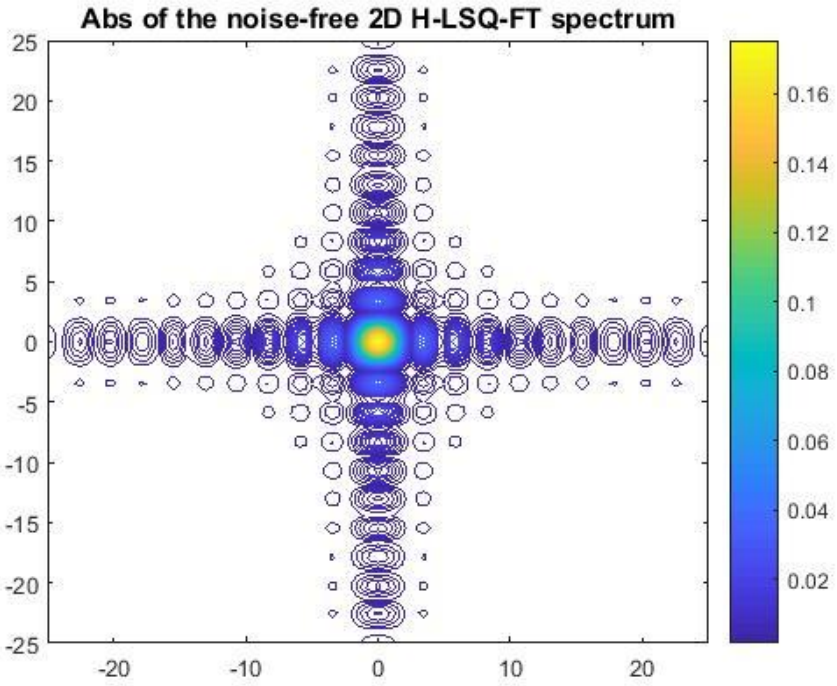

Figure 37; Processed 2D H-LSQ-FT spectrum of the noise-free Morlet waveform

DOI: 10.14750/ME.2020.008

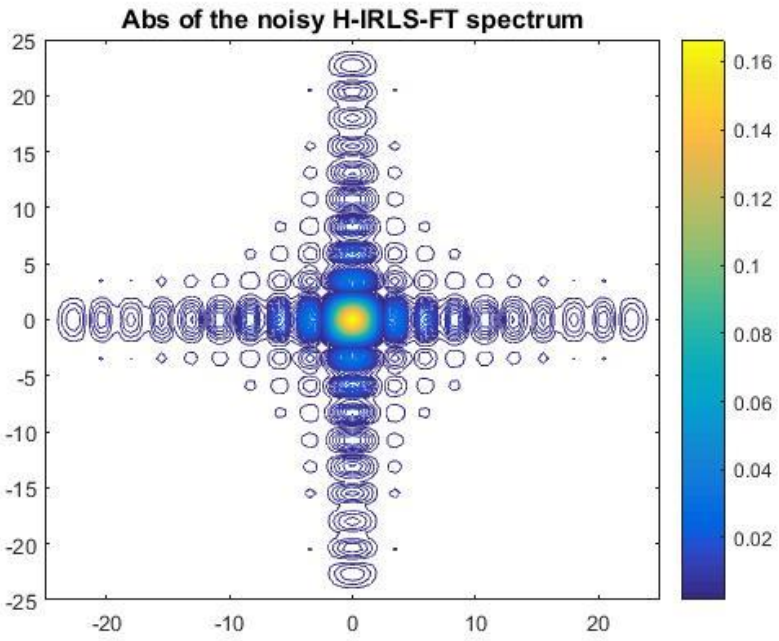

Figure 38; Processed 2D H-IRLS-FT spectrum of the noise-free test surface

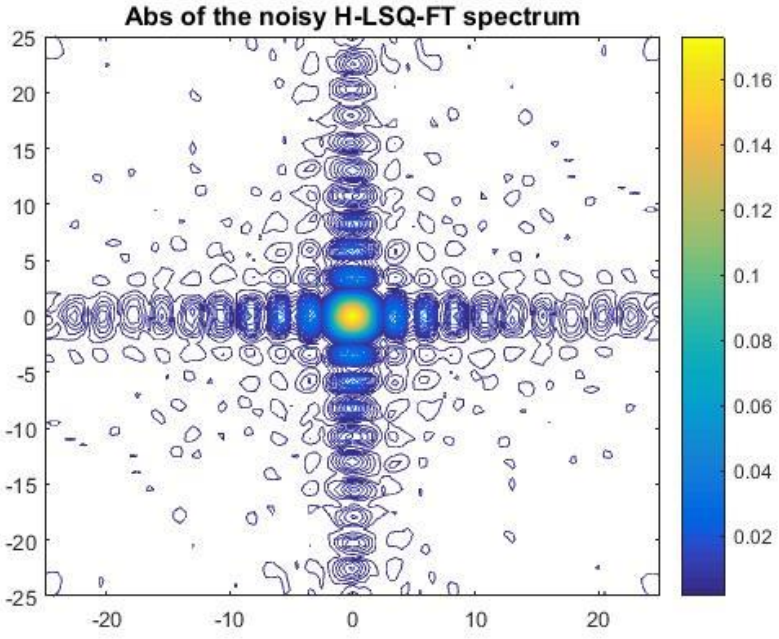

Figure 39; Processed 2D H-LSQ-FT spectrum of the Gaussian noisy data (D=8.6322*10^{-4})

DOI: 10.14750/ME.2020.008

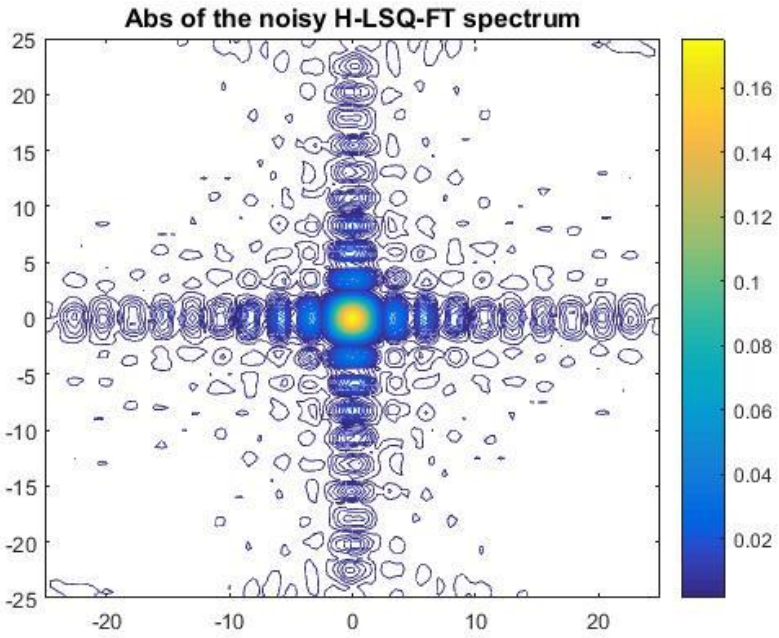

Figure 40; Processed 2D H-LSQ-FT spectrum of the Cauchy noisy data (D=9.147*10^{-4})

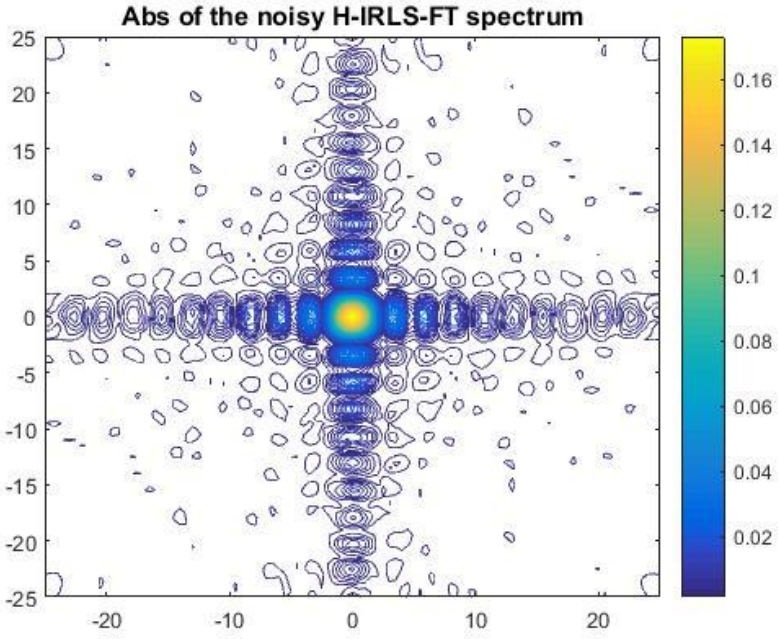

Figure 41; Processed 2D H-IRLS-FT spectrum of the Gaussian noisy Data (D=8.6563*10^{-4})

DOI: 10.14750/ME.2020.008

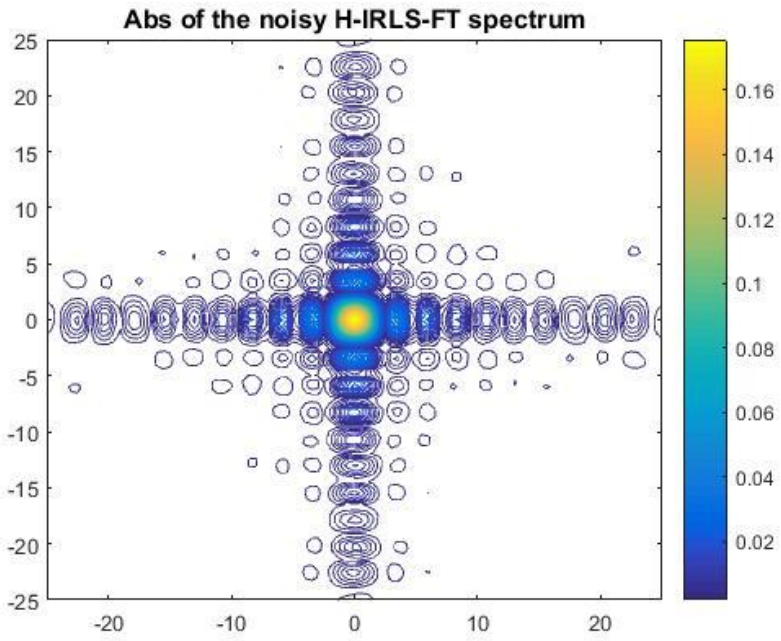

Figure 42; Processed 2D H-IRLS-FT spectrum of the Cauchy noisy Data (D=6.8676*10^{-4})

To quantitatively differentiate and relate the results of the newly developed two dimensional Legendre polynomial based Fourier transformation and the accepted two dimensional Hermite-based Fourier transformation, we applied the RMS distance between two data sets in the space domain as well as the model or spectra distance. Table 3 shows the calculated spectra distances after various noise was introduced into the noise-free test data and processed. It was practical that the 2D Hermite based inverse Fourier transformation methods were more efficient in its Least Square and iteratively reweighted forms than the 2D Legendre polynomial based Fourier transformation. As explained in the 1D example, this is so because the applied Hermite functions are the Eigenfunctions of the inverse Fourier transform while the Legendre polynomials are not. Comparatively, the newly developed 2D L-LSQ-FT and 2D L-IRLS-FT can be considered as a better alternative to the traditional 2D DFT as demonstrated above but an additional backing to the original 2D H-LSQ-FT and the 2D H-IRLS-FT.

In conclusion, a new robust and resistant, inversion based Fourier transformation (L-LSQ-FT and L-IRLS-FT) is presented in 1D and 2D where the spectrum is discretized by series expansion using Legendre polynomials as basis functions. First, the series expansion coefficients are given by the solution of a linear inverse problem (L-LSQ-FT) and the elements of the Jacobian matrix calculated faster and easily, without integration. The procedure is robustified using Iteratively Reweighted Least Squares (IRLS) method with Steiner weights

DOI: 10.14750/ME.2020.008

resulting in a very efficient robust and resistant inversion procedure (L-IRLS-FT). Its applicability is demonstrated in 1D and 2D through the numerical testing of synthetic data. It was observed in the test that the Discrete Fourier Transformation (DFT) is a common data processing method but incorporate some level of noise in the transformation process. Comparatively, the newly introduced Legendre polynomial inversion-based Fourier transformation algorithm has a higher noise rejection capability than the traditional DFT Method and can be used as an alternative to the original Hermite inversion-based Fourier transformation.

Table 3. Calculated Spectra Distance for 2D L-LSQ-FT, 2D H-LSQ-FT, 2D L-IRLS-FT, and the 2D H-IRLS-FT Methods

	CALCULATED SPECTRA DISTANCE	
APPLIED METHOD	**GAUSSIAN NOISE**	**CAUCHY NOISE**
2D L-LSQ-FT	0.001037	0.001336
2D H-LSQ-FT	0.000863	0.000914
2D L-IRLS-FT	0.001231	0.000676
2D H-IRLS-FT	0.000865	0.000687

DOI: 10.14750/ME.2020.008

Chapter 4

DEVELOPMENTS OF THE H-LSQ-FT AND H-IRLS-FT BY OPTIMIZING THE SCALE PARAMETERS

4. 1 Method Development In 1D

As discussed in chapter 2, the basic theory of the H-LSQ-FT method development involves some distinct steps which are the formulation of Fourier transformation as an over-determined inverse problem, discretization of the Fourier spectrum using series expansion and the calculation of the Jacobi matrix using Hermite functions as the Eigenfunctions of the Fourier transform. The applied Hermite functions are orthonormal and square-integrable between the interval -∞ to ∞, hence, the need to be modified by scaling to cover specific frequencies. This necessitated the introduction of a scaling parameters 'α' into equation (44) above. Inappropriately, the value of the scale parameter is inserted into the algorithm from practical experience (trial and error basis), which is problematic, making it difficult to assume. There is, therefore, a real need to determine 'α' in an exact algorithm. In this chapter, we introduce a meta-algorithm that improves the H-LSQ-FT procedure by giving the optimal values of the scale parameters. The calculated Jacobian matrix for the 1D inverse Fourier transformation is given by Eq. (37)

$$G_{k,i} = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \Psi_i(\omega) e^{j\omega t_k} d\omega ,$$

where $G_{k,i}$ is an element of the Jacobian matrix of the size N-by-M. The Jacobian matrix is the inverse Fourier transform of the Ψ_i basis function. Parameterization of the model is achieved by introducing a special feature of the Hermite functions, thus, by making them the eigenfunctions of the inverse Fourier transform as

$$\mathrm{F}^{-1}\left\{H_n^{(0)}(\omega)\right\} = (j)^n H_n^{(0)}(t) ,$$

Where $H_n^{(0)}$ is the Hermite function which is modified by scaling. By introducing 'α' as a scale parameter, it can be shown that the scaled Hermite functions are given by

$$H_n(\omega_x, \alpha) = \frac{e^{-\frac{\alpha \omega_x^2}{2}} h_n(\omega_x, \alpha)}{\sqrt{\sqrt{\frac{\pi}{\alpha}}\, n!(2\alpha)^n}} \quad .$$

Scaling was important because, in geophysical applications, the frequency covers wider ranges depending upon the problem discussed.

41.2 A Meta-Algorithm To Optimize The Scale Parameter

To develop a meta-algorithm to optimize the scale parameter, we use the Simulated Annealing Method which is a Global Optimization Technique. It is a method for solving unconstrained and bound-constrained optimization problems. Simulated annealing is an optimization process that is an analogy to the process of physical annealing with solids, in which a crystalline solid is heated and then allowed to cool very slowly until it achieves its most regular possible crystal lattice configuration. The method creates a relationship between this type of thermo-dynamic behavior and the search for global minima for a discrete optimization problem. It further provides an algorithmic means for exploiting such a connection. For each iteration, the objective function generates values for two solutions for comparison. Improved results are accepted, while some non-improving results are also accepted in the hope of escaping the local optima in search of global optima. Thus, the algorithm accepts all new solutions that lower the objective function, but also with a certain probability, solutions that raise the objective function. By accepting solutions that raise the objective function, the algorithm avoids being trapped in local minima and is able to explore globally for more possible solutions. An annealing schedule is selected to systematically decrease the temperature as the algorithm proceeds. As the temperature decreases, the algorithm reduces the extent of its search to converge to a minimum.

To optimize the alpha parameter and determine it from the real statistics of the data, we first looked at the relationship between alpha and the data frequency. From the exponential sequence in the Hermite function,

$$e^{\frac{-\alpha\omega^2}{2}} \Longrightarrow \alpha = \frac{1}{\omega_0^2} = \frac{1}{(2\pi f_0)^2} \quad \text{giving} \quad e^{-\frac{1}{2}\left(\frac{\omega}{\omega_0}\right)^2}$$

Since, $'\alpha'$ changes by smaller values, the values of alpha were optimized through the frequency $'f_0'$. We defined three sets of objective or energy functions to be minimized independently which include the data distance given by the difference between the measured and calculated data from the H-LSQ-FT (inverse FT), the model or spectrum distance with the DFT spectrum as reference spectrum and a combination of the data and model distances. The minimized energy functions are given below:

$$Ene = \sum_{k=1}^{N} (d^{measured} - d^{calculated})^2 = Min \tag{72}$$

DOI: 10.14750/ME.2020.008

$$Ene = \sum_{k=1}^{N}(|U_n^{cal}(\omega) - U_n^{DFT}(\omega)|)^2 = Min \quad (73)$$

$$Ene = \sum_{k=1}^{N}(d^{measured} - d^{calculated})^2 + \sum_{k=1}^{N}(|U_n^{cal}(\omega) - U_n^{DFT}(\omega)|)^2 = Min \quad (74)$$

We define a total number of 150 SA iterations with 30 random internal loop search. The search interval for $'f_0'$ was $2 \le f_0 \le 6$. In each SA iteration, the inversion-based Fourier transform is called (with a given α regulated by the SA algorithm) to estimate the energy function which is either accepted or rejected based on the change of the energy function. Below is the flow chart demonstrating the operational sequence of the alpha optimizing meta-algorithm.

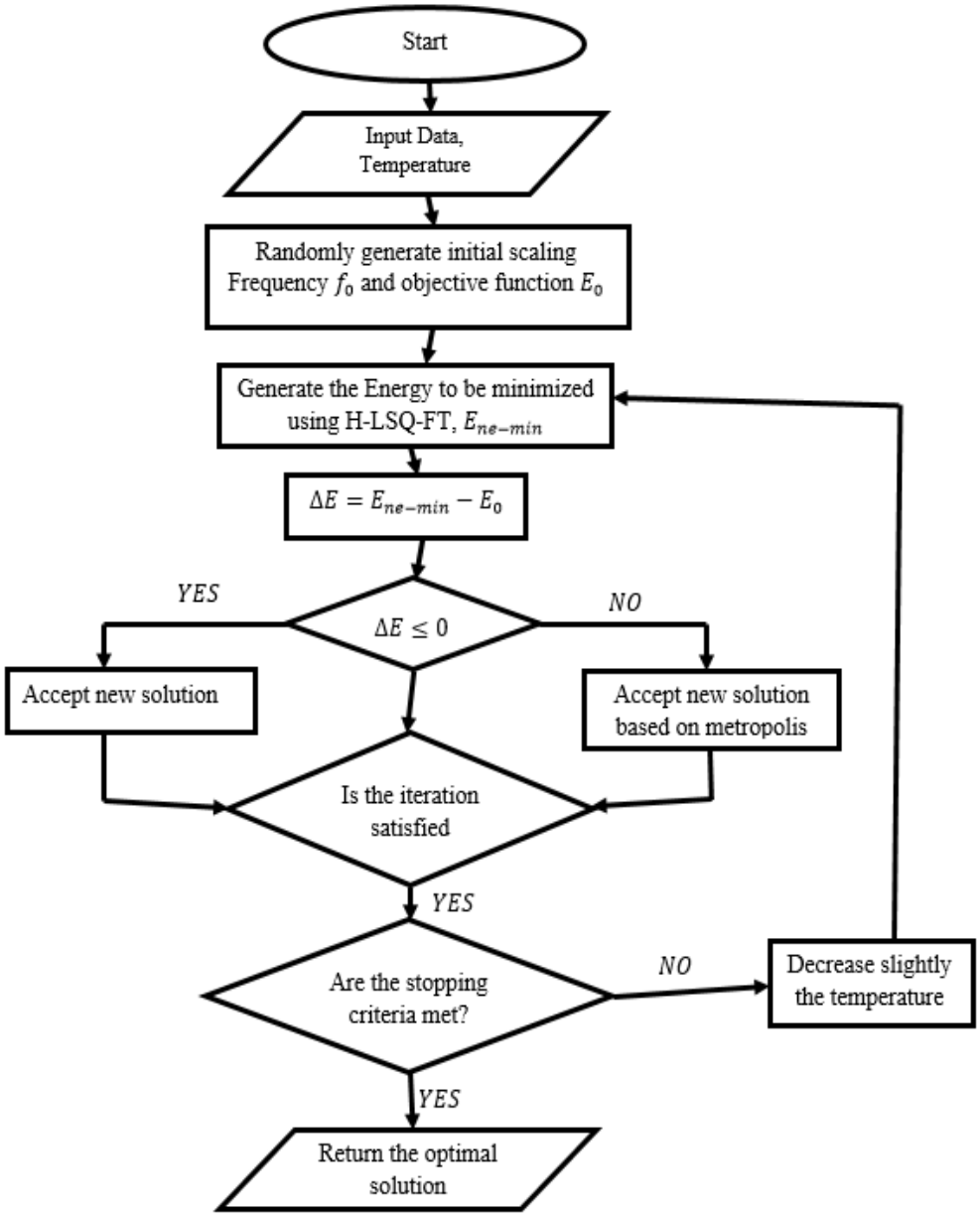

Fig 43, Flowchart of the Alpha Optimization Simulated Annealing Meta Algorithm

4.1.3 Numerical testing

The time-domain signal (Figure 2) was used again to test the optimization capability of the newly developed meta-algorithm by applying the H-LSQ-FT and the traditional DFT in one dimension. The noiseless time function of the test data as described earlier by the formula below

$$u(t) = \kappa t^{\eta} e^{-\lambda t} \sin(\omega t + \varphi)$$

where specified fixed values for the signal parameters are as follows: $\kappa \approx 738.91$, $\eta = 2$, $\lambda = 20$, $\omega = 40\pi$, $\varphi = \pi/4$. The noise-free waveform was sampled at regular intervals of 0.005 (sec) measurement points ranging over the time interval of [-1, 1] to produce 401 data points.

The algorithm estimates the inverse Fourier transform of the input signal using the H-LSQ-FT methods. The H-LSQ-FT was calculated with a series expansion coefficient of m=150. An objective function was then created in the form of an energy function to be optimized. The SA program randomly generates a scaling frequency $'f_0'$ and energy $'Ene\_min'$ to be minimized in each step of the iteration. The inversion based Fourier transform is called into the algorithm in each iteration to re-calculate the energy function to be minimized and the cycle continues. If the estimated energy is less than zero, the algorithm accepts the new solution but otherwise, the algorithm accepts the results based on the Metropolis acceptance condition. The input temperature decreases correspondingly for each iteration until the stop criteria are met and the algorithm returns the optimal solution.

The program first runs with noise-free data and an introduction of data noise (following Gaussian and Cauchy distributions). For the noise-free data, we optimize the scaling frequency using the data distance (eq. 72) to give Figure 44. The energy function reached a minimum of $8.1737*10^{-6}$ after 60 iterations to give a scaled frequency of 3.9618 Hz. A similar optimization procedure was applied to the spectrum or model distance (defined in eq. 73) to give Figure 45. The results show the energy function arriving at a minimum of $8.158*10^{-5}$ after 70 iterations to give an optimized frequency of 3.9592 Hz. Figure 46 shows the resultant frequency, energy reduction trajectory and the number of iteration when we optimized the scaling frequency from the combined data and model distances (given in eq.74). The input energy reached its minimum after 60 iterations to give a frequency of 3.9549 Hz. From a quantitative comparison of the output scaled frequencies, thus, 3.9618 Hz, 3.9592 Hz, and 3.9549 Hz, it can be concluded that the algorithm has the ability to derive similar alpha values ($\alpha = \frac{1}{(2\pi f_0)^2}$) from the three different energy functions generated by the method.

DOI: 10.14750/ME.2020.008

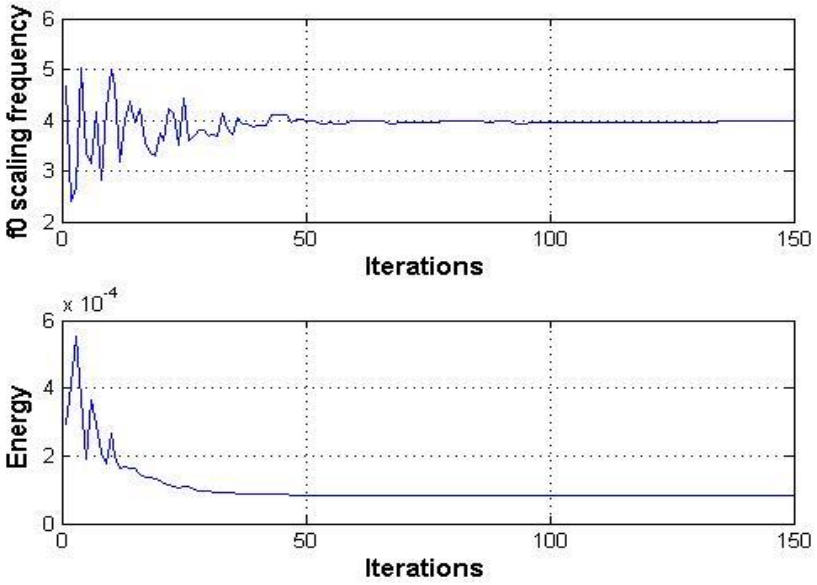

Fig 44; Noise-free optimization of alpha from the data distance

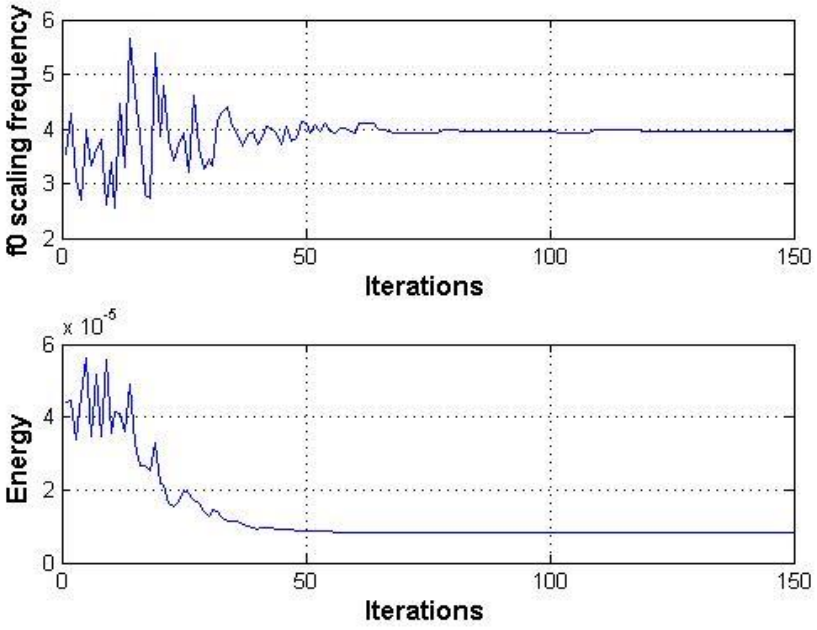

Fig 45; Noise-free optimization of alpha from the spectrum distance

DOI: 10.14750/ME.2020.008

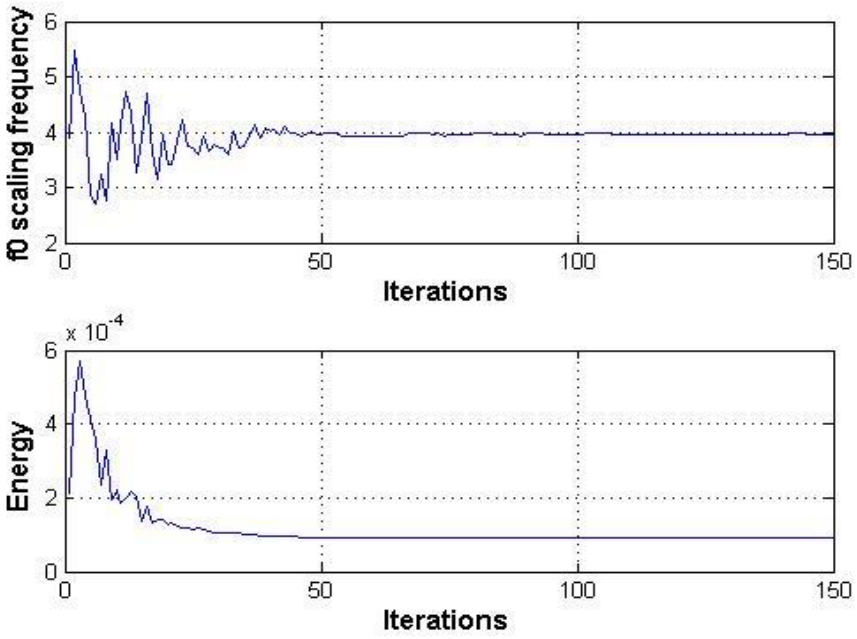

Fig 46; Noise-free optimization of alpha from the combined data and spectrum distance

To further demonstrate the efficacy of the method, we introduced Gaussian noise into the noise-free data to obtain Gaussian noisy data (Figure 5 above). We then optimized the scaling frequency in the data (eq. 72) and model domains (eq. 73) as well as in their combination (eq. 74) to give Figures 47, 48 and 49, respectively. By computing the input energy function from the Gaussian noisy data using data distance, the algorithm estimated a scaled frequency of 2.9548 Hz after 100 iterations (Figure 47) while the input energy reached its minimum after 60 iterations. Similarly, a scaling frequency of 2.9514 Hz was obtained when optimization was effected in the model or spectrum domain (Figure 48). Figure 49 shows the output results when the scaling frequency was optimized from the combined data and model distances with Gaussian noise. The input energy reached its minimum or a constant value of $8.45*10^{-2}$ after 60 iterations to optimize a scaled frequency of 2.9319 Hz after 80 iterations. Comparably, the optimized frequencies are almost the same in each domain with a calculated mean value of 2.94 Hz.

DOI: 10.14750/ME.2020.008

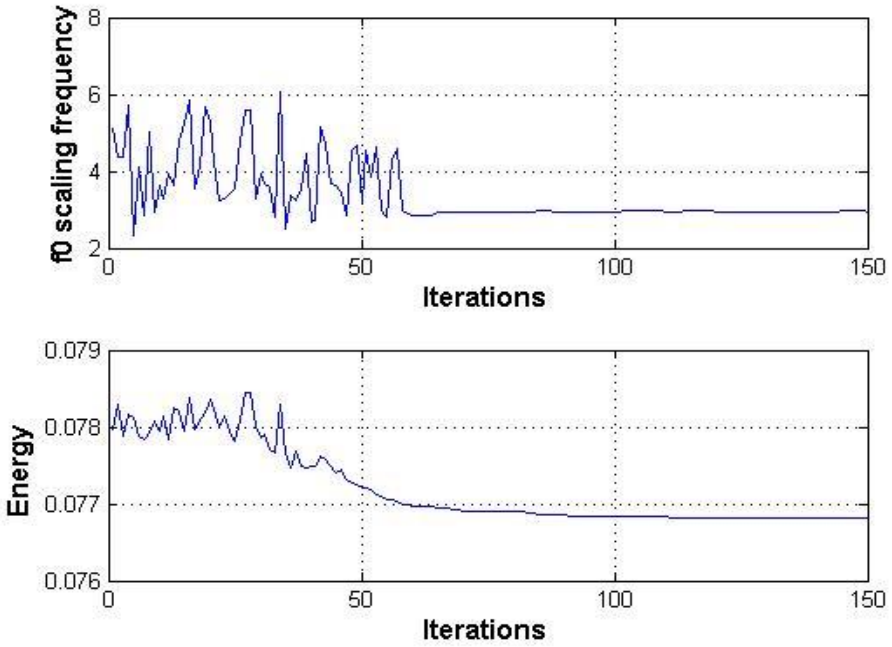

Fig 47; Optimization of alpha from data distance with Gaussian Noisy

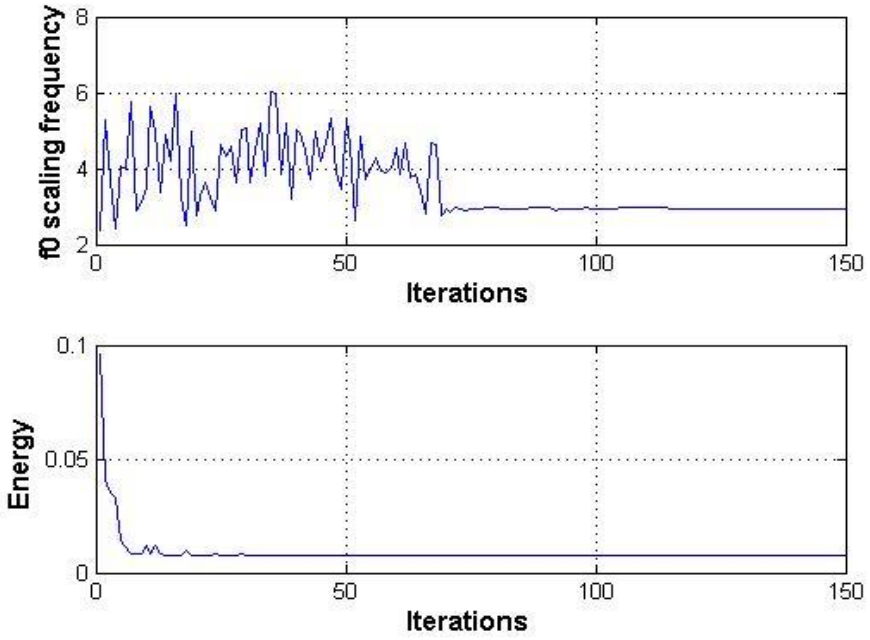

Fig 48; Optimization of alpha from spectrum distance with Gaussian Noisy

DOI: 10.14750/ME.2020.008

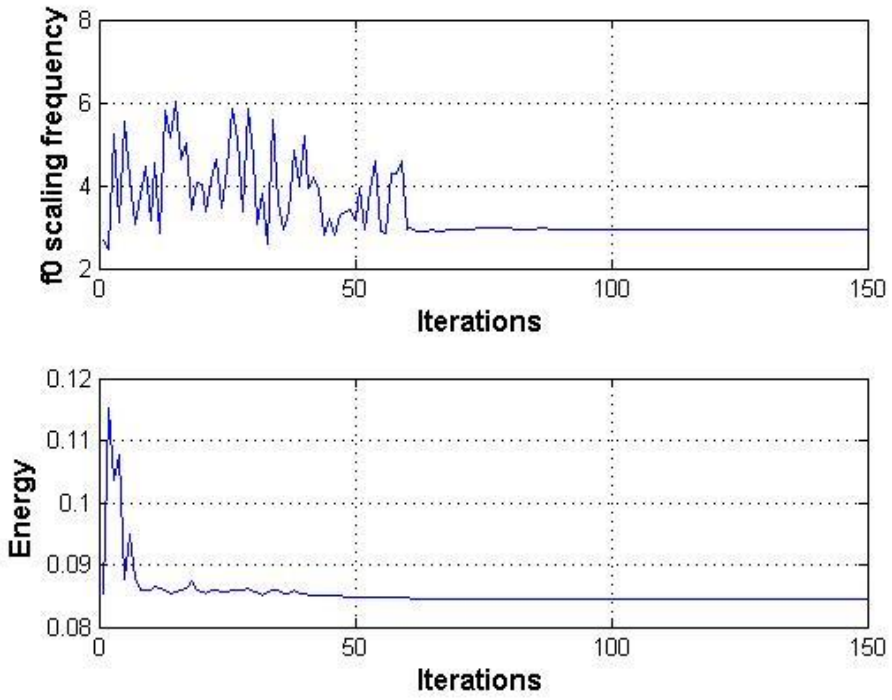

Fig 49; Optimization of alpha from combined data and spectrum distances with Gaussian noisy data

Since in practice Cauchy noise is normally encountered in acquired field data instead of Gaussian noise, we introduced Cauchy noise into the noise-free data to obtain Cauchy noisy data (Figure 6 above) for optimization. Output images showing frequency and energy trajectories for the optimized alpha values in the data and model domains are given in Figures 50, 51 and 52 respectively. By using data distance, the algorithm estimated a scaled frequency of 2.4132 Hz after 80 iterations (Figure 50) while the input energy reached its minimum after 75 iterations. Correspondingly, a scaling frequency of 2.7404 Hz was also attained when optimization was realized in the model or spectrum domain (Figure 51). Figure 52 shows the output results when the scaling frequency was optimized from the combined data and model distances with Cauchy noise. The input energy reached its minimum or a constant value of $4.127*10^{-1}$ after 80 iterations to optimize a scaling frequency of 2.7538 Hz after 125 iterations. Comparably, the variations in the optimized scaling frequencies are insignificant with an estimated mean value of 2.64HZ. These tests show that, in practice, the alpha parameter can be adequately optimized using only the data distance in the face of any type of noise to give a correct estimate. Table 4 shows the final optimized frequencies and their calculated mean values in the different domains after the simulated annealing procedure.

DOI: 10.14750/ME.2020.008

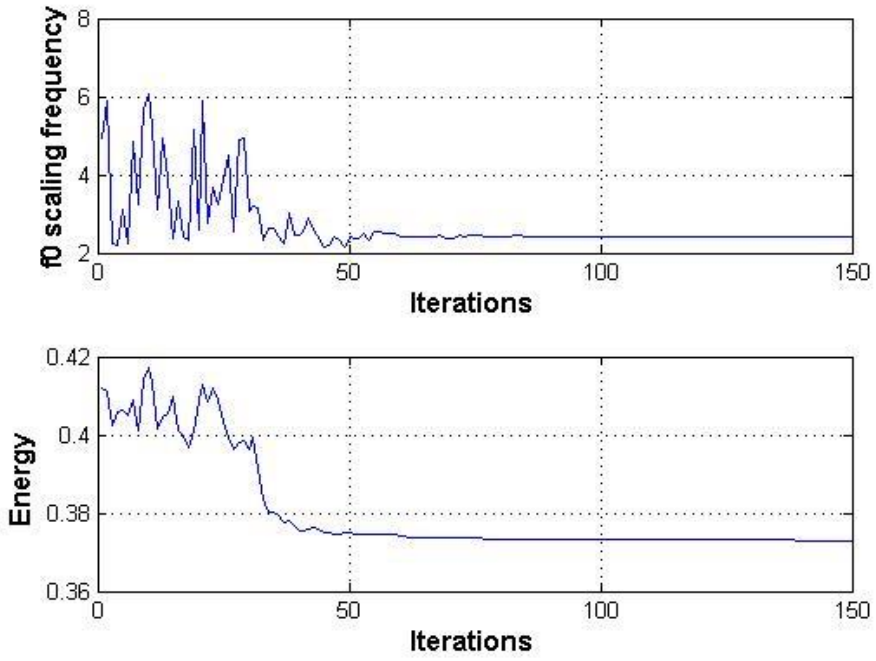

Fig 50; Optimization of alpha from data distance with Cauchy Noisy

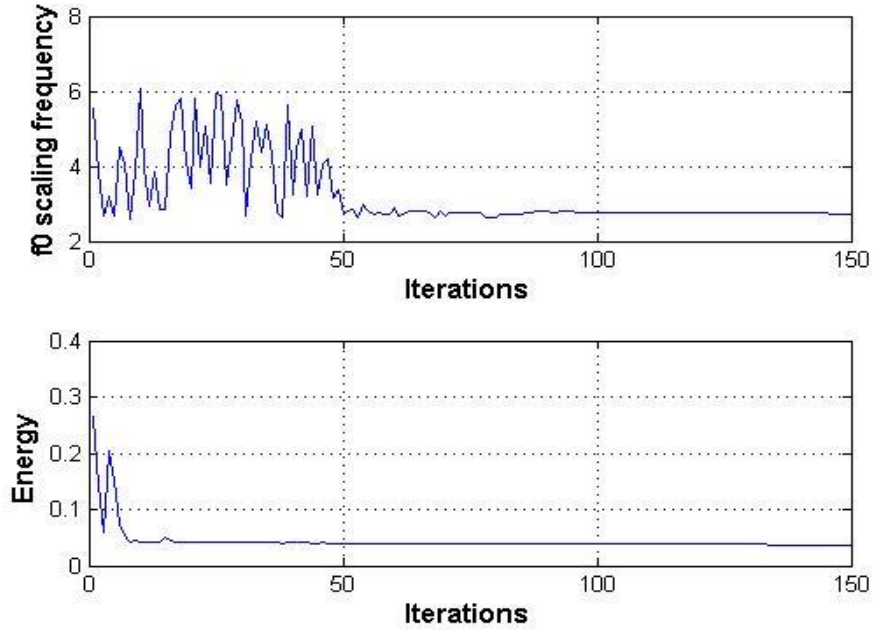

Fig 51; Optimization of alpha from spectrum distance with Cauchy Noisy

In conclusion, we have developed an iterative procedure for optimizing the scaled parameter alpha, in the H-LSQ-FT Method. This procedure uses the simulated annealing technique to optimize the alpha parameter from the real statistics of the input data or output spectrum or a combination of both, hence, eliminating the human error component associated

DOI: 10.14750/ME.2020.008

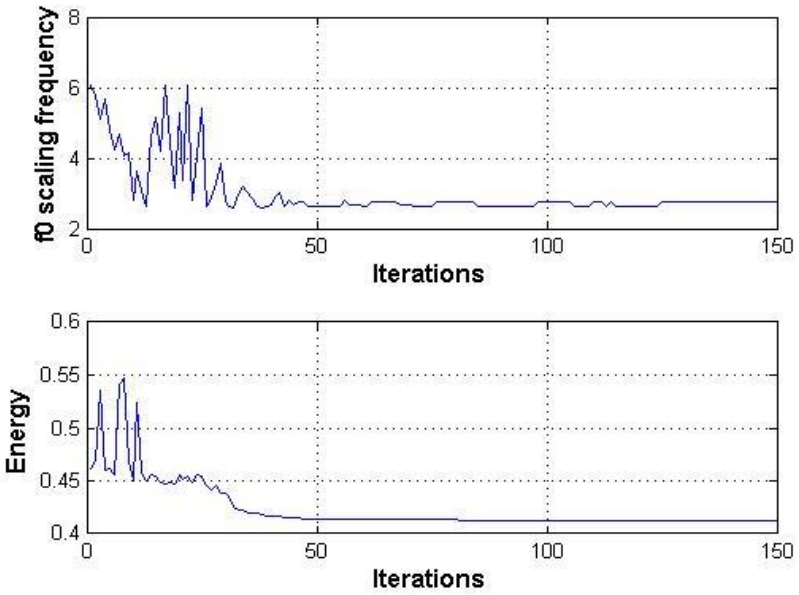

Fig 52; Optimization of alpha from combined data and spectrum distances with Cauchy Noisy

Table 4, Results from Alpha Optimization Test Using Simulated Annealing

	Characteristic Frequency f_0			
	Data Distance Minimized	Spectrum Distance Minimized	Data And spectral Distance Minimized	Mean Value
Noiseless	3.9618	3.9592	3.9549	3.96
Gaussian Noisy	2.9545	2.9514	2.9319	2.94
Cauchy noisy	2.4132	2.7404	2.7538	2.64

with defining the alpha parameter in the H-LSQ-FT Method. The results demonstrated that it is enough to use data distance in case of noisy data to optimize the alpha parameter as the algorithm gave similar output in the data space, spectrum space or a combination of both. This note is important because, in practice, optimization in the noise-free case is impossible since acquired geophysical data always has a noise component. Also, it is reasonless in practice to use comparison in spectral-domain with DFT, as it is highly noise-sensitive (this feature motivated the development of the inversion based FT method). With the above alpha optimizing algorithm, it is certain that the H-LSQ-FT Method will exhibit a significant improvement in its noise reduction capabilities.

DOI: 10.14750/ME.2020.008

Chapter 5

THE CONCEPT OF RANDOM-WALK SAMPLING

5.1 Preliminary Investigations

A random walk survey is a known stochastic or random process, that describes a path that consists of a succession of random steps along a survey line. With the advancement in survey tools which incorporates the global positioning system, the concept of random walk survey enable geophysicist to sample randomly in the field knowing the exact geographic coordinates of the acquired data. Conducting a geophysical survey is a multi-step process including survey planning and design, fieldwork, data processing and analysis, and preparing and presenting findings. The aim of a rigorous geophysical field measurement design is to obtain a representative survey data which can be interpreted to reveal subsurface bodies. The representative survey should be less expensive and less time consuming to conduct and if thoughtfully designed and carefully conducted, can yield results that accurately and reliably reflect subsurface characteristics relevant to advancing interpretation. Data acquisition in geophysical field surveys typically requires a grid of measurement stations (Parasnis, 1986). The data point density does not vary throughout the field and may not take into account whether anomalies are present. Therefore, the idea arose to develop survey systems to sample field data adapted to specific problems and accumulate the data for storage in the field. The traditional surveys enable samples to be taken on pre-defined equally spaced lines or grids. The concept of random-walk involves a survey pattern where data are collected on even in random walk paths along pre-determined grids. In the case of magnetic surveys, a well-organized geomagnetic survey system is presented, which avoids separate geodetic measurements by using a Global Positioning System (GPS) in combination with a magnetic system. The data is sent to the survey magnetic system for onward storage and monitoring which offers the possibility of adaptive data sampling in the exploration area of interest.

Random-Walk survey design is a procedure for selecting a sample location in which all positions along the survey grid are given an equal probability of being selected or sampled as one takes a definite step along the survey line. Even in random-walk field surveys guarantees unbiasedness at the level of geometric sampling, but does not guarantee adequate precision. Other ways to implement random surveys include simple random measurements where data is taken randomly across the field of survey and systematic random sampling where measurements are taken at a definite interval at a non-regular step along a pre-determined grid.

Systematic uniform random surveys often offer an appropriate balance between the estimator precision and the time spent to obtain samples. It is easy to execute without mistakes along a survey line, and in many situations, is more precise than simple random measurements for a given survey area. Sampling on the corners of a regular square grid (lattice design) or equivalently at fixed intervals of random-walk long equally spaced parallel transects (Mulla and Bhatti, 1997) is the most commonly used Systematic uniform random surveys design used in geophysical field surveys.

Technological advances in geospatial data have the potential to change how survey data are collected. Being affected by high costs, limited capacity, and difficulties in supervision, geophysical field data acquisition are often designed using non-probability approaches. As geospatial technology has improved and become more widespread, costs have decreased substantially and the number of available survey tools upgraded, making even in random-walk field-based sampling approaches accessible to more geophysicists. The practical possibility of the random-walk technique associated with its ease of implementation is seen from the continual development of the method, dissemination of associated techniques and algorithms in many areas of science and engineering (including biology, statistical physics, chemical engineering, astronomy, and geosciences). In the area of geoscience, an application of random-walk measurements can be found in the field of hydrology, oil and gas exploration and subsurface repository of various wastes. Random-walk techniques have been employed as a direct simulation technique for passive tracer or pollutant transport in aquifers (Kinzelbach, 1988; Kinzelbach and Unk, 1991; Zimmermann et al, 2001; Hoteit et al, 2002b; Delay et al, 2005). At the laboratory scale, the technique has also been applied in post-processing intended at inferring effective transport properties in porous media such as permeability, porosity, and electrical conductivity. For example, electrical conductivity was first computed using a random walk by Kim and Torquato (1992) and Sahimi (2011). Similarly, at the same pore-to-sample scale, random walk techniques were also used to interpret Nuclear Magnetic Resonance (NMR) data (Neel et al, 2011, 2014; Guillon et al, 2013, 2014; Fleury et al, 2015). Various random-walk data processing algorithms have been developed for anomaly detection using original non-gridded data points. The possibility of sampling at random intervals with successful processing tools will eliminate the apparent misrepresentations sampling out of the grid creates in the data processing. To achieve this, we tested the processing strength of the newly developed robust 2D H-IRLS-FT method and the traditional 2D DFT on synthetic magnetic anomaly generated on a regular and random walk geometries.

DOI: 10.14750/ME.2020.008

5.2 Application in Reduction to Pole (RTP)

Magnetic data processing, also known as data reduction enable geophysicist to eliminate spikes and noise unrelated to the underlying geology from field measurements. Among the magnetic data reduction techniques applied in data processing include magnetic compensation, data checking and editing, diurnal variations removal, geomagnetic reference field removal, tie line leveling, micro leveling and reduction to pole (RTP). Each of these processing techniques performs specific functions to improve the quality of the acquired data. For instance, data checking and editing are performed on survey lines to help remove cultural anomalies from the data. Cultural anomalies are a serious problem in the geologic interpretation of magnetic data, especially modern surveys that typically fly low above cultural sources. Diurnal correction removes temporal variations of the earth's magnetic field and is achieved by subtracting the time-synchronized signal recorded at a stationary base magnetometer from the measured data. In addition, Tie line leveling utilizes supplementary data recorded on the lines to further adjust the data and is based on the fact that data recorded at the intersection of traverse and tie lines should be equal after data reduction, while Micro leveling eliminates subtle errors caused by variation in terrain clearance or elevation. After the above corrections, the data is gridded to obtain the total magnetic intensity (TMI) map containing various anomalies which are then reduced to the pole. The RTP filter transformed the anomalies such that they are centered over their subsurface causative bodies which compensate for latitudinal variations.

Baranov (1957), Baranov and Naudy (1964) proposed the mathematical approach known as a reduction to the pole to simplify magnetic anomaly shape during interpretation. Reduction to the pole converts the magnetic field from magnetic latitude where the Earth's field is inclined, to the field at a magnetic pole, where the inducing field is vertical (Luo et al., 2010). It is a data processing technique used to recalculate total magnetic intensity data as if the inducing magnetic field had a 90^0 inclination. This filter transforms dipole magnetic anomalies to monopole anomalies centered over their causative bodies which can simplify the interpretation of data. A magnetic anomaly depends on the inclination and declination of the body's magnetization, the inclination and declination of the local earth's magnetic field, and the orientation of the body with respect to the magnetic north (Nabighian et al. 2005). When the Earth's field is inclined, magnetic anomalies due to induction have forms that are asymmetrically related to their sources, but when the inducing field is vertical, the induced anomalies are directly over their sources (Milligan and Gunn, 1997). Reduction to pole requires knowledge of the direction of magnetization, often assumed to be parallel to the ambient field.

DOI: 10.14750/ME.2020.008

The RTP concept makes the simplifying assumption that the rocks in the survey area are all magnetized parallel to the earth's magnetic field and is true in the case of rocks with an induced magnetization only, however remnant magnetization will not be correctly dealt with if the direction of remanence is different to the direction of the earth's magnetic field. Reduction to the pole was achieved by applying the formula in the frequency domain given as

$$R(u,v) = T(u,v)S(u,v), \tag{75}$$

where *T(u,v)* is the 2D Fourier transform of the magnetic data set, *S(u,v)* is the frequency domain operator of pole reduction and *R(u,v)* is the reduced data set after the data reduction process. In practice, RTP operation becomes unstable at lower magnetic latitudes because of a singularity that appears when the azimuth of the body and the magnetic inclination both approach zero (Nabighian et al. 2005). Numerous approaches have been proposed by various researchers to overcome this problem. Leu (1982) suggested reducing anomalies measured at low magnetic latitudes to the equator rather than the pole. Pearson and Skinner (1982) proposed a whitening approach that strongly reduced the peak amplitude of the RTP filter, thus reducing noise. Mendonca and Silva (1993) used a truncated series approximation of the RTP operator. Gunn (1972, 1995) designed Wiener filters in the space domain by determining filter coefficients that transform a known input model at the survey location to a desired output at the pole. Keating and Zerbo (1996) also used Wiener filtering by introducing a deterministic noise model, allowing the method to be fully automated.

5.2.1 Numerical test in 1D using Morlet signal

The Morlet waveform (Figure 2) was used to test the applicability of the H-IRLS-FT method on random measurements in one dimension. As a first step, the Morlet waveform was sampled equidistantly in 401 points ranging over the time interval of [-1, 1] and processed using the traditional DFT method to give both the real and imaginary parts of Fourier Transform. The same equidistantly sampled waveform was also processed using the H-IRLS-FT method. The result is shown in Figure 55. Both the traditional DFT and the H-IRLS-FT showed similarities in the output processed real and imaginary parts of the Fourier spectrum.

DOI: 10.14750/ME.2020.008

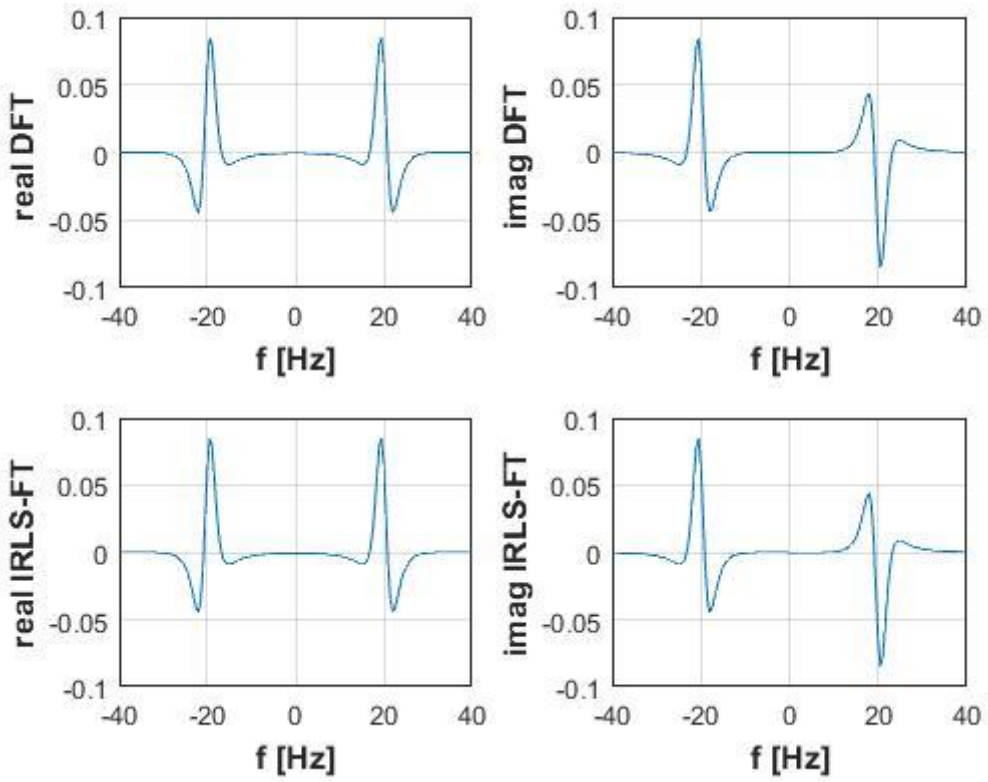

Figure 53; The spectrum of the equidistantly sampled Morlet Waveform processed with DFT and also with H-IRLS-FT method

To test the efficiency of the H-IRLS-FT method in random-walk measurements, the Morlet waveform was sampled randomly for processing. In this experiment, the same number of samples were used with randomly selected positions in the whole time interval. The randomness of the sampling is demonstrated in Figure 54 whilst Figure 55 shows the H-IRLS-FT spectrum of the signal. It can be seen, that both its real and imaginary parts are exactly the same as that, found in regularly sampled cases (Figure 53). This proves that the inversion based Fourier Transform method gives the same results in processing both regularly and non-regularly sampled data set.

DOI: 10.14750/ME.2020.008

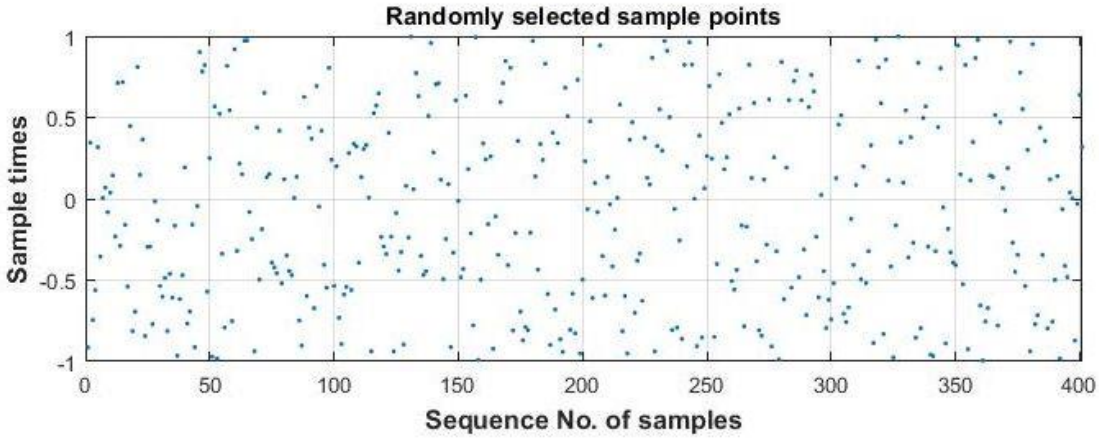

Figure 54. The randomly selected sample points used for sampling the Morlet Waveform

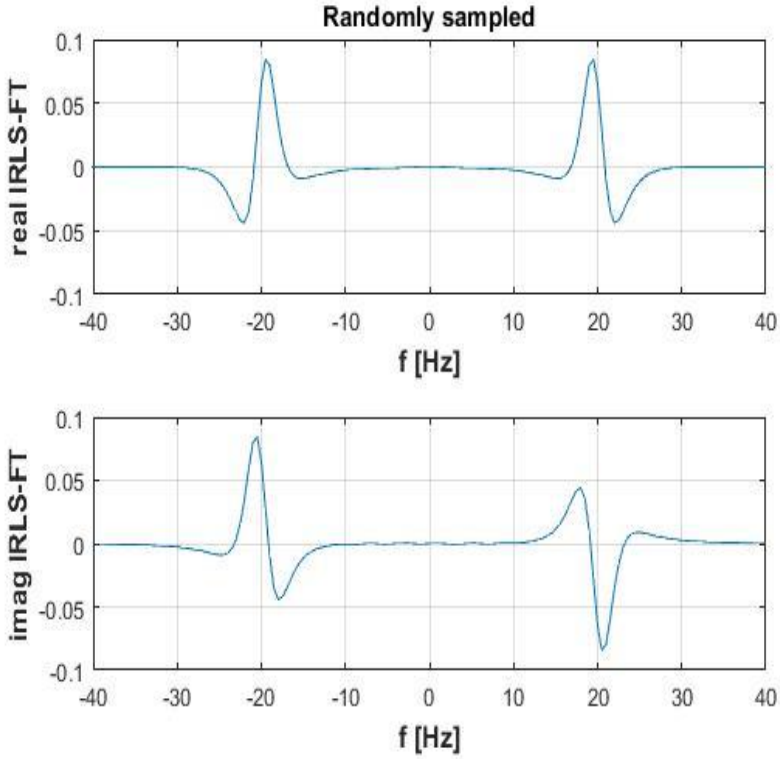

Figure 55; The H-IRLS-FT spectrum of the randomly sampled Morlet Waveform

5.3.1 A magnetic dipole example with equidistant sampling

For practical purposes, an upgraded program code was written to further test and evaluate the effectiveness of the 2D DFT and the 2D H-LSQ-FT methods for processing equidistant and non-equidistant (random-walk) geometry dataset. A grid-based sampling area of 6000m in both X and Y directions was created with a regular sampling interval of 400m in both directions as shown in Figure 56. An analytic magnetic dipole source was generated at a depth of 1000m below an initial surface elevation of zero degrees (sea level). The Total Magnetic Intensity (T)

DOI: 10.14750/ME.2020.008

of the dipole source used was based on Karoly I. Kiss (2009), Magnetic Methods of Applied Geophysics. To calculate the Total Magnetic Intensity of the dipole source, fixed source parameters included magnetic moment, inclination and declination of the earth which were given as $10^9 A \cdot m^2$, 60°, and 0° respectively. Remnant magnetization was not taken into consideration. The calculated total magnetic field in 2D contoured and 3D surface maps are shown in Figures 57 and 58 respectively. The mapped anomaly clearly shows magnetic variations in highs (red) and lows (blues) demonstrating the dipole nature of the magnetic field.

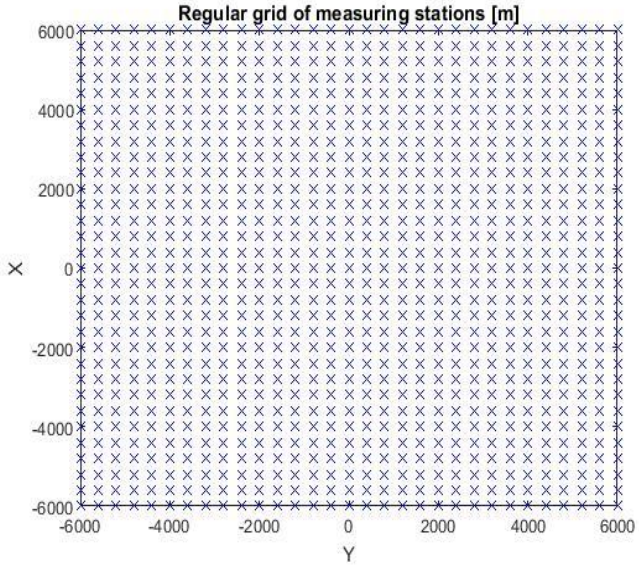

Figure 56; Regular grid of measuring stations

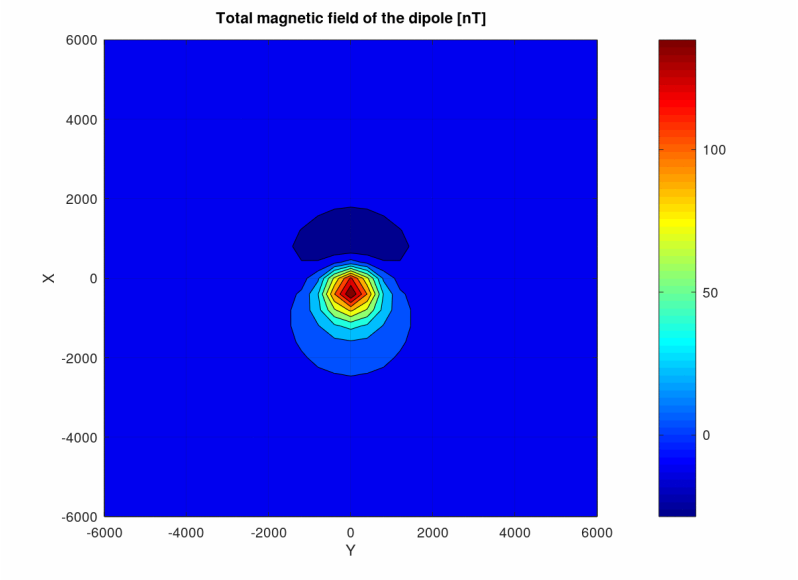

Figure 57; Total Magnetic Field of the Dipole from Equidistant grid

DOI: 10.14750/ME.2020.008

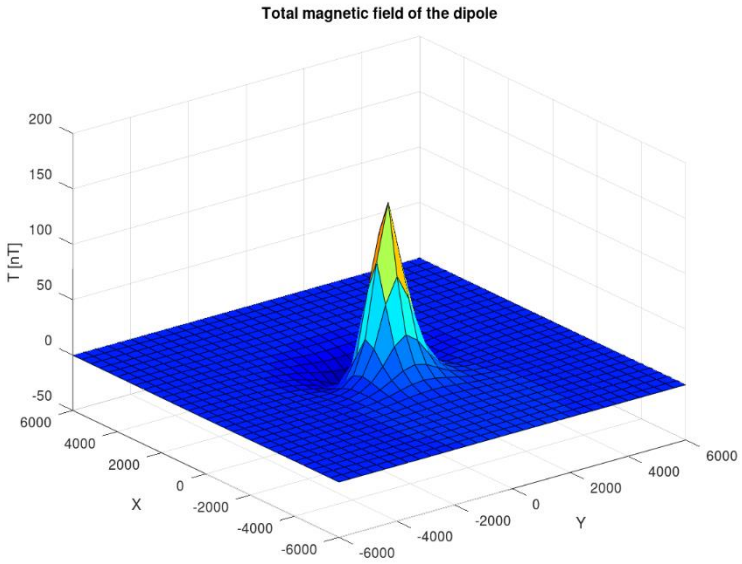

Figure 58; 3D Total Magnetic Field of the Dipole from Equidistant grid

Prior to applying the traditional 2D DFT and the Robust 2D H-LSQ-FT, a 2D frequency spectrum of the dipole source based on the analytical formula was first calculated. Also, the data was transformed between -1 and 1 to obtain appropriate values in the space domain. To effectively evaluate and assess the appropriateness of the two methods in processing the data, well-proven analytical techniques such as the Mean Deviation and Root Mean Square (RMS) Deviation were estimated in other to quantify differences in the applied methods. Figures 59 and 60 show the real and imaginary parts of the spectrum produced from the 2D DFT whilst Figures 61 and 62 show similar spectrums produced from the 2D H-LSQ-FT as applied to the data. A visual comparison between the real and imaginary parts of the spectrum from both methods shows no significant variations or differences. This observation was highly supported by the estimated mean deviation between the exact real spectrum and the real spectrum from the two methods as shown in table 5 below.

Table 5. Analytical output spectra values for Root mean square deviation and Mean deviation for equidistant grid

Spectrum type	Root mean square deviation	Mean deviation
2D DFT Imaginary Spectrum	0.00036976	0.00031870
2D H-LSQ-FT Imaginary Spectrum	0.00031043	0.00029932
2D DFT Real Spectrum	0.0018364	0.00031870
2D H-LSQ-FT Real Spectrum	0.0019361	0.00029932
RTP map of 2D DFT	0.16861	0.1103
RTP map of 2D H-LSQ-FT	0.11038	0.1104

DOI: 10.14750/ME.2020.008

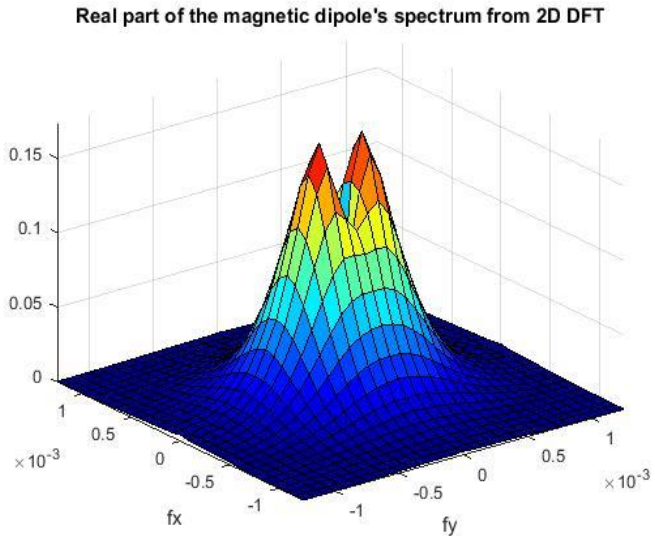

Figure 59; Real part of the magnetic dipole spectrum from the 2D DFT

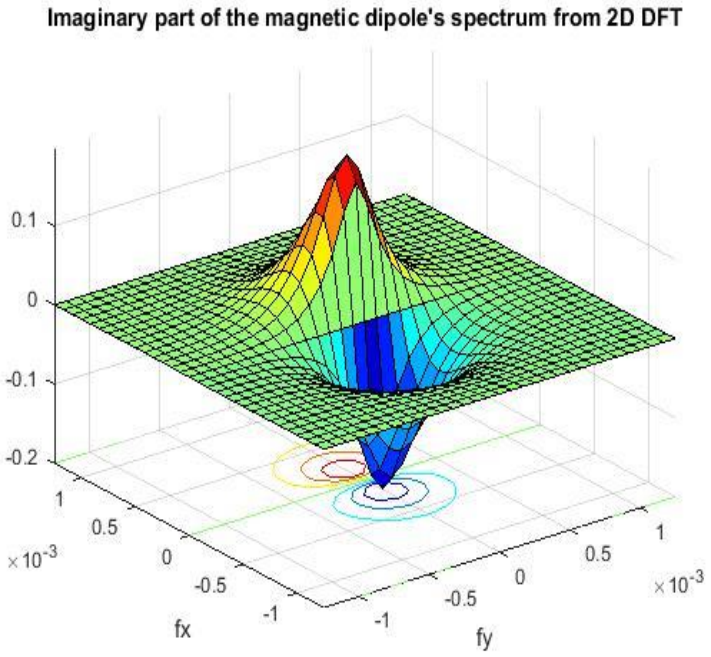

Figure 60; Imaginary part of the magnetic dipole spectrum from the 2D DFT

DOI: 10.14750/ME.2020.008

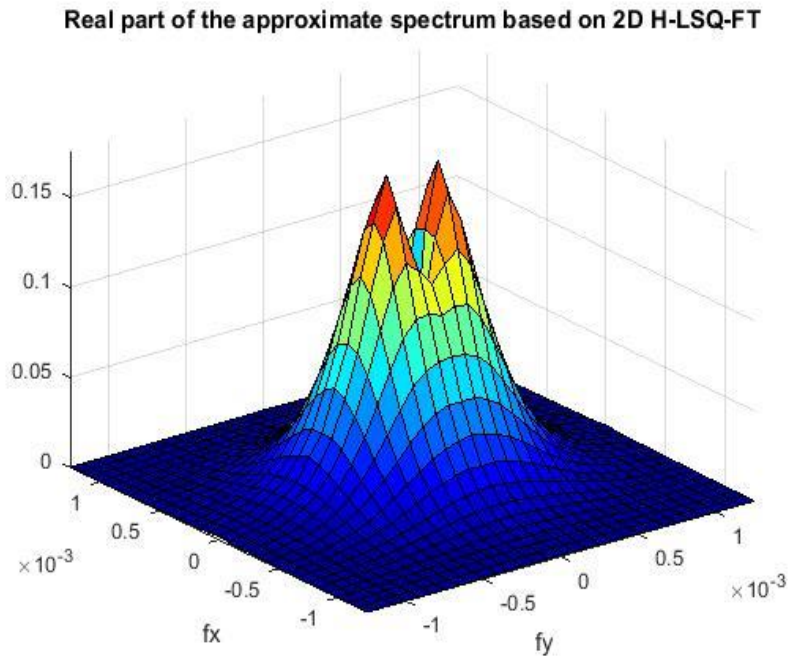

Figure 61; Real part of the magnetic dipole spectrum from the 2D H-LSQ-FT

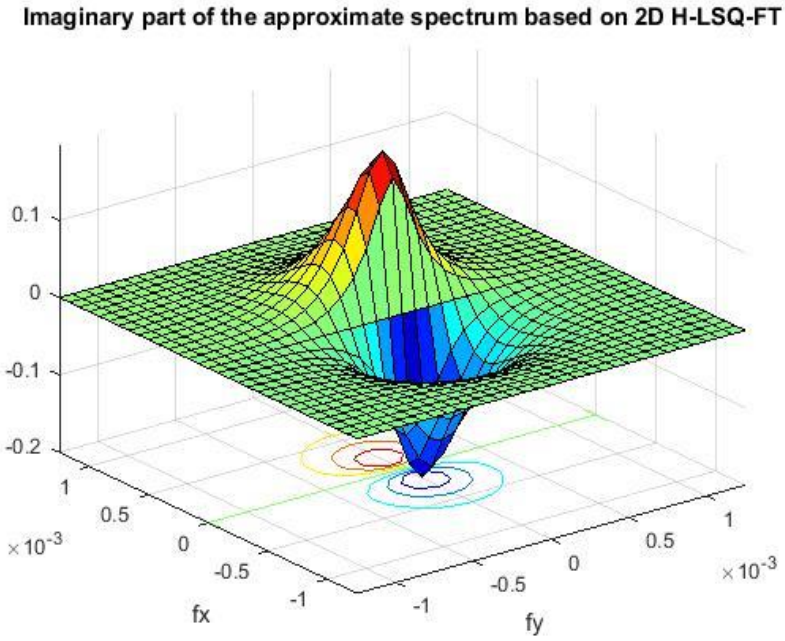

Figure 62; Imaginary part of the magnetic dipole spectrum from the 2D H-LSQ-FT

In other to achieve statistical robustness, accuracy, and stability in the application of the 2D H-LSQ-FT method, Hermite functions of maximum order of '23' was used. The data were scaled using 0.8m in both X and Y directions. This was observed to be a good compromise between accuracy and stability. Using the frequency spectrums produced from the DFT and the H-LSQ-

DOI: 10.14750/ME.2020.008

FT methods, the equidistantly sampled data was reduced to the pole to give Figure 63. A visual comparison shows no significant differences between the two RTP maps, as the anomalies are both concentrated. Analytically, the Root Mean Square Deviation of the exact and approximated RTP for the DFT is 0.16861 (Table 5) whilst that of the H-LSQ-FT is 0.11038. This shows the similarities in the output RTP maps from both methods. The results demonstrate that the traditional 2D DFT and the newly developed Robust 2D H-LSQ-FT were successful in processing datasets acquired on an equidistant grid.

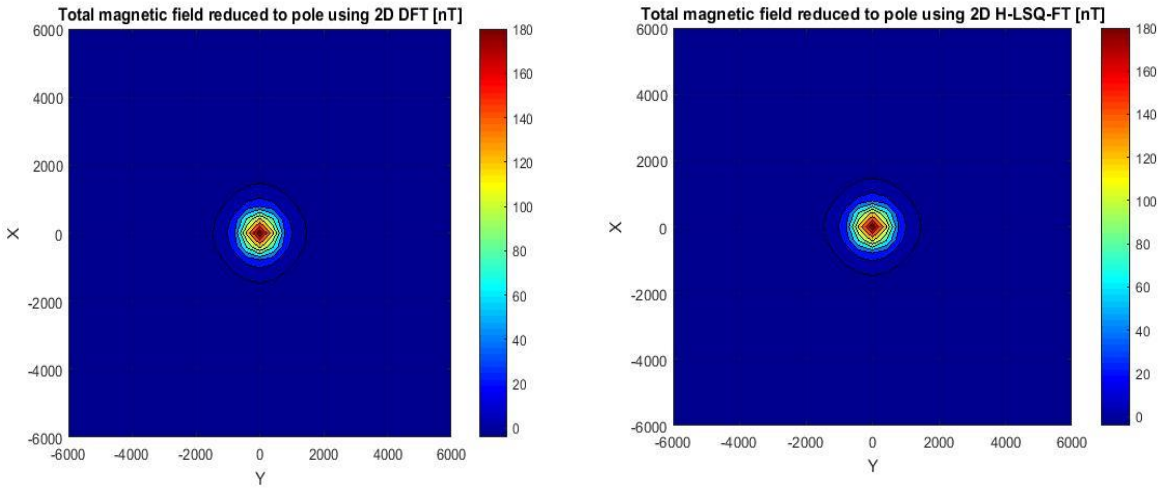

Figure 63; Reduced to the pole maps from 2D DFT (left) and 2D H-LSQ-FT (right) using an equidistant grid.

5.3.2 A magnetic dipole example with non-equidistant sampling

Practically, the traditional DFT Method is ineffective and inefficient in processing geophysical measurement taken in irregular (random) grid, unless additional processing techniques are introduced into the program development, hence the DFT was not applied in this step. To test the applicability of the 2D H-LSQ-FT method on randomly sampled data, the regular 400 m sampling interval as applied in the equidistant survey (Figure 56) was randomized to produce irregular measurements across the survey area as shown in Figure 64 below. Fixed source parameters (inclination, declination and magnetic moment) for the dipole source were equal to that of the equidistant example. The calculated total magnetic field in 2D contoured and 3D surface maps are shown in Figures 65 and 66 respectively. Comparing these maps to that of the equidistant grid maps (Figures 57 and 58 above), it can be observed that the dipole source in the case of non-equidistant is slightly irregular (Figure 65) and does not fill the entire survey area. This is not different for the 3D surface map (Figure 66), the random nature of the sampling interval resulted in a much rougher and unbalanced surface.

DOI: 10.14750/ME.2020.008

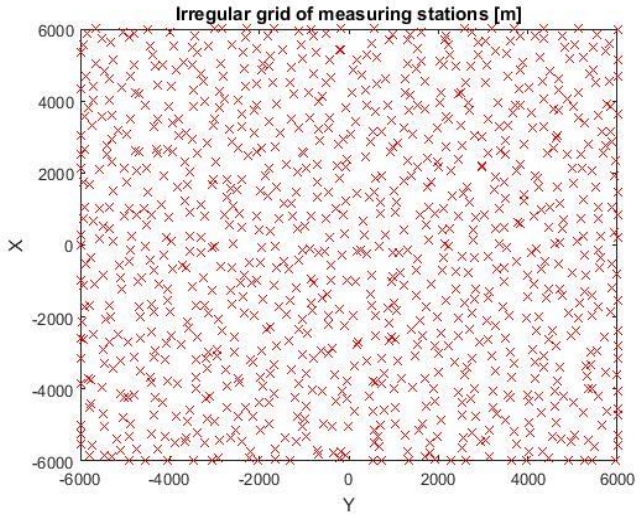

Figure 64; Random grid of measuring stations

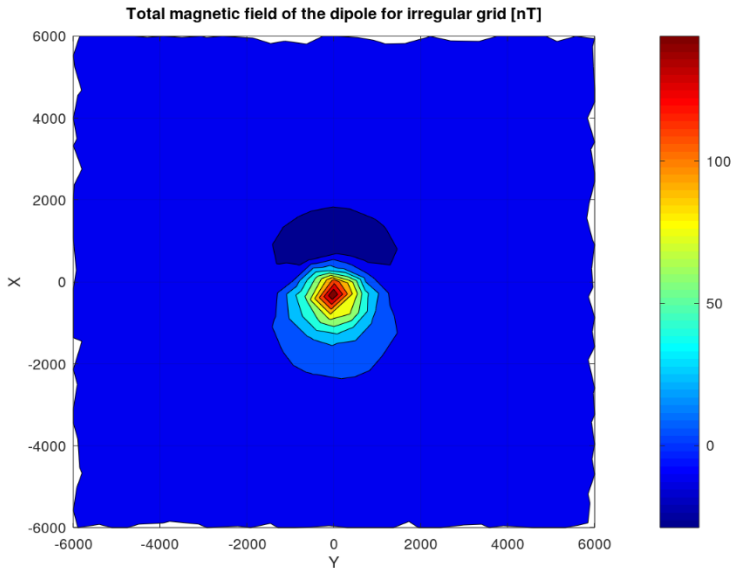

Figure 65; Total Magnetic Field of the Dipole from Non-equidistant grid

DOI: 10.14750/ME.2020.008

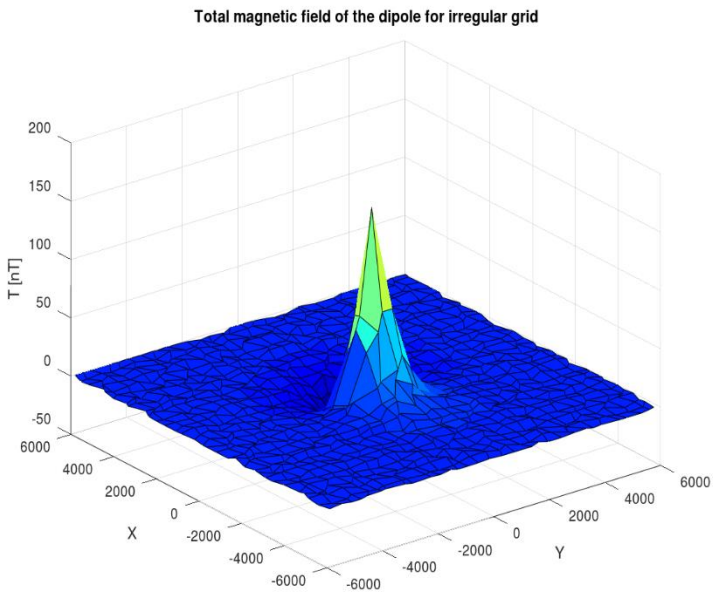

Figure 66; 3D Total Magnetic Field of the Dipole from Non-equidistant grid

The data was re-arranged into columns to facilitate the application of the 2D H-LSQ-FT method. This was followed by a selection of the parameters of the series expansion and weighting function. An inverse problem was formulated using the Gaussian Least Squares approach where the coefficient of the series expansion was estimated. As explained in the equidistant example, Hermit functions of maximum order of 23 were used with a scaling factor of 0.8 m. Figure 67 shows the constrained space-frequency interval as figures 68 and 69 shows the real and imaginary parts of the 2D H-LSQ-FT spectrum in the space domain for the random sampling. Table 6 below shows estimated values for comparing the output spectra from 2D H-LSQ-FT for equidistant and non-equidistant sampling.

Table 6. Analytical output spectra values for equidistant and non-equidistant sampling (H-LSQ-FT)

Spectrum Type	**Root Mean Square Deviation (RMSD)**	
	Equidistant	Non-equidistant
H-LSQ-FT Imaginary Spectrum	0.00031043	0.00031043
H-LSQ-FT Real Spectrum	0.0019361	0.0019361

DOI: 10.14750/ME.2020.008

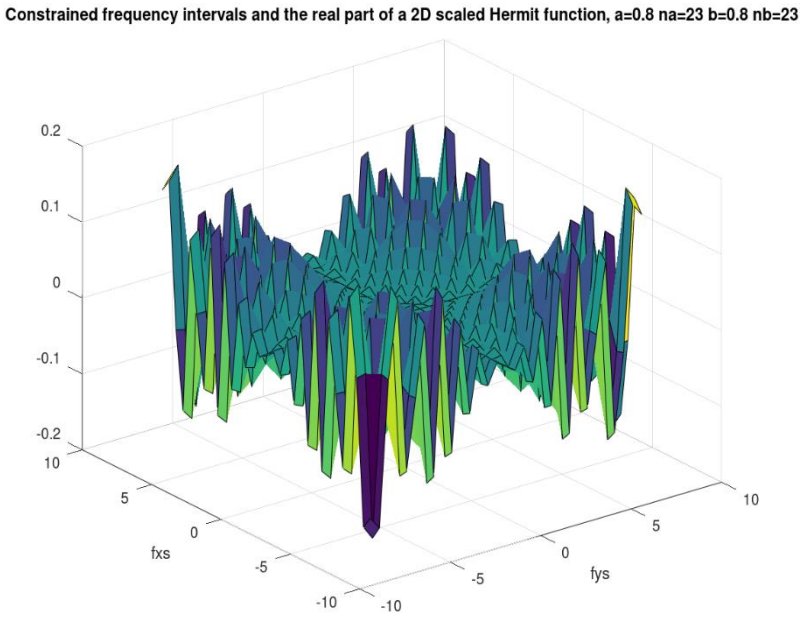

Figure 67; Constrained Frequency Intervals from 2D Scaled Hermit Function

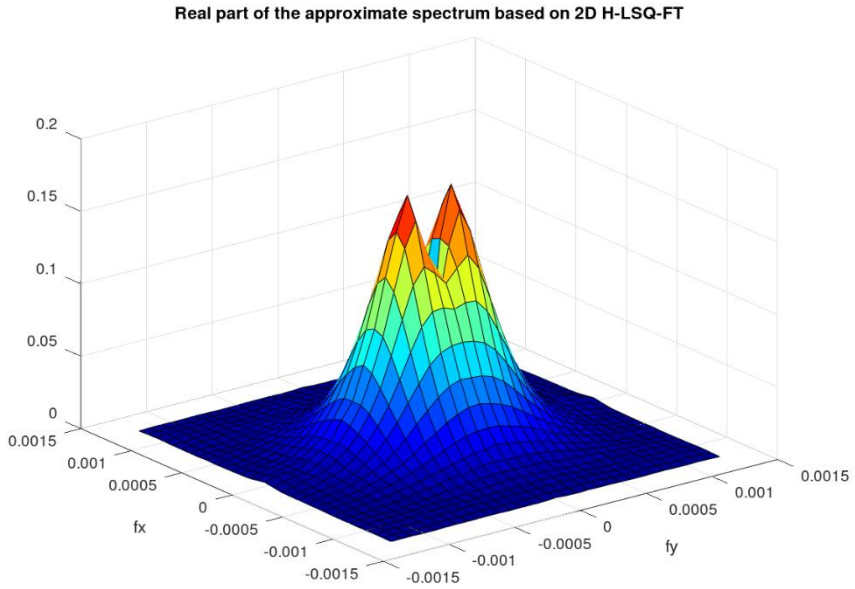

Figure 68; Real part of the magnetic dipole spectrum from the 2D H-LSQ-FT

DOI: 10.14750/ME.2020.008

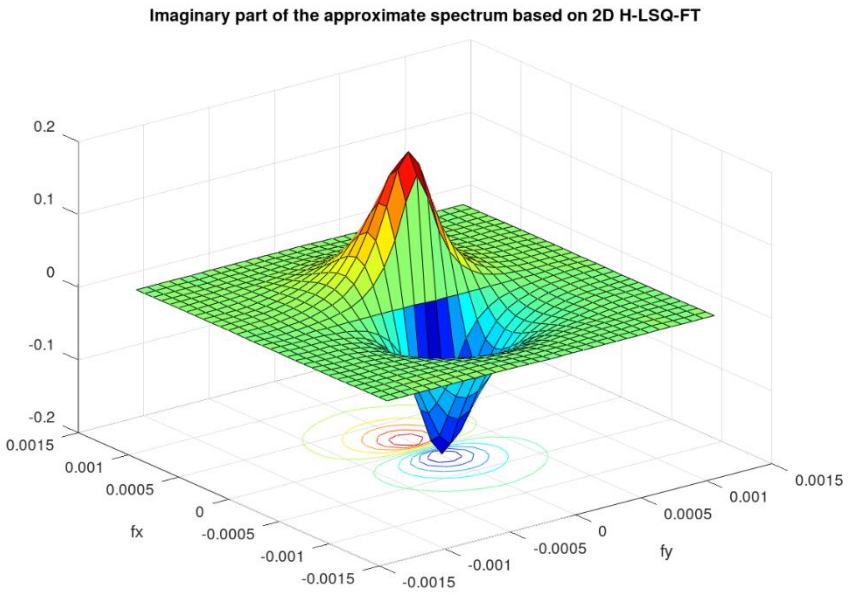

Figure 69; Imaginary part of the magnetic dipole spectrum from the 2D H-LSQ-FT

A visual comparison of figures 68 and 69 to the output spectra of the 2D H-LSQ-FT method for equidistant sampling (Figures 61 and 62 above) shows approximately the same spectra for the real and imaginary parts of the Fourier spectrum in regular and non-regular sampling. The root means square deviations as estimated for the 2D H-LSQ-FT method (Table 6) are 0.00031043 and 0.0019361 respectively for the imaginary and real parts of the spectrum for equidistant and non-equidistant sampling. This demonstrates the 2D H-LSQ-FT ability to derive similar output spectra for both regular and randomly sampled data. From this success, the randomly sampled magnetic dipole data was reduced to the pole as shown in Figure 70. The anomaly as seen from the RTP map is concise and the same as that produced from the equidistant sampling (Figure 63). This indicates that the newly created Robust 2D H-LSQ-FT method gave the same result or output RTP maps for the equidistant and non-equidistantly sampled datasets. The Root means square deviation of the exact and approximated reduction to pole is 0.24752 while the Mean deviation of the exact and approximated reduction to pole is 0.15670.

DOI: 10.14750/ME.2020.008

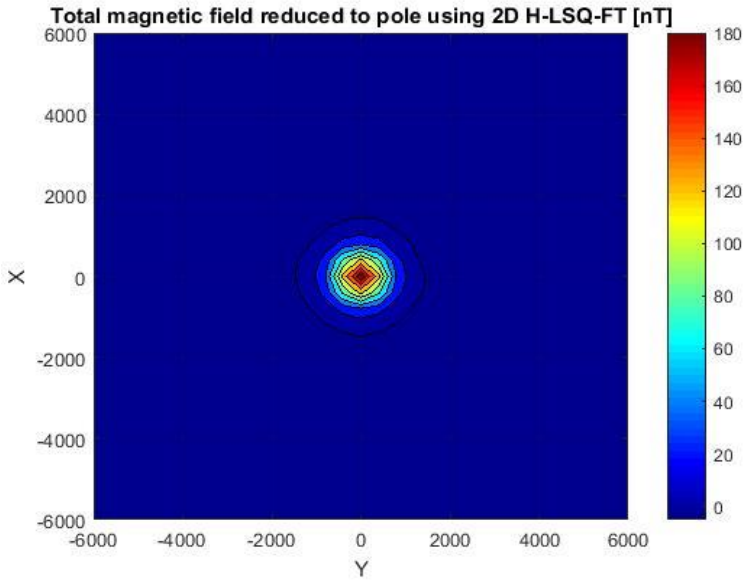

Figure 70; Reduced to the pole map by 2D H-LSQ-FT using Non-equidistant grid

5.1.4 Numerical test using a synthetic magnetic data

To prove the applicability of the 2D inversion based H-IRLS FT method on randomly sampled data, it was tested on synthetic magnetic data sets sampled at regular equidistant and non-equidistant intervals. In all, 41 x 41 measurement points were sampled along a 5 m x 5 m regular grid and further randomized to obtain non-equidistant measurements. Data were generated for a surface between +/-100 m both in the x and y directions above a 'C_L' shaped magnetic body (inclination I=63°, declination D=3°, magnetization 200 nT). The surface magnetic data were calculated by the Kunaratnam (1981) method and was subsequently reduced to the pole (I=90°) by applying the formula in the frequency domain using Eqn 75. First, the reduction to the pole was performed by using the conventional 2D DFT algorithm on equidistantly sampled magnetic data. The map of noiseless magnetic data on the equidistant grid and its reduced to pole version are given in Figure 71. The 2D spectrum of the regularly sampled data set was calculated by means of 2D DFT and 2D H-IRLS-FT (Figure 72). For numeric reasons, the calculated spectra were made on the data set transformed to [-1,1] in both x and y coordinates resulting in an appropriate scale in the wavenumber domain. The 2D H-IRLS-FT spectrum was calculated using Hermite functions of the (maximal) order of M=28. The similarity of the two results are obvious and can be increased by using higher-order basis functions. This has a consequence in increased computation time and in a rapid change of the condition number. The M=28 was found as a good compromise between accuracy and stability.

DOI: 10.14750/ME.2020.008

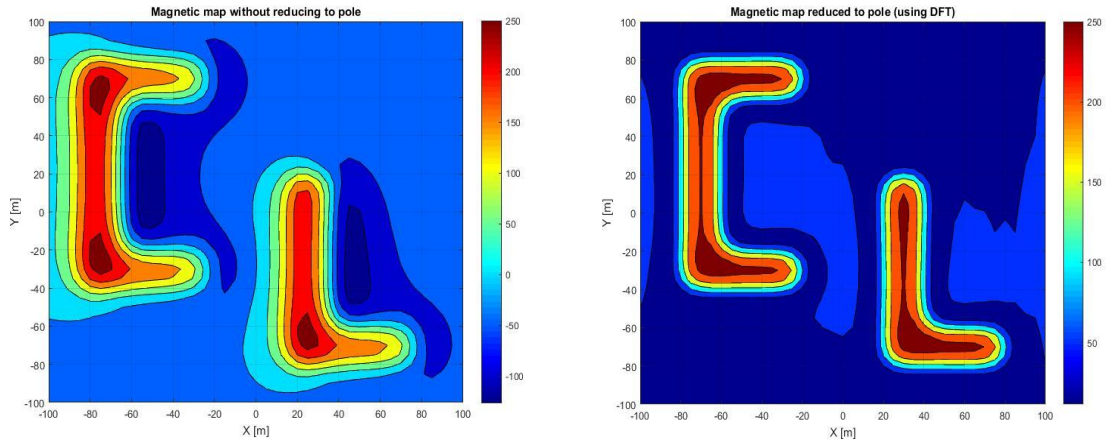

Figure 71; Noise-free Magnetic Map calculated on Equidistant grid (left) and it Reduced to the Pole map using the conventional DFT(right).

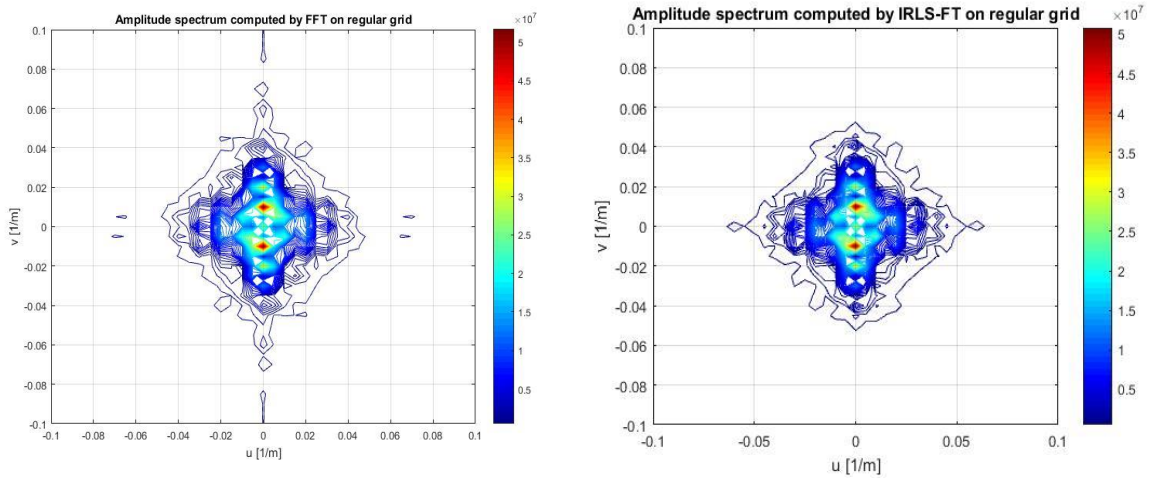

Figure 72; DFT spectrum of the Magnetic data set calculated on the Equidistant grid (left) and the H-IRLS-FT spectrum of the Magnetic data set calculated on the Non-Equidistant grid (right).

The total magnetic intensity data was further calculated on a randomized set of sample points. The randomly selected x and y coordinates are shown in Figure 73, the sample points are defined by ordering all y coordinates to all of the x position resulting in 41x41 sample points.

DOI: 10.14750/ME.2020.008

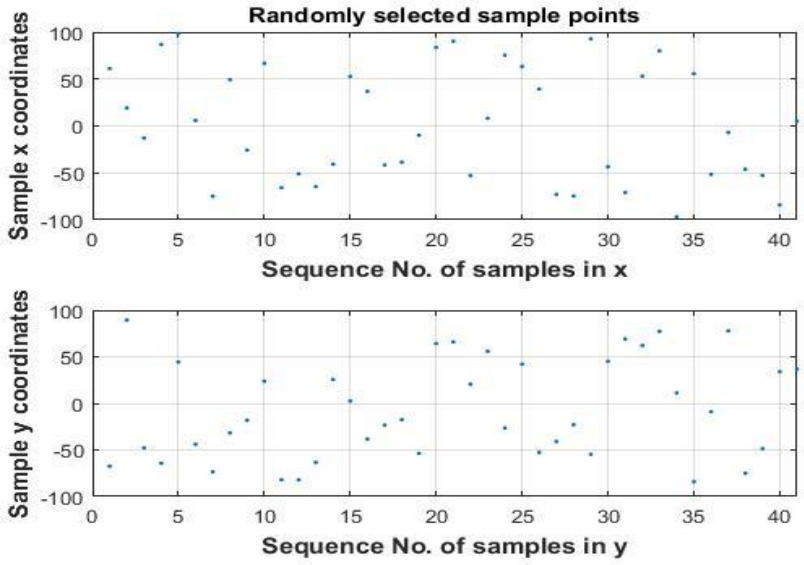

Figure 73; The randomly selected x and y coordinates used for generating 41x41 sample points

The magnetic field data calculated in these non-regular positions were processed using the 2D H-IRLS-FT algorithm, the resulting spectrum is given in Figure 74 (for the sake of comparability, the IRLS-FT spectrum calculated in the regular grid is also shown, Figure 72 right). As can be seen, the two spectra are the same (both calculations were performed with the same inversion parameters using series expansion by means of Hermite functions of M=28 order). The produce spectra prove that the 2D H-IRLS-FT algorithm has a very important feature, that it works on both regularly or irregularly sampled data sets.

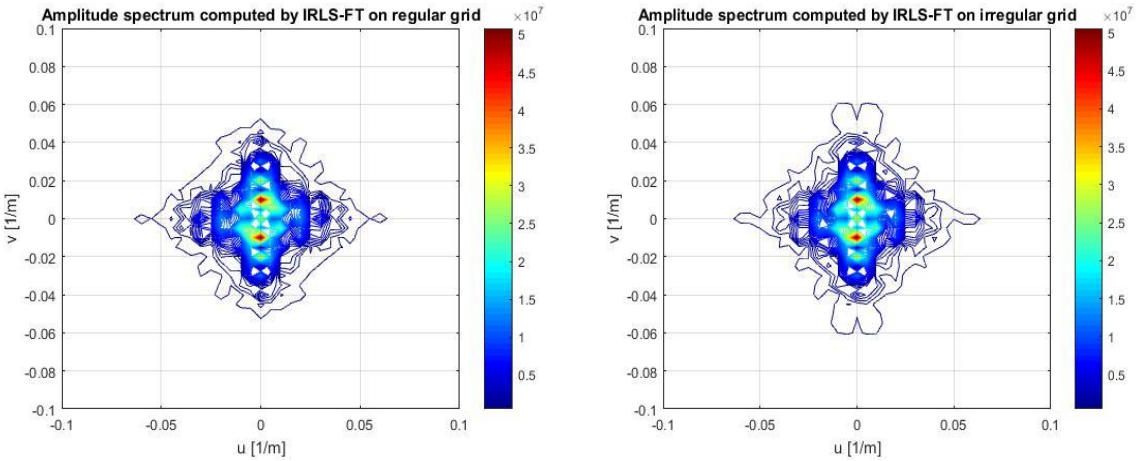

Figure 74; 2D H-IRLS-FT spectrum of the Magnetic data set calculated on Equidistant grid (left) And non-Equidistant grid (right).

DOI: 10.14750/ME.2020.008

From this fact, it is straightforward to expect that the method gives the same result in reducing to pole for both regularly sampled (in Figure 75, left) and randomly sampled data sets (Figure 75, right). As was observed in the above 1D example, the inversion based 1D H-IRLS-FT algorithm was successfully applied in processing a non-regularly sampled data set. This important observation has been extended to the two-dimensional case by testing the 2D H-IRLS-FT method on the randomly selected set of sample points to give a similar result. The efficiency of the new 2D H-IRLS-FT algorithm in processing random-walk sampled data is clearly observable, demonstrating that the method gives accurate results even in "random walk" measurements.

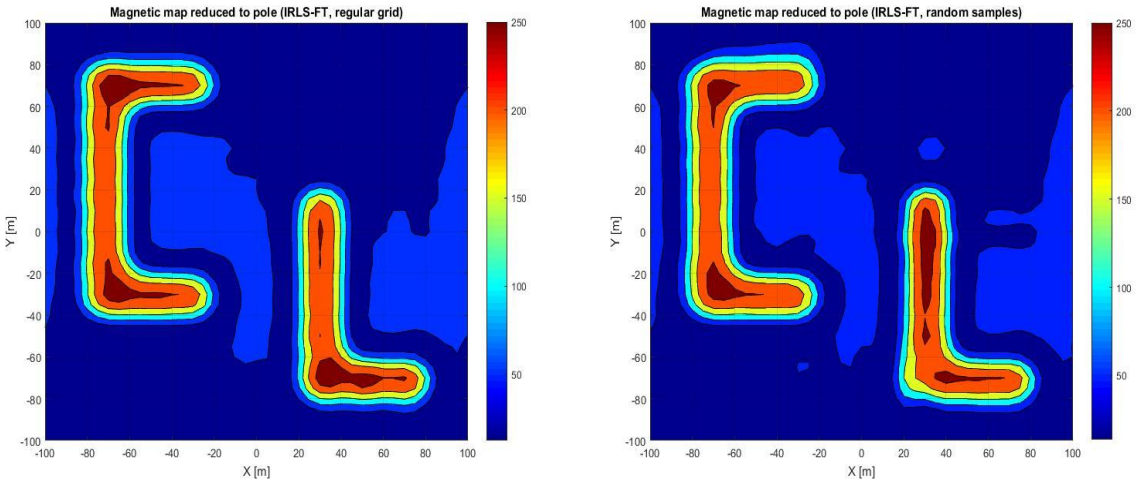

Figure 75; The pole reduced Magnetic data sets using 2D H-IRLS-FT for Equidistant sampling (left) and non-Equidistant or random sampling (right).

DOI: 10.14750/ME.2020.008

Chapter 6

FIELD EXAMPLES USING RANDOM-WALK GEOMETRY

6.1 Geology of the Study Area

The area from which the magnetic dataset was selected for processing is located in the north-eastern region of Hungary (Fig. 76). The subsurface geological structures and the rock formations with which a local magnetic anomaly occurs is situated near the town called Mezőkövesd, to the south of the Bükk Mountains (Figure 77). The shaded relief map (Figure 77) indicates the main mass of the Bükk Mountains (darker green area) and the positions of the towns (pink areas) in the area (Miskolc, Mezőkövesd, Eger). The Pre-Cenozoic basement outcrops close to the surface over the area of the Bükk Mountains and gradually dips to the southern direction due to a fault zone where it is covered by Neogene clastic sedimentary and pyroclastic beds. Faults with strike direction of north-east to the south-west and dip direction of the north-west are dominant in the southern part of the Bükkalja mountains. To the south-east of the Bükkalja mountains, an elongated depression, the Vatta-Maklár trough occurs in the basement with a strike direction parallel to that of the above-mentioned faults bordering it. The depth of the buried basement may reach as deep as 3000 m below the vertical datum (Baltic Sea vertical datum) along the axis of the trough (Fig. 78). The trough is bordered to the south-east by a threshold running below the area near Mezőkövesd (Mezőkövesd threshold). Along this threshold, the basement is in an uplifted position compared to the trough. The top of the threshold near Mezőkövesd (Fig. 78) is about 800 m below the vertical datum (Pelikán et al. 2005).

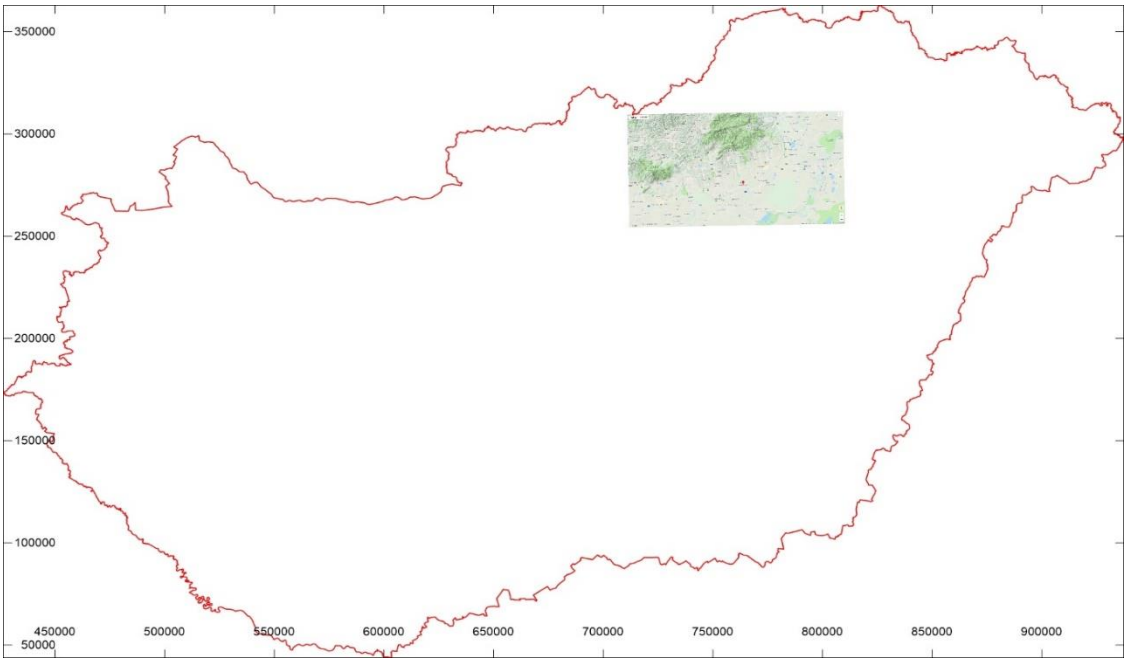

Figure 76: The location of the area in Hungary.

DOI: 10.14750/ME.2020.008

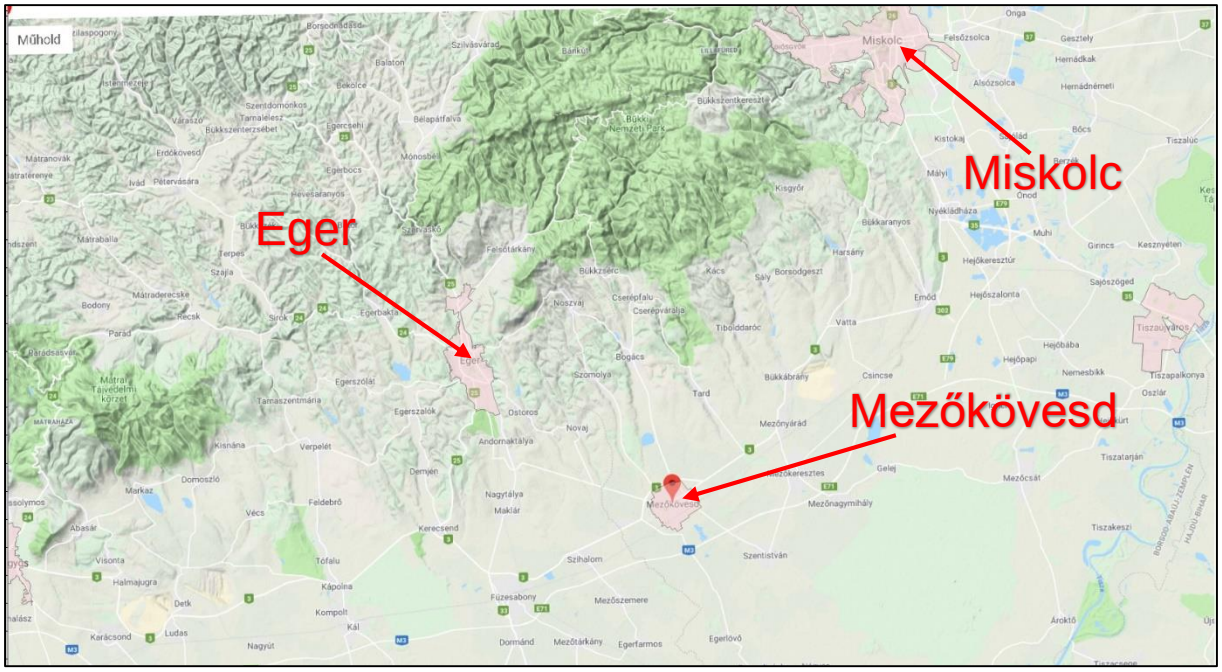

Figure 77: Geographical surroundings of the selected area. (www.google.hu)

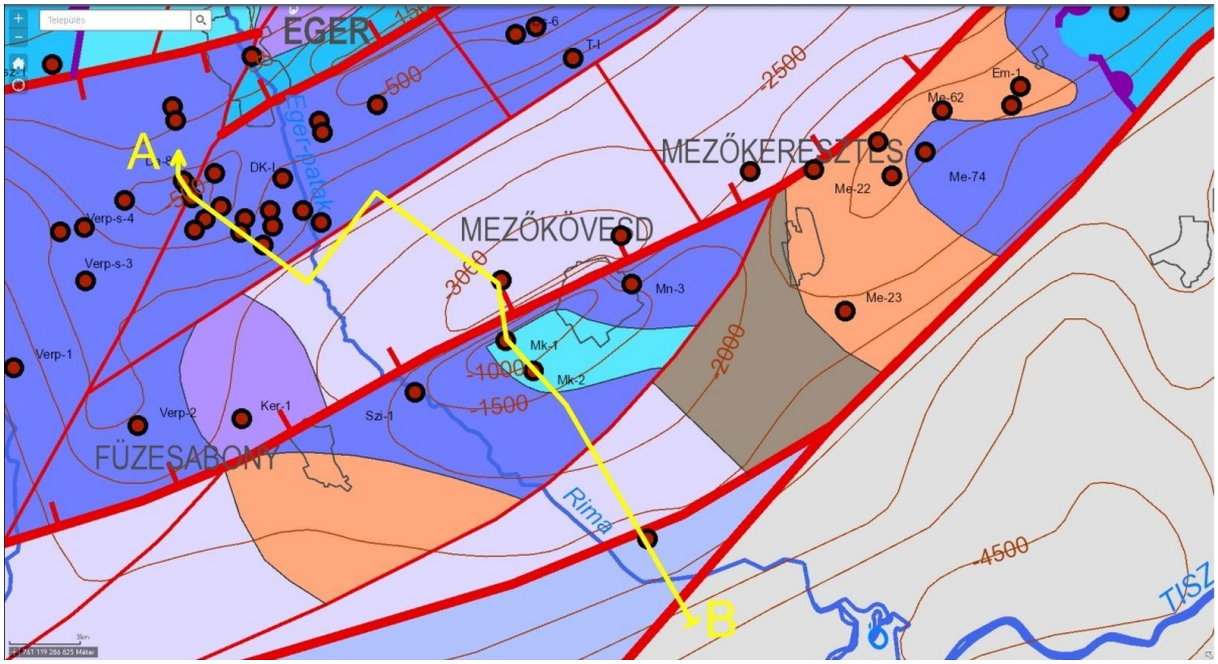

Figure 78: Pre-Cenozoic basement below the southern foreland of the Bükk mountains. The brown contour lines indicate the depth levels of the basement. The coloured areas mostly bordered by tectonic lines (red lines) show the arrangement of different rock formations. The red points with black rims represent the positions of former drill sites. (web maps of Mining and Geological Survey of Hungary, https://map.mbfsz.gov.hu)

DOI: 10.14750/ME.2020.008

This uplift of the basement causes excess mass in its surroundings because its bulk density is significantly greater than that of the Neogene beds bordering it laterally. The effect of this lateral anomaly in the bulk density was detected by gravity measurements. The Bouguer anomaly map of the area presented in Figure 79 directly indicates both the Vatta-Maklár trough with a stripe of high negative values (dark green colour) and the Mezőkövesd threshold with less negative values (light green colour). The area filled with brown (positive Bouguer anomaly) colour shows the main mass of the Bükk Mountains where the high-density basement is above the level of its foreland.

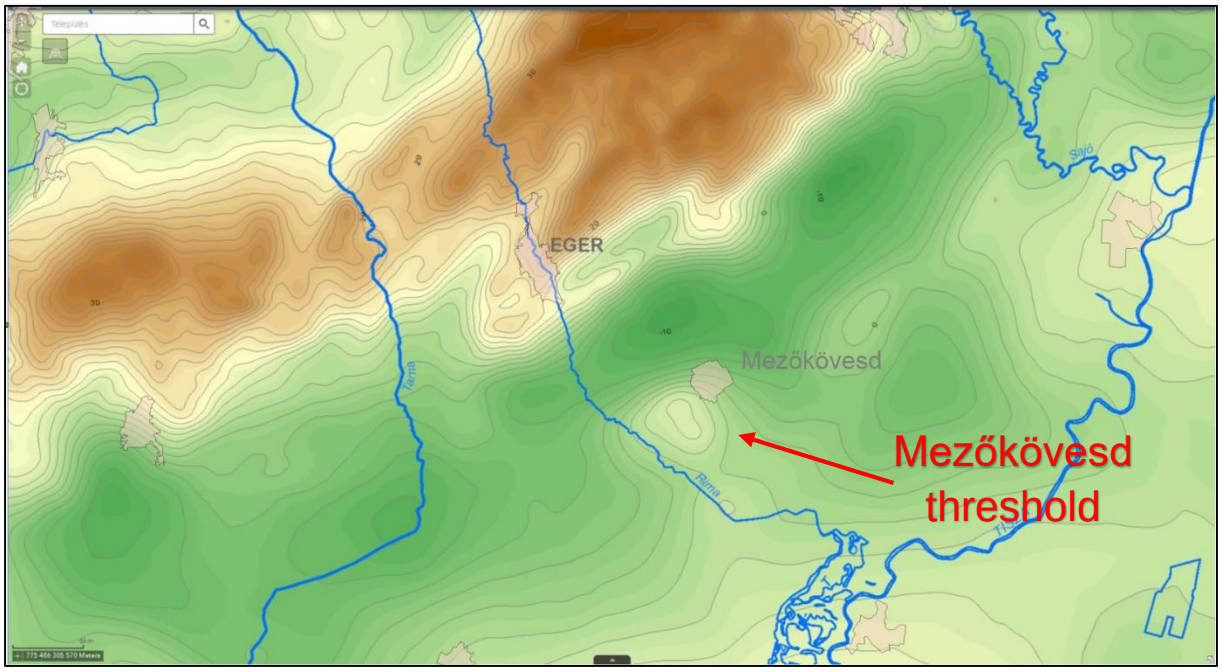

Figure 79: Bouguer anomaly map of the selected area with the effect of Mezőkövesd threshold located to the south-east of Vatta-Maklári trough. The difference between the neighboring isogal lines is 2 mgal. (web maps of Mining and Geological Survey of Hungary, https://map.mbfsz.gov.hu)

The narrower environment of the Bouguer anomaly caused by the Mezőkövesd threshold can be studied in more detail in Figure 80. A refracted cross-section line between two points of the area (A and B) is also displayed. The curve of Bouguer anomaly and the geological cross-section along this line are presented in Figure 81. The route of the cross-section line is also shown in Figure 78 (yellow line).

DOI: 10.14750/ME.2020.008

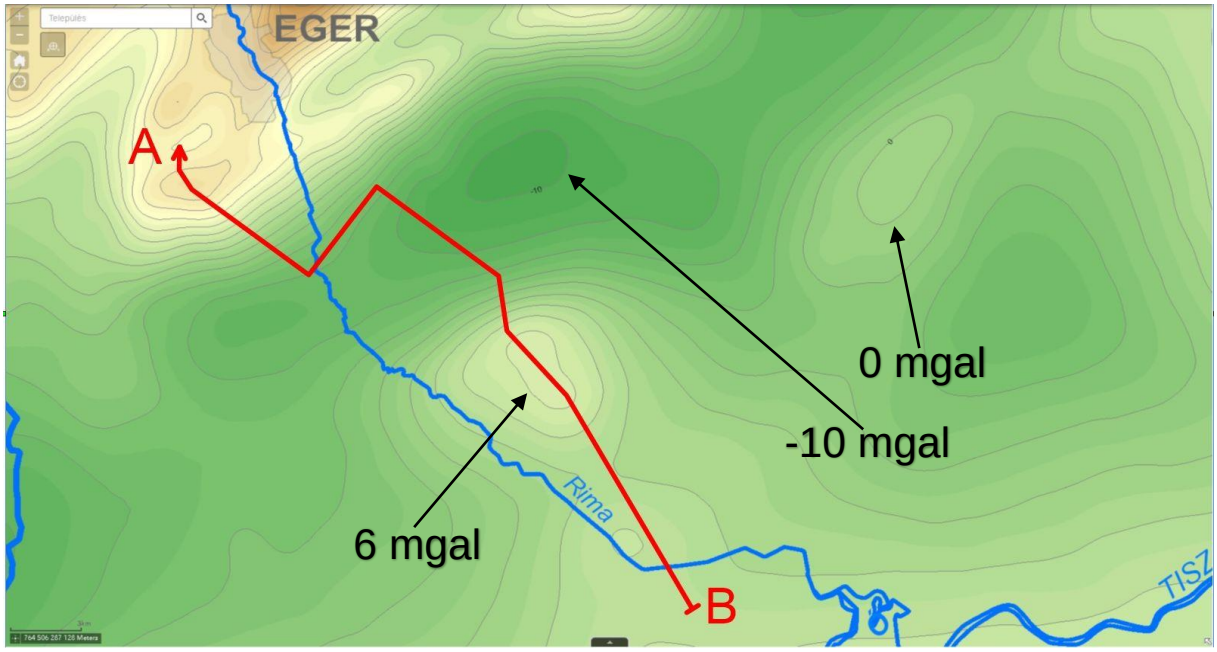

Figure 80: Bouguer anomaly map with a refracted cross-section line. The values of some isogal lines are also marked. (web maps of Mining and Geological Survey of Hungary, https://map.mbfsz.gov.hu)

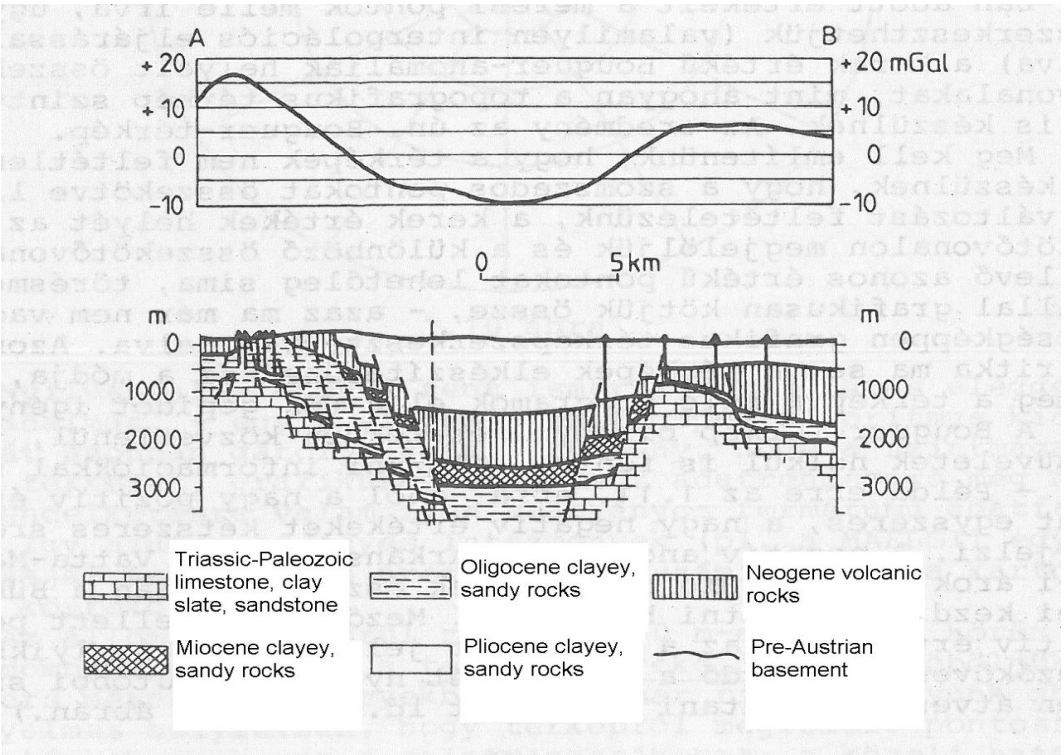

Figure 81: Bouguer anomaly curve (above) and geological cross-section (below) along a refracted cross-section line between points A and B. (Takács et al., 1993)

DOI: 10.14750/ME.2020.008

The geological cross-section in Figure 81 shows that a significant amount of Neogene (Miocene) volcanic rocks, mostly pyroclastic rocks, were deposited on the older formations. The sites of the deep boreholes formerly drilled above and near the Mezőkövesd threshold can be studied in Figure 82. One of these boreholes (Mk-4) traversed this Miocene stratigraphic unit between 440 m and 1496 m and penetrated into the Oligocene sediment. The volcanic rock was described as dark grey andesite tuff. The depth was measured from the drill floor with an elevation of 111,77 m above the vertical datum (Adriatic Sea vertical datum).

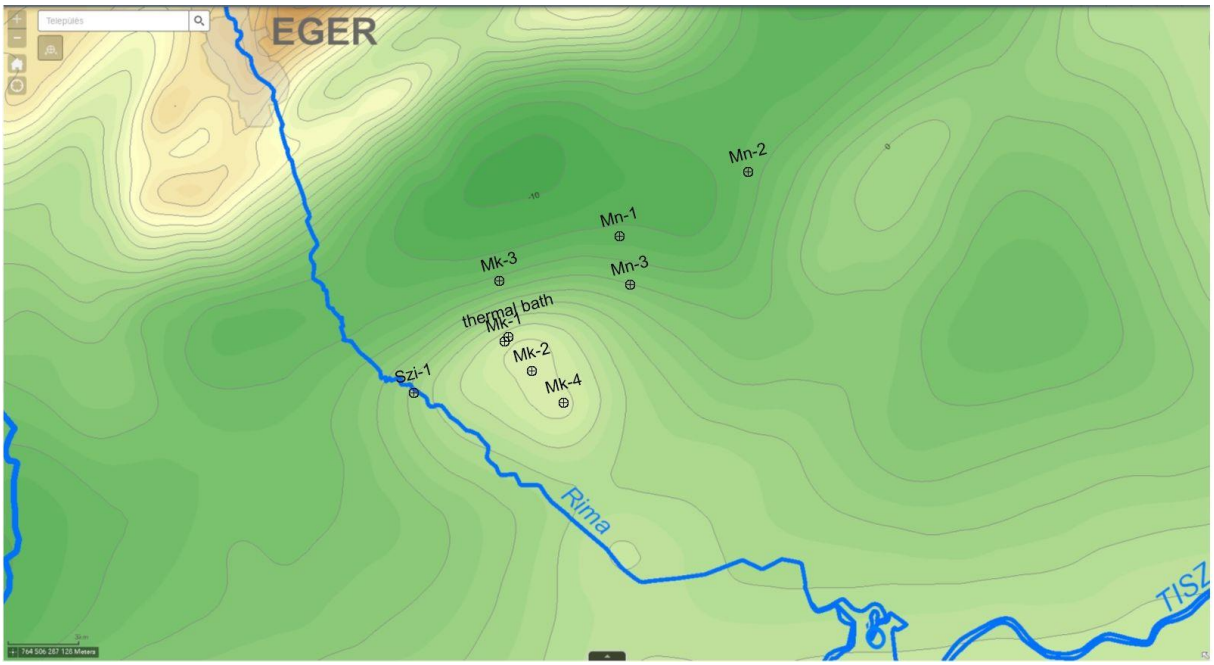

Figure 82: Bouguer anomaly map with the sites of deep boreholes drilled above and near the Mezőkövesd threshold. (the abbreviations in the short names come from the names of the near settlements: Mk = Mezőkövesd, Mn = Mezőnyárád, Szi = Szihalom) (web maps of Mining and Geological Survey of Hungary, https://map.mbfsz.gov.hu)

The presence of ferromagnetic minerals in this volcanic rock is the reason why not only a Bouguer anomaly but also a local magnetic anomaly was detected above the Mezőkövesd threshold. The dipolar anomaly in the vertical component of the magnetic field which is associated with the geological structure is presented in Figure 83. The route of the refracted cross-section line previously introduced is also shown to help in comparing the positions of gravity and magnetic anomalies. A square area with a side length of 10 km was selected from the magnetic anomaly map, which is shown in Figure 84. This area includes primarily the dipolar anomaly in the vertical component of the magnetic field connected to the Mezőkövesd threshold. The regional effect coming from deeper and lateral geological objects functioning as magnetic causative bodies is not too significant within this delimited volume.

DOI: 10.14750/ME.2020.008

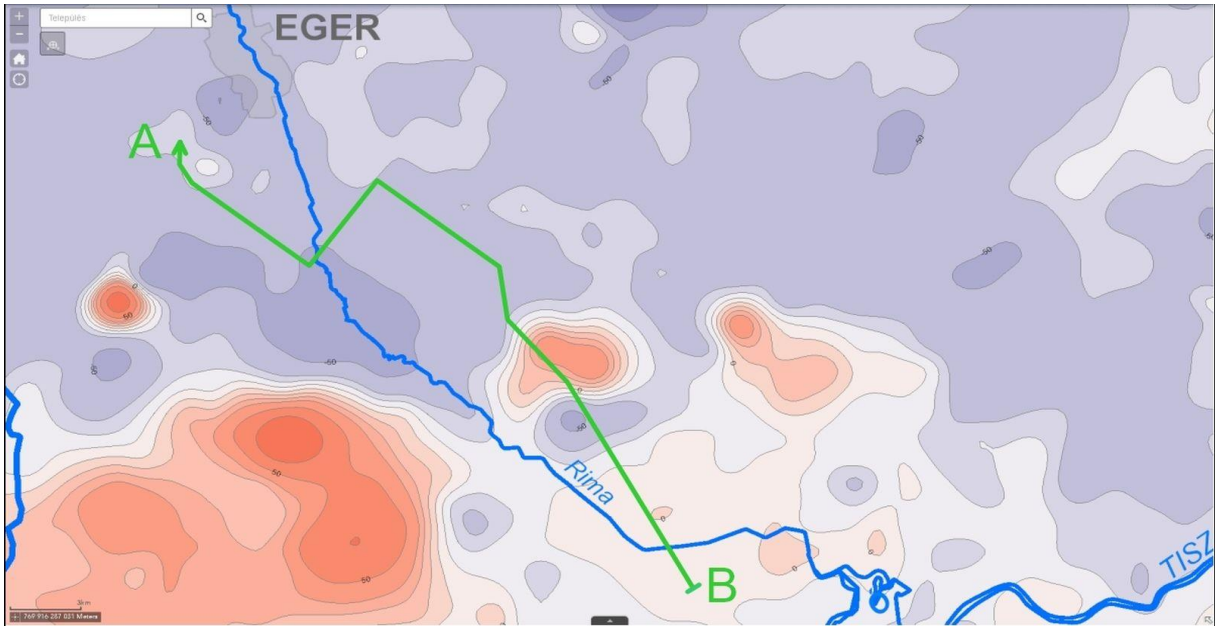

Figure 83: Map of the vertical component of the local magnetic field with the cross-section line previously introduced. The cross-section line passes a dipolar magnetic anomaly associated with the Miocene volcanic rock deposited onto the uplifted block of the Pre-Cenozoic basement and Paleogene sediments. The difference between the neighboring isodynamic lines is 12.5 nT. (web maps of Mining and Geological Survey of Hungary, https://map.mbfsz.gov.hu)

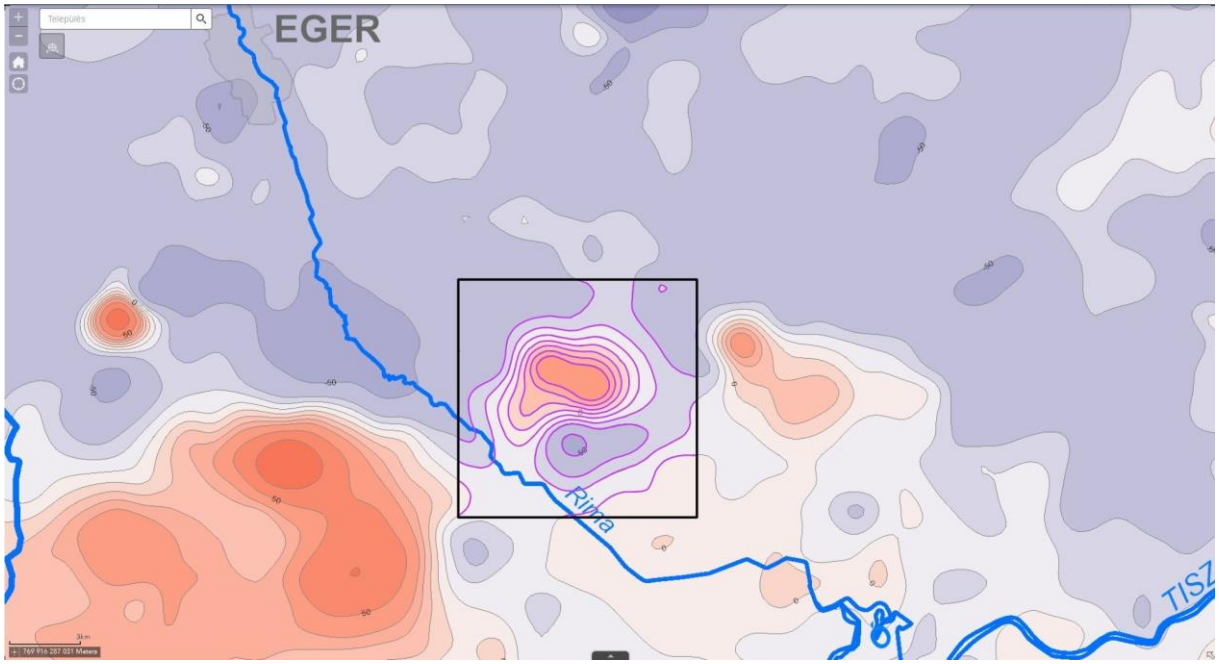

Figure 84: Map of the vertical component of the local magnetic field with the selected local anomaly bordered by a square area. (web maps of Mining and Geological Survey of Hungary, https://map.mbfsz.gov.hu)

DOI: 10.14750/ME.2020.008

6.2 A field example with Equidistant sampling

we digitized the magnetic map shown in Figure 84 along its isodynamic lines, the magnetic anomaly was approximated over regular grids of imaginary measuring stations. Two grids with different sampling intervals (250 m and 500 m) are illustrated in Figure 85. The approximated magnetic anomaly based on these grids are represented on the contour maps of Figure 86 and the surface maps of Figure 87. At first, the 2D DFT algorithm was applied for producing the spatial frequency spectrum of the anomaly, then it was filtered with the transfer function of the reduction to the magnetic pole. The values of the input parameters necessary for the transfer function were determined by means of experiments performed with the magnetic anomaly of a point-like dipole source. Since the analytical formulae describing the anomaly of such a source are known for both spatial and spatial frequency domains, the suitably chosen values of the parameters resulted in very similar orientation and polarity of the dipole's anomaly to those of the selected local anomaly. The resultant magnetism of the anomaly was not treated as the sum of remanent and induced magnetism, because no exact information was available for the inclination and declination of former magnetic fields as well as the displacement and the rotation of the piece of basement carrying the magnetic anomaly during the last few millions of years. The selected values of the input parameters derived from the above-mentioned experiments were the following: an inclination of both the remanent and the external magnetic fields = -47 ° whereas Declination of both the remanent and the external magnetic fields= -10 °

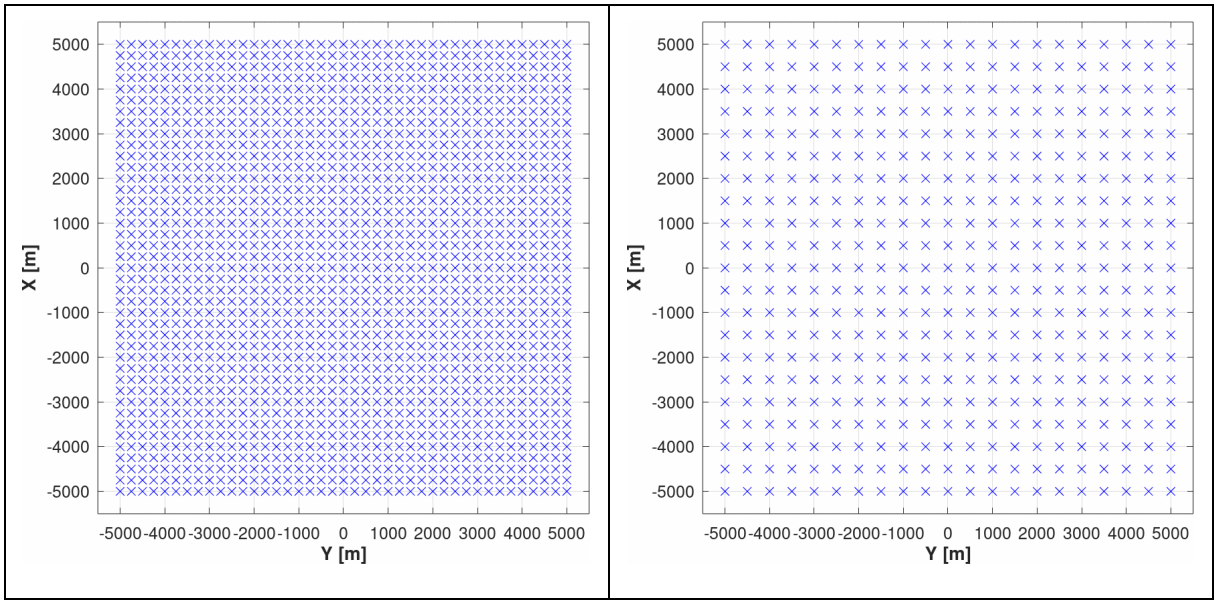

Figure 85: Regular grids of imaginary measuring stations with sampling intervals of 250 m (left) and 500 m (right) in both horizontal directions.

DOI: 10.14750/ME.2020.008

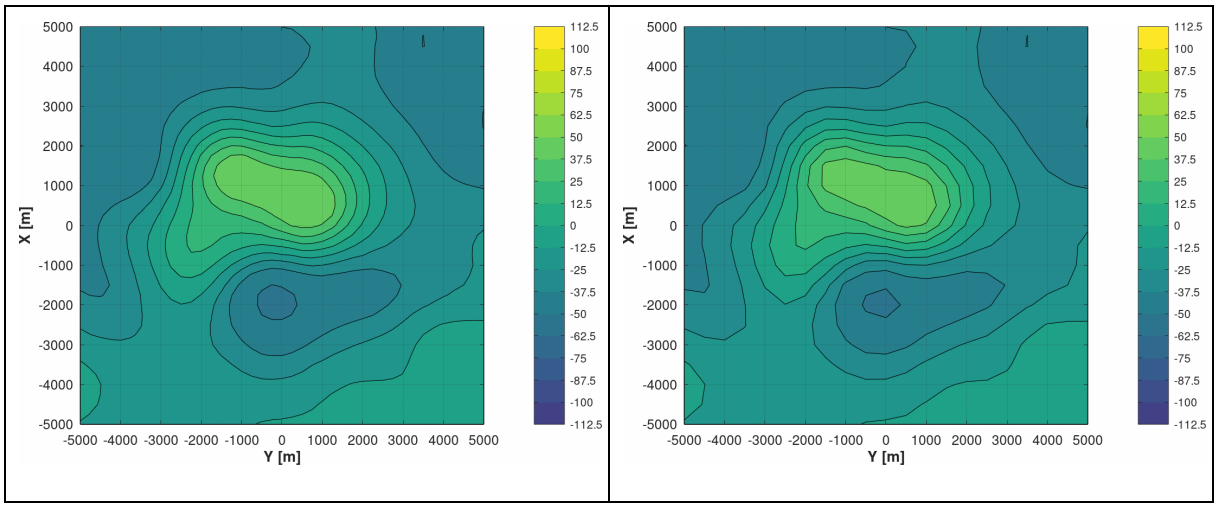

Figure 86 Contour maps of the approximated magnetic anomaly based on the regular grids with sampling intervals of 250 m (left) and 500 m (right). The unit of isodynamic lines is in nT.

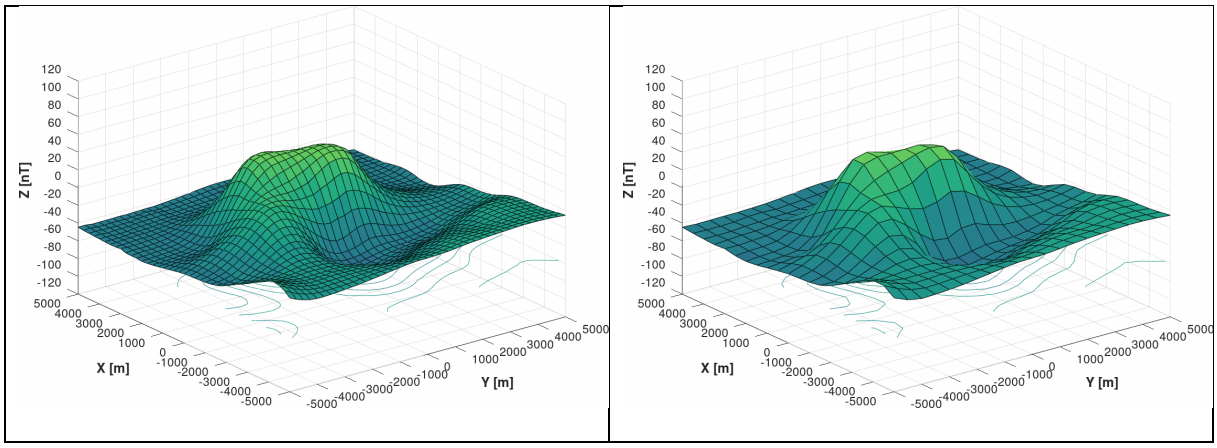

Figure 87: Surface maps of the approximated magnetic anomaly based on the regular grids with sampling intervals of 250 m (left) and 500 m (right).

After the 2D DFT of the magnetic anomaly had been filtered with the transfer function of the reduction to pole, the result of this operation was transformed back to the spatial domain by means of the 2D IDFT algorithm. Essentially, the result of this inverse transform is regarded as the reduction of the magnetic anomaly to the magnetic pole. It is shown for the magnetic anomaly sampled with two different intervals on the contour maps (Figure 88) and the surface maps (Figure 89). Although these figures show the success of the filtering operation, the quality of the result is not too good. The output shows an anomaly of monopole nature dominating the maps. The roughness of the results is due to the computational noise connected with the 2D DFT algorithm. Clearly, the rarer sampling causes a lower resolution of the filtered anomaly.

DOI: 10.14750/ME.2020.008

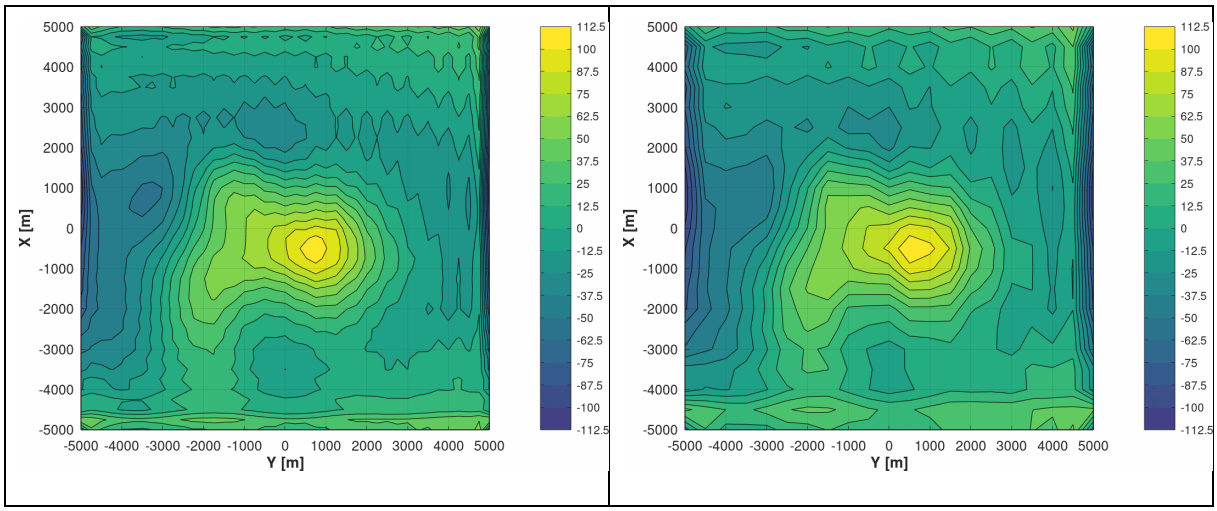

Figure 88: Contour maps representing the reduction of the magnetic anomaly to the magnetic pole for the regular grids with sampling intervals of 250 m (left) and 500 m (right). The filtering operation was executed on the 2D DFT of the anomaly.

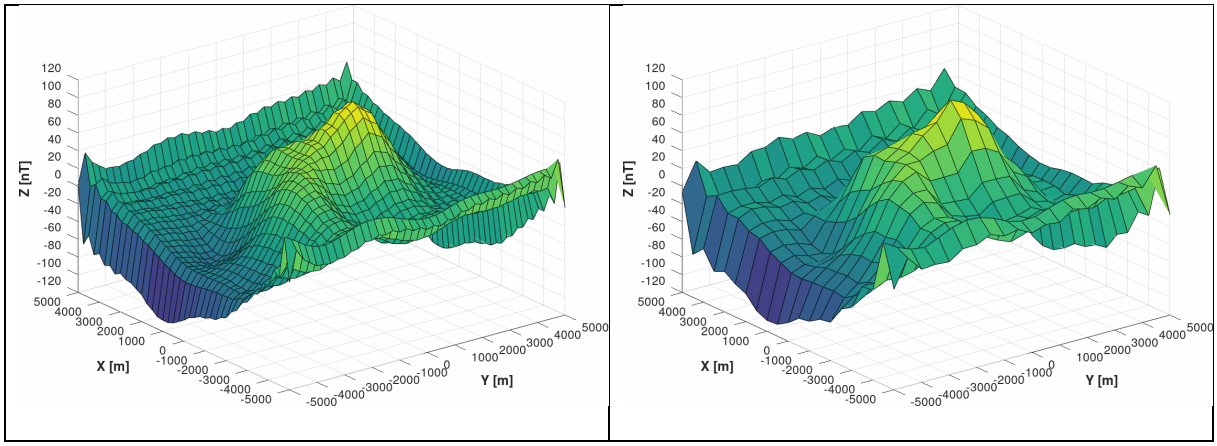

Figure 89: Surface maps representing the reduction of the magnetic anomaly to the magnetic pole for the regular grids with sampling intervals of 250 m (left) and 500 m (right). The filtering operation was executed on the 2D DFT of the anomaly.

Next, the 2D H-LSQ-FT algorithm was applied for producing the spatial frequency spectrum of the anomaly. The application of this algorithm requires the suitable setting of the parameters (the scaling factor of the weighting function and the number of terms involved in the series expansion) controlling the series of Hermit functions used for approximating the Fourier spectrum. The optimal values of these parameters highly depend on the input dataset. After adjusting the parameters, the approximation of the spatial frequency spectrum of the anomaly was computed by means of the 2D H-LSQ-FT. This intermediate result was filtered with the transfer function of the reduction to pole. The filtering operation provided the final result, the reduction of the magnetic anomaly to the magnetic pole, which is represented for the magnetic

DOI: 10.14750/ME.2020.008

anomaly sampled with two different intervals on the contour maps of Figure 90 and the surface maps of Figure 91. A quick visual comparison of the results with those ones based on the 2D DFT of the anomaly is enough to decide that the application of 2D H-LSQ-FT for the reduction to magnetic pole is much more advantageous.

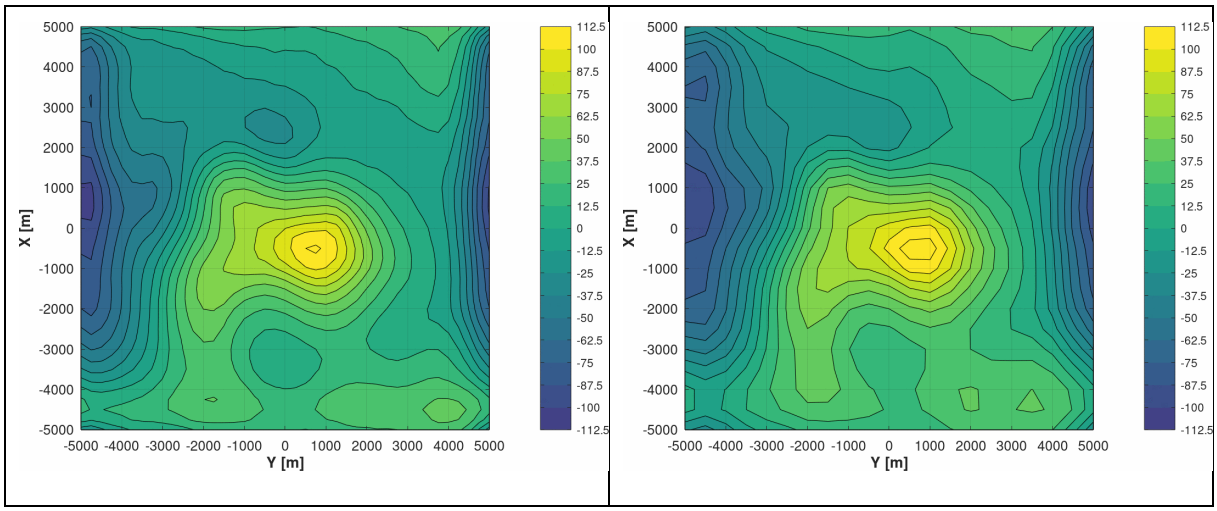

Figure 90: Contour maps representing the reduction of the magnetic anomaly to the magnetic pole for the regular grids with sampling intervals of 250 m (left) and 500 m (right). The unit of isodynamic lines is nT. The filtering operation was executed on the 2D HLSQFT of the anomaly.

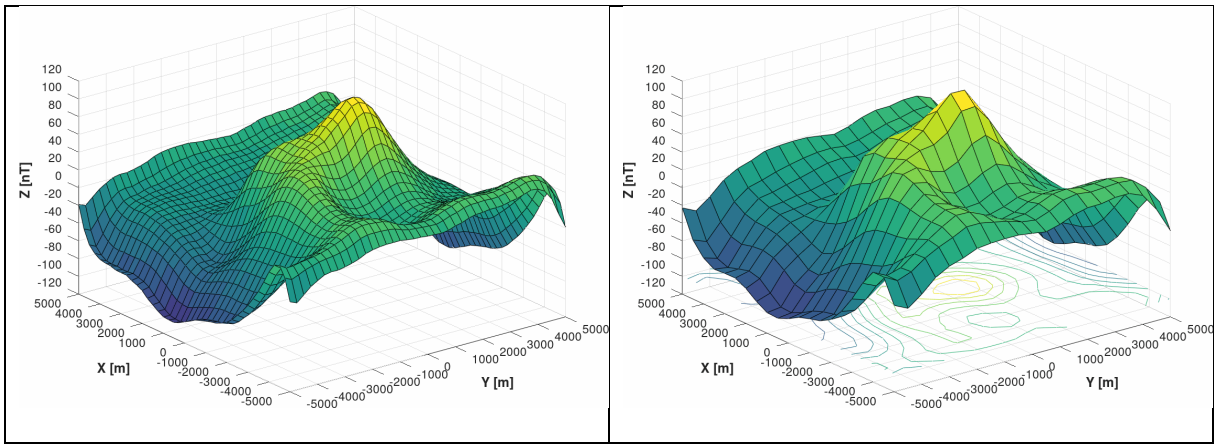

Figure 91: Surface maps representing the reduction of the magnetic anomaly to the magnetic pole for the regular grids with sampling intervals of 250 m (left) and 500 m (right). The filtering operation was executed on the 2D HLSQFT of the anomaly.

DOI: 10.14750/ME.2020.008

6.3 A Field Example with Non-Equidistant Sampling

As the 2D H-LSQ-FT proved its success for 2D magnetic datasets with a regular grid of measuring stations, there is every reason to suppose that the algorithm is also effective when the measuring stations are located irregularly in the field. The regular grids shown in Figure 85 were distorted with random numbers using linear interpolation to produce the irregular distributions of measuring stations of Figure 92. The approximated magnetic anomaly based on these sets of points is represented on the contour maps of Figure 93 and the surface maps of Figure 94. It is worth comparing these figures with Figures 86 and 87 which show the anomaly over regular grids. The effect of point distribution and density on the quality of anomaly maps (visual representation of input datasets) can be observed in such a way.

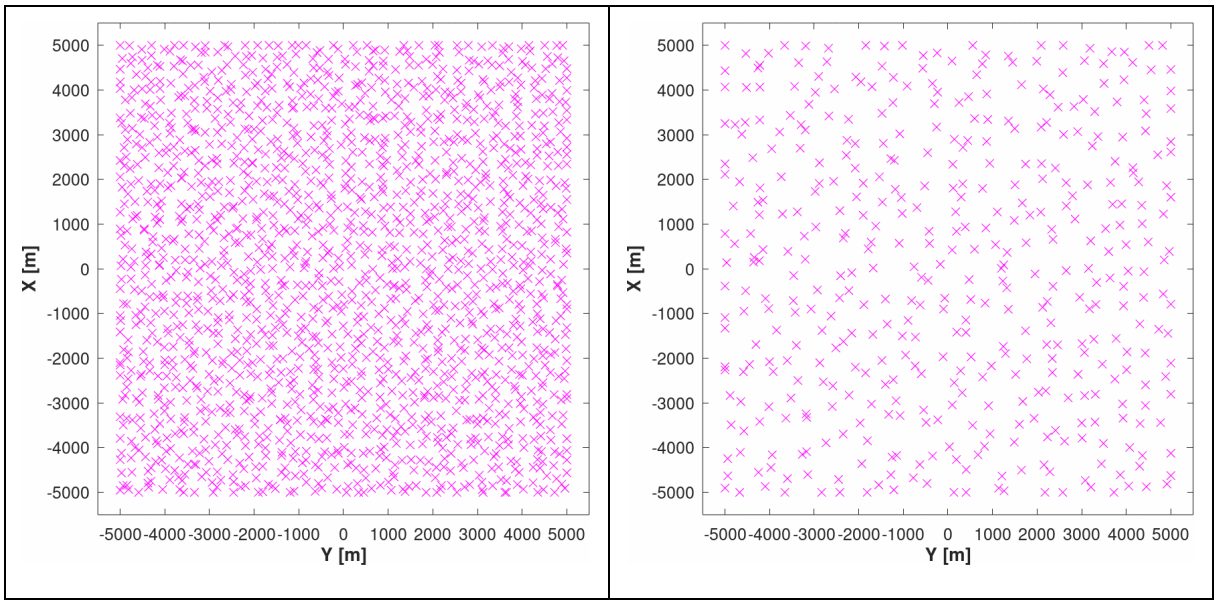

Figure 92: Irregular distributions of measuring stations derived from the regular grids shown in Fig. 10. by adding random values to the coordinates.

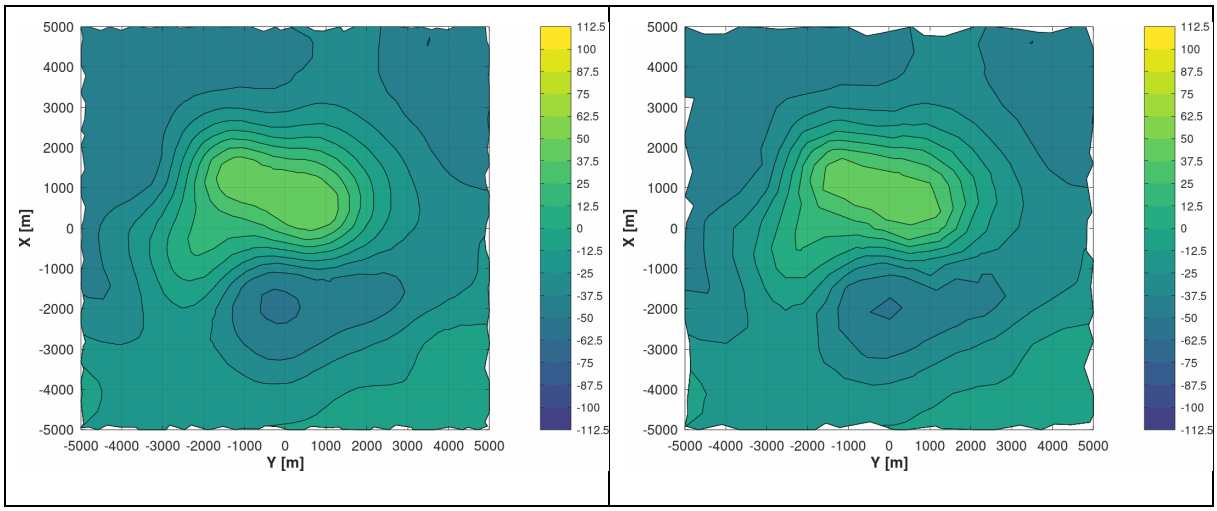

Figure 93: Contour maps of the approximated magnetic anomaly based on the irregular distributions of measuring stations shown in Fig. 17. The unit of isodynamic lines is nT.

DOI: 10.14750/ME.2020.008

Since the 2D H-LSQ-FT algorithm does not require an equal-spaced grid of data, it is able to provide a fine approximation of the Fourier spectrum even for irregular grids. The filtering operation was executed using the 2D H-LSQ-FT on these input data sets. The reduction of the magnetic anomaly to the magnetic pole coming from the filtered spectrum is shown for the magnetic anomaly sampled at two different irregular grids on the contour maps with Figure 95 and the that of the surface maps (Figure 96). After comparing these maps with those of Figures 90 and 91, one can easily state that the distortion of regular grids did not cause significant worsening in the qualities of the results. The demonstrated maps prove the effectiveness of the 2D H-LSQ-FT in the execution of the reduction of the magnetic anomaly to the magnetic pole for both regularly and irregularly sampled anomalies.

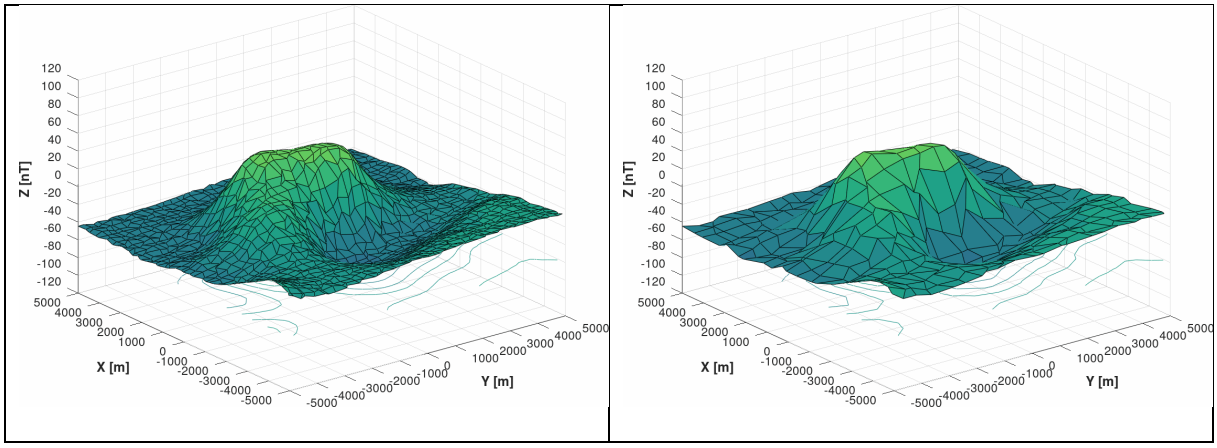

Figure 94: Surface maps of the approximated magnetic anomaly based on the irregular distributions of measuring stations shown in Fig. 17.

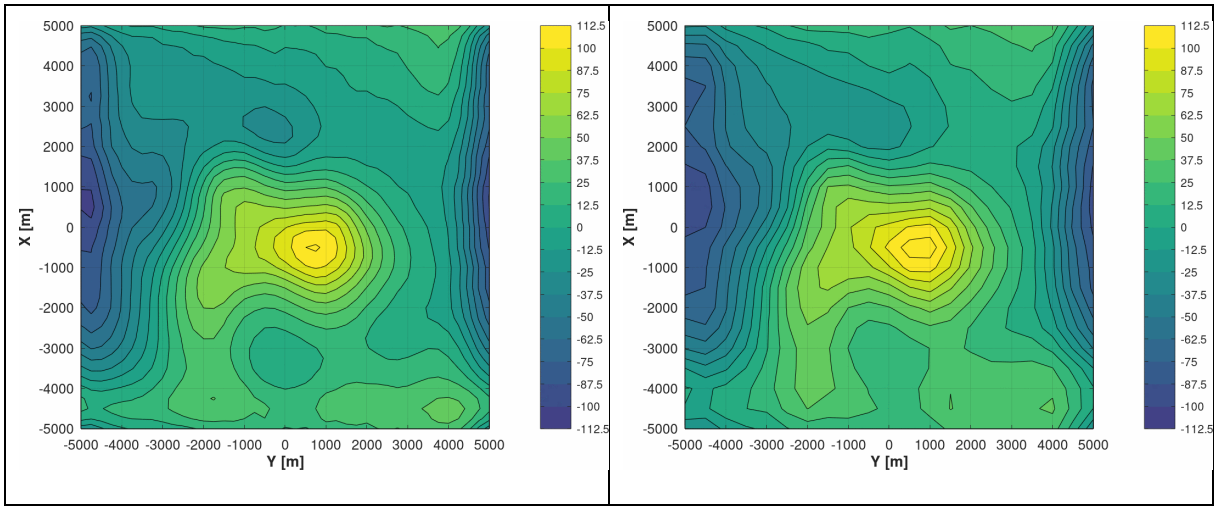

Figure 95: Contour maps representing the reduction of the magnetic anomaly to the magnetic pole for the irregular grids of measuring stations shown in Fig. 17. The unit of isodynamic lines is nT. The filtering operation was executed on the 2D HLSQFT of the anomaly.

DOI: 10.14750/ME.2020.008

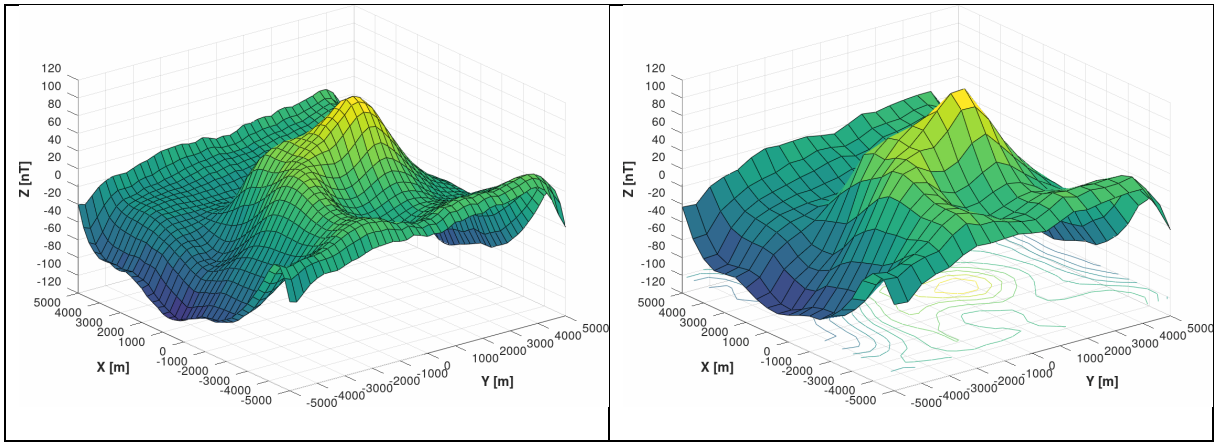

Figure 96: Surface maps representing the reduction of the magnetic anomaly to the magnetic pole for the irregular grids of measuring stations shown in Fig. 17. The filtering operation was executed on the 2D HLSQFT of the anomaly.

DOI: 10.14750/ME.2020.008

Chapter 7

Summary

Data and information processing is an essential aspect of geophysics, hence the continual need to develop methods of scientific value to enhance data interpretation. An ideal processing method should have the ability to reduce substantially or eliminate completely data noise and recover important information from randomly occurring outliers. Fourier transformation is a processing tool used in continuous and discrete data processing in geophysics, but unfortunately, incorporate data noise in its transformational process. Though effective in processing certain data types, the traditional methods are inefficient when input data contains randomly occurring data noise and outliers, hence the need for a more stable, robust and less noise-sensitive method to be developed. In the field of inverse problem theory, a variety of numerous procedures are available for noise rejection, therefore, Dobroka et al, 2012 formulated the Fourier transformation as an overdetermined inverse problem using Hermite functions as basis functions of discretization (the H-LSQ-FT and H-IRLS-FT) which proved to have a higher noise reduction capability even in the face of randomly occurring outliers than the traditional Discrete Fourier Transform (DFT). The discretization of the Fourier spectrum required the numerical scaling of the Hermite functions used as basis functions of discretization. Unfortunately, the values of the scale parameters were inserted manually into the H-LSQ-FT and H-IRLS-FT algorithms from practical experience based on one's own discretion, hence, problematic. This thesis seeks to:

- Develop a new inversion-based Fourier transformation method using Legendre polynomials as basis functions of discretization thereby eliminating the scaling parameter component of the algorithm associated with the use of scaled Hermite functions.
- Improve the existing H-LSQ-FT and the H-IRLS-FT algorithms by introducing a meta-algorithm to optimize the scale parameter thereby eliminating the human component.
- Practically, demonstrate the applicability of the series expansion based inverse Fourier transform in processing non-equidistantly (randomly) acquired geophysical data.

A new 1D inversion based Fourier transformation method known as the Legendre-Polynomials Least-Squares Fourier Transformation (L-LSQ-FT) and the Legendre-Polynomials Iteratively Reweighted Least-Squares Fourier Transformation (L-IRLS-FT) is presented. The introduced L-LSQ-FT method treats the Fourier transformation as an overdetermined inverse problem. The spectrum is discretized by series expansion using Legendre polynomials as basis function and

the inversion problem is solved for the series expansion coefficients by the LSQ method using Steiner weights. We, therefore, introduce the new Jacobi matrix as

$$G_{k,n} = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} P_n(\omega) e^{j\omega t_k} d\omega$$

or in a more formal notation

$$G_{k,n} = \mathrm{F}_k^{-1} \left\{ P_n(\omega) \right\}.$$

As a tool for calculating the inverse Fourier transform the FFT algorithm has been used. The practice of geophysical inversion shows that the least square solutions are very sensitive to sparsely distributed large errors, thus, outliers in the data set and the estimated model parameters may even be completely non-physical. A more flexible method was defined by modifying the weights with the help of Steiner's Most Frequent Value method. The Fourier transformation was handled as a robust inverse problem using the IRLS algorithm with Cauchy-Steiner weights to give the L-IRLS-FT method. The discretization of the continuous Fourier spectra is given by a series expansion with the Legendre polynomials as basis functions. The expansion coefficients are determined by solving an overdetermined inverse problem. The traditional DFT, L-LSQ-FT and the L-IRLS-FT were tested numerically using a Morlet waveform in the presence of both Gaussian and Cauchy noise. The results fully demonstrate the reduced outlier and random noise sensitivity of the newly developed L-LSQ-FT and L-IRLS-FT methods compared to the traditional DFT. It was further concluded that the newly developed L-LSQ-FT and L-IRLS-FT can be considered as a better alternative to the traditional DFT but complementary support to the original H-LSQ-FT and H-IRLS-FT methods.

After successful application in 1D, the concept of using Legendre polynomials as basis functions of discretization was extended to the 2D domain where a new 2D inversion based Fourier transformation method known as the 2D Legendre-Polynomials Least-Squares Fourier Transformation (2D L-LSQ-FT) and the 2D Legendre-Polynomials Iteratively Reweighted Least-Squares Fourier Transformation (2D L-IRLS-FT) was developed. The introduced 2D L-LSQ-FT method also treated the Fourier transformation as an overdetermined inverse problem. The spectrum is discretized by series expansion using Legendre polynomials as basis function and the inversion problem is solved for the series expansion coefficients by the 2D LSQ method using Steiner weights. We, therefore, introduce the new Jacobi matrix as

$$G_{nm}^{kl} = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} P_n(\omega_x) e^{j\omega_x x_k} d\omega_x \; . \; \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} P_m(\omega_y) e^{j\omega_y y_l} d\omega_y$$

or in a more formal notation

$$G_{n,m}^{kl} = \mathcal{F}_k^{-1}\{P_n(\omega)\}.\mathcal{F}_l^{-1}\{P_m(\omega)\}$$

The Fourier transformation was additionally handled as a robust inverse problem using the 2D IRLS algorithm with Cauchy-Steiner weights to give the 2D L-IRLS-FT method. The discretization of the continuous Fourier spectra is given by a series expansion with the Legendre polynomials functions as basis functions. The traditional 2D DFT, 2D L-LSQ-FT, and the 2D L-IRLS-FT were tested numerically using a generated 2D test surface in the presence of both Gaussian and Cauchy noise. The result showed that newly developed 2D L-LSQ-FT and the 2D L-IRLS-FT were more efficient in eliminating data outlier and randomly occurring noise compared to the traditional 2D DFT method. Comparatively, the newly developed 2D L-LSQ-FT and 2D L-IRLS-FT were better alternatives to the traditional DFT but complementary support to the original 2D H-LSQ-FT and 2D H-IRLS-FT methods.

To develop a meta-algorithm to optimize the scale parameter used in the H-LSQ-FT Method, the Simulated Annealing Method which is a Global Optimization Technique is applied. An objective function was generated in the form of an energy function to be optimized. The SA program unsystematically creates a scaling frequency $'f_0'$ and energy $'Ene\_min'$ to be minimized in each step of the iteration. The inversion based Fourier transform is called into the algorithm in each iteration to re-calculate the energy function for the cycle to continue. The algorithm accepts a new solution if the estimated energy is less than the condition number 'one', otherwise the acceptance is based on the Metropolis acceptance condition. The input temperature declines continually for each iteration until the stop criteria are met and the algorithm proceeds to the optimal solution. This procedure uses the simulated annealing technique to optimize the alpha parameter from the real statistics of the input data or output spectrum or a combination of both, hence, eliminating the human error component associated with defining the alpha parameter. The numerical testing of the meta-algorithm in the noiseless and noisy case demonstrated that it is enough to use data distance in case of noisy data to optimize the alpha parameter as the algorithm gave similar output in the data space, spectrum space or a combination of both. With the above alpha optimizing algorithm, it is certain that the H-LSQ-FT Method will exhibit a significant improvement in its noise reduction capabilities.

There is an important new feature of the inversion-based Fourier Transformation method (H-LSQ-FT and 2D H-LSQ-FT), in that, the measurement array or field survey should not be at equal intervals. With the advancement in survey equipment that incorporates the Global Positioning System, a random field survey has been made possible. We tested the H-

DOI: 10.14750/ME.2020.008

LSQ-FT and the 2D H-LSQ-FT on synthetic and field data sampled at equidistant and non-equidistant intervals. The out maps showed the efficiency of the inversion-based Fourier Transformation method in processing randomly acquired data. This new feature will go a long way to ease geophysical field surveys as field measurements are not necessary to be taken on an equal grid, reducing the ensuing cost of geodetic related works in planning a geophysical field survey. In this thesis, it has been adequately demonstrated that the inversion-based Fourier transformation algorithm can be effectively used in processing data set collected in non-equidistant (even in random walk) measurement geometry synthetic and field data.

DOI: 10.14750/ME.2020.008

Acknowledgment

My greatest gratitude goes to the Almighty God, whose mercies, guidance and protection have made it possible for me to accomplish this work.

I wish to express my sincerest thanks to my main supervisor, Prof. Dr. Mihaly Dobroka for his support, guidance, and contribution towards the successful completion of this thesis. Without his guidance and constant feedback, this success would not have been achievable. He has undoubtedly thought me a great deal about both scientific research and life in general. He showed me by his example, what a good researcher and person should be. Thanks for teaching me to be kind and good to the people I encounter in this life. I really enjoyed every minute I was your student.

A very special thanks to my co-supervisor Assoc. Prof. Peter Vass for his invaluable advice and contributions to my research and for always being supportive of my work. I would like to thank the entire faculty members of the Department of Geophysics, the University of Miskolc for the love and support shown me during my stay at the department. Especially, Prof. Norbert Szabo and Dr. Anett Kiss for the personal advice offered me at different stages of my work. To the PhD students, I appreciate the times we spent together discussing issues of academic importance. My sincere gratitude to my mate, Armand Abordan for his timely assistance as we progressed together.

I also wish to express my heartfelt thanks to my mother, Madam Agnes Ebbah, my uncle's, Mr. Paul Kofi Ebbah and Mr. Michael Kojo Ebbah for their financial support, advice, and dedication to my education. Special thanks to my siblings, especially Gifty Nuamah and the entire Ebbah-Nyamekye family. To my friends, P. Boakye, F. Addae and K. O. Yeboah, I say God bless you for being there for me throughout my life.

The research was carried out in the framework of the GINOP-2.3.2-15-2016- 00010 "Development of enhanced engineering methods with the aim at utilization of subterranean energy resources" project of the Research Institute of Applied Earth Sciences of the University of Miskolc in the framework of the Széchenyi 2020 Plan, funded by the European Union, co-financed by the European Structural and Investment Funds.

NUAMAH DANIEL ODURO BOATEY

REFERENCES

1. Aggarwal C.C (2013). Outlier analysis. Springer, New York.
2. Al-Dossary S. and Marfurt K.J. (2007). Lineament-preserving filtering. *Geophysics* 72, P1–P8.
3. Amundsen L. (1991), Comparison of the least-squares criterion and the Cauchy criterion in frequency-wavenumber inversion. Geophysics. ;56:2027-2038.
4. Askari R. and Siahkoohi H.R. (2008). Ground-roll attenuation using the s- and x-f-k-transforms. *Geophysical Prospecting* 56, 105–11
5. Bagaini, C. (2010) Acquisition and processing of simultaneous vibrosis data, Geophysical prospecting, 58(1), 81–99.
6. Baranov, V., (1957), A new method for interpretation of aeromagnetic maps pseudo gravimetric anomalies: Geophysics, 22, 359–383.
7. Baranov, V., and H. Naudy, (1964), Numerical calculation of the formula of reduction to the magnetic pole: Geophysics, 29, 67–79.
8. Barhen, J. Berryman, J.G., Borcea, J. Dennis, C. de Groot-Redlin, F. Gilbert, P. Gill, M. Reinkenschloss, 1. Johnson, T. McEvilly, J. More, G. Newman, D. Oldenburg, P. Parker, B. Porto, M. Sen, V. Torczon, D. Vasco, and N.B. Woodward (2000). Optimization and Geophysical Inverse Problems. Ernest Orlando Lawrence Berkeley National Laboratory, Earth Science Division, LBNL-46959, Pp 1-37
9. Barnett V, Lewis T. (1994). Outliers in statistical data, 3rd edn. Wiley, Chichester
10. Berryman, J. G., (2000). Analysis of approximate inverses in tomography 1. Resolution analysis of common inverses, *Optimization and Engineering*, 1, 87-115.
11. Boschetti, F., M. C. Dentith, And R. D. (1996). List, Inversion of Seismic Refraction Data Using Genetic Algorithms, *Geophysics*, 61,1715-1727.
12. Bouguer anomaly map of Hungary: https://map.mbfsz.gov.hu/gravitacios_anomalia/
13. Butler K.E and Russel R.D. (1993), Subtraction of powerline harmonics from geophysical records. Geophysics 58(6), pp898-903

14. Butler, K. E. and Russell, R. D. (2003). Cancellation of multiple harmonic noise series in geophysical records, Geophysics, 68(3), 1083–1090.

15. Deighan A.J. and Watts D.R. (1997). Ground-roll suppression using the wavelet transform. *Geophysics* 62, 1896–1903.

16. Delay F, Ackerer P, Danquigny C (2005) Simulating solute transport in porous or fractured formations using random walk particle tracking: A review. Vadose Zone Journal 4(2):360-379, DOI 10.2136/vzj2004.0125

17. Dennis, J. E., And R. B. Schnabel, (1996). *Numerical Methods for Unconstrained Optimization and Nonlinear Equations*, SIAM, Philadelphia, 378 Pp.

18. Dobróka M , H. Szegedi, P. Vass (2017) Inversion-based Fourier transform as a new tool for noise rejection. Fourier Transforms: High-tech Application and Current Trends, 1

19. Dobróka M, Gyulai A, Ormos T, Csókás J, Dresen L. (1991), Joint inversion of seismic and geoelectric data recorded in an under-ground coal mine. Geophysical Prospecting. ;39:643-665.

20. Dobróka M, Szabó N (2011) Interval inversion of well-logging data for objective determination of textural parameters Acta Geophysica 59 (5), 907-934

21. Dobróka M, Szabó NP, Tóth J, Vass P (2016) Interval inversion approach for an improved interpretation of well logs. Geophysics 81: D155–D167

22. Dobroka M, Szegedi H, and Vass P. (2017) Inversion-based Fourier transform as a new tool for noise rejection, INTECH 2017, DOI: 10.5772/66338

23. Dobróka M, Szegedi H, Vass P, Turai E. (2012), Fourier transformation as inverse problem - an improved algorithm. Acta Geodaetica et Geophysica Hungarica. ;47(2):185-196. DOI: 10.1556/AGeod.47.2012.2.7

24. Dobroka M. and Vass P. (2006). Fourier transform as an Inverse Problem**.**, In: Szerk.: EAGE Business Office Extended Abstracts of the 64th EAGE Conference & Exhibition: Towards Integration of Geosciences. Houten: European Association of Geoscientists and Engineers (EAGE). pp. P069/1-4.

25. Dobroka M. and Volgyesi L. (2010). Series expansion based inversion IV. Inversion reconstruction of the gravity potential (in Hungarian). *Magyar Geofizika*, 51, 143–149.

26. Dobróka, M., H. Szegedi, J. Somogyi Molnár, And P. Szűcs, (2015), On the Reduced Noise Sensitivity of a New Fourier Transformation Algorithm: Mathematical Geosciences, 47, 679–697, Doi: 10.1007/ S11004-014-9570-X.

27. Dobróka, M., Prácser, E., Turai, E. and Kavanda, R. (2013). Quick imaging of MT data using an approximate inversion algorithm. Acta Geod Geophys 48: 17. https://doi.org/10.1007/s40328-012-0011-3

28. Everett, M. E., And A. Schultz, (1993). Two-Dimensional Nonlinear Magnetotelluric Inversion Using a Genetic Algorithm, *J. Geomag. Geoelectr.*, 45, 1013-1026.

29. Fleury M, Bauer D, Neel M (2015) Modeling of super-dispersion in unsaturated porous media using NMR propagators. Microporous and Mesoporous Materials 205:75-78

30. Goldberg, D. E. (1989). *Genetic Algorithms in Search, Optimization and Machine Learning*, Addison-Wesley, Reading, MA.

31. Goswami, J.C., Mydur, R., Wu, P., Heliot, D., (2004). A Robust Technique for Well-Log Data Inversion. Ieee Transactions on Antennas and Propagation 52, 717-724.

32. Guillon V, Fleury M, Bauer D, Neel MC (2013) Super dispersion in homogeneous unsaturated porous media using NMR propagators. Physical Review E 87(4), DOI 10.1103/PhysRevE.87.043007

33. Guitton, A. (2005). Multiple attenuation in complex geology with a pattern-based approach, Geophysics, 70(4), V97–V107.

34. Guitton, A. and Symes, W. (2003): Robust inversion of seismic data using the Huber norm, Geophysics, 68(4), 1310–1319.

35. Gunn, P. J., (1972), Application of Wiener filters to transformations of gravity and magnetic fields: Geophysical Prospecting, 20, 860–871.

36. Gunn, P. J., (1975), linear transformations of gravity and magnetic fields: Geophysical Prospecting, 23, 300–312.

37. Gyulai, Á., Ormos, T. and M. Dobróka, (2010), A quick 2-D geoelectric inversion method using series expansion: Journal of Applied Geophysics, 72, 232–241

38. Haines, S., Guitton, A., and Biondi, B. (2007), Seismo-electric data processing for surface surveys of shallow targets, Geophysics, 72(2), G1–G8.

39. Holland J.H (1975) Adaptation in natural and artificial systems. University of Michigan Press, Ann Arbor

40. Hoteit H, Mose R, Younes A, Lehmann F, Ackerer P (2002). Three-dimensional modeling of mass transfer in porous media using the mixed hybrid finite elements and the random-walk methods. Mathematical Geology 34(4):435{456, DOI 10.1023/A:1015083111971

41. Jeffryes, P. (2002). A method of seismic surveying with overlapping shot times, US Patent 7,050,356.

42. Jeng, Y., Li, Y.-W., Chen, C.-S., and Chien, H.-Y. (2009). Adaptive filtering of random noise in near-surface seismic and ground-penetrating radar data, J. Appl. Geophys., 68, 36–46, doi: 10.1016/j.jappgeo.2008.08.013.

43. Keating, P., and L. Zerbo, (1996), An improved technique for reduction to the pole at low latitudes: Geophysics, 61, 131–137.

44. Kennett, B. L. N., And M. S. Sambridge, (1992). Earthquake Location - Genetic Algorithms for Teleseisms, *Phys. Earth Planet. Interiors,* 75, 103-110.

45. Kim IC, Torquato S (1992). Effective conductivity of suspensions of overlapping spheres. Journal of Applied Physics 71(6):2727-2735, DOI 10.1063/1.351046

46. Kinzelbach W (1988). The random walk method in pollutant transport simulation. In: Groundwater flow and quality modelling, Springer, pp 227-245.

47. Kinzelbach W, Uffink G (1991). The random walk method and extensions in groundwater modelling, vol; Transport Processes in Porous Media. Springer, Netherlands

48. Kirkpatrick, S., C. D. Gelatt, Jr., And M. P. Vecchi, (1983). Optimization by Simulated Annealing, *Science*, 220, 671-680.

49. Kiss Karoly I. (1981). Magnetic methods of applied Geophysics. ISBN 978-963-284-057-4

50. Kunaratnam K. (1981). Simplified expressions for the magnetic anomalies due to vertical rectangular prisms. Geophysical Prospecting. ;29:883-890.

51. Lee, W. H. K., And S. W. Stewart, (1981). *Principles and Applications of Microearthquake Networks,* Academic Press, New York, 293 Pp.

52. Leu, L., (1982), Use of reduction-to-the-equator process for magnetic data interpretation: Geophysics, 47, 445

53. Liu C., Liu Y., Yang B., Wang D. and Sun J. (2006). A 2D multistage median filter to reduce random seismic noise. *Geophysics* 71, V105–V110.

54. Liu Y., Liu C. and Wang D. (2009). A 1D time-varying median filter for seismic random, spike-like noise elimination. *Geophysics* 74, V17–V24.

55. Luo, Y., Xue, D. J., & Wang, M. (2010). Reduction to the Pole at the Geomagnetic Equator. *Chinese Journal of Geophysics, 53*(6), 1082-1089.

56. Magnetic anomaly map of Hungary: https://map.mbfsz.gov.hu/magneses_anomalia/

57. Map of Pre-Cainozoic basement of Hungary: https://map.mbfsz.gov.hu/preterc500/

58. Mendonca, C. A., and J. B. C. Silva, (1993), A stable truncated series approximation of the reduction-to-the-pole operator: Geophysics, 58, 1084–1090

59. Menke W, (1984). Geophysical Data Analysis – Discrete Inverse Theory. Academic Press, Inc. London Ltd.

60. Menke, W., (1989). Geophysical Data Analysis – Discrete Inverse Theory. Academic Press, San Diego

61. Metropolis N, Rosenbluth A, Rosenbluth M, Teller A, Teller E, (1953). Equation of State Calculations by Fast Computing Machines. J. Chem. Phys. 21, Pp. 1087-1092.

62. Meunier, J. and Bianchi, T. (2002). Harmonic noise reduction opens the way for array size reduction in vibroseis operations, 72nd Annual International Meeting, SEG, Expanded Abstr., 70–73.

63. Milligan, P., & Gunn, P. (1997). Enhancement and presentation of airborne geophysical data. *AGSO Journal of Australian Geology & Geophysics, 17*(2), 63-75

64. Mosegaard, K., (1998). Resolution Analysis of General Inverse Problems Through Inverse Monte Carlo Sampling, *Inverse Problems*, 14, 405-426.

65. Mosegaard, K., And A. Tarantola, (1995). Monte-Carlo Sampling of Solutions to Inverse Problems, *J. Geophys. Res.*, 100, 12431-12447.

66. Mulla, D. J., and Bhatti, A. U. (1997). An evaluation of indicator properties affecting spatial patterns in N and P requirements for winter wheat yield. Spatial Variability in Soil and Crop. In J. V. Stafford (Ed.), *Precision Agriculture '97: Spatial Variability in Soil and Crop* (pp. 145-154). Oxford, UK: BIOS Sci.

67. Nabighian M. N., Grauch V. J. S., Hansen R. O., LaFehr T. R., Li Y., Peirce J. W., Phillips J. D., and Ruder M. E. (2005). The historical development of the magnetic method in exploration. Geophysics, Vol. 70, No. 6

68. Neel M.C, Rakotonasyl S.H, Bauer D, Joelson M, Fleury M (2011). All order moments and other functional of the increments of some non-markovian processes. Journal of Statistical Mechanics: Theory and Experiment DOI 10.1088/1742-5468/2011/02/P02006

69. Newman, G. A., And D. L. Alumbaugh, (1997). Three-Dimensional Massively Parallel Electromagnetic Inversion - 1. Theory, *Geophys. J. Int.*, 128, 345~354.

70. Nolet, G., (1985). Solving or Resolving Inadequate and Noisy Tomographic Systems, *J. Compo Phys.*, 61, 463~482.

71. Nolte, B., And L. N. Frazer (1994)., Vertical Seismic Profile Mverslon with Genetic Algorithms, *Geophys. J. Int.*, 117,162-178.

72. Nyman, D. C. and Gaiser, J. E. (1983). Adaptive rejection of highline contamination: 53rd Annual International Meeting, SEG, Expanded Abstr., 321–323.

73. Oldenburg Et Al., (1998). Nanoengineering of Optical Resonance. Chemical Physics Letters 288(1998)243-247, Elsevier.

74. Osborne J.W, Overbay A. (2004). The power of outliers (and why researchers should always check for them). Pract Assess Res Eval 9(6):1–12

75. Paige, C. *C.*, And M. A. Saunders, (1982). LSQR: An Algorithm for Sparse Linear Equations and Sparse Least Squares, *ACM Trans. Math. Softw.*, 8, 43-71.

76. Parasnis D. S. (1986). Principles of Applied Geophysics, 4th ed. x+402 pp. London, New York: Chapman & Hall. ISBN 0 412 28320 4.

77. Parker, P. B., (1999). *Genetic Algorithms and Their Use in Geophysical Pmblems,* Ph.D. Thesis, University of California, Berkeley, 202 Pp.

78. Parker, R.L., (1994). Geophysical Inverse Theory. Princeton University Press, Princeton, Nj.

79. Pearson, W. C., and C.M. Skinner, (1982), Reduction-to-the-pole of low latitude magnetic anomalies: 52nd Annual International Meeting,SEG, Expanded Abstracts, 356.

80. Pelikán P., Budai T. (Szerk.), Lessgy., Kovács S., Pentelényi L., Sásdi L. (2005) Geology of the Bükk Mountains, Budapest 2005, ISBN 963 671 2530

81. Rothman, D. H. (1986). Automatic Estimation of Large Residual Statics Correction, *Geophysics*, 51, 332-346.

82. Rothman, D. H., (1985). Nonlinear Inversion, Statistical Mechanics, And Residual Statics Estimation, *Geophysics*, 50, 2784-2796.

83. Sahimi M (2011). Flow and transport in porous media and fractured rock: from classical methods to modern approaches. John Wiley & Sons.

84. Sambridge, M., And G. Drijkoningen, (1992). Genetic Algorithms in Seismic Waveform Inversion, *Geophys, J. Int.*, 109, 323-342.

85. Sambridge, M., And K. Gallagher, (1993). Earthquake Hypocenter Location Using Genetic Algorithms, *Bull. Seism. Soc. Am.*, 83, 1467-1491.

86. Saucier, A., Marchant, M., and Chouteau, M. (2006). A fast and accurate frequency estimation method for canceling harmonic noise in geophysical records, Geophysics, 71(1), V7–V18.

87. Scales J. A, Gersztenkorn A, Treitel S. (1988), Fast Lp solution of large, sparse, linear systems: Application to seismic travel time tomography. Journal of Computational Physics. ;75:314-333.

88. Scales, J. A., M. L. Smith, And T. L. Fischer, (1992). Global Optimization Methods for Multimodal Inverse Problems, *J. Comput. Phys.*, 103, 258-268.

89. Sen, M. *K.,* And P. 1. Stoffa, (1992). Rapid Sampling of Model Space Using Genetic Algorithms: Examples from Seismic Waveform Inversion, *Geophys, J. Int.*, 108, 281-292.

90. Sen, M. K., And P. L. Stoffa, (1995). *Global Optimization Methods in Geophysical Inversion*, Elsevier, Amsterdam, 281 Pp.

91. Sen, M.K., Bhattacharya, B.B., Stoffa, P.L., (1993). Non-linear inversion of resistivity sounding data. Geophysics 58,496–507

92. Steiner F. (1988), Most Frequent Value procedures (a short monograf). Geophysical Transactions. ;34:139-260.

93. Steiner F. (1997), Optimum methods in statistics. Budapest: Academic Press .

94. Stoffa, P. L., And M. K. Sen, (1991). Nonlinear Multiparameter Optimization Using Genetic Algorithms: Inversion of Plane-Wave Seismograms, *Geophysics,* 56, 1794-1810.

95. Szabó Norbert Péter, Dobróka Mihály (2019). Series Expansion-Based Genetic Inversion of Wireline Logging Data, MATHEMATICAL GEOSCIENCES 51: (6) pp. 811-835.

96. Szabó N. P, Balogh G. P, Stickel J (2018). Most frequent value-based factor analysis of direct-push logging data. Geophys Prospect 66:530–548

97. Szabó, N. P., (2004). Global Inversion of Well Log Data. Geophysical Transactions 44, 313-329.

98. Szegedi H, Dobróka M. (2014), On the use of Steiner's weights in inversion-based Fourier transformation: robustification of a previously published algorithm. Acta Geophysica. ;49(1):95-104. DOI: 10.1007/s40328-014-0041-0

99. Szűcs P, Civan F, Virág M. (2006), Applicability of the Most Frequent Value method in groundwater modelling. Hydrogeology Journal. ;14:31-43. DOI: 10.1007/s10040-004-0426-1

100. Szűcs, P., F. Civan, (1996), multi-layer well log interpretation using the stimulated annealing method. Journal of Petroleum Science and Engineering,14, pp209-220

101. Takács et al., (1993). Bevezetés az alkalmazott geofizikába (Introduction to Applied Geophysics), Nemzeti Tankönyvkiadó, 1993

102. Turai E, Dobroka M, Herczeg A. (2010). Series expansion based inversion III. Procedure for inversion processing of induced polarization (IP) data (in Hungarian). *Magyar Geofizika*, 51, 88–98.

103. Turai E. and M. Dobróka (2001) A New Method for the Interpretation of Induced Polarization Data - the TAU-Transform Approach Conference Proceedings, 63rd EAGE Conference & Exhibition, Jun 2001, cp-15 DOI: https://doi.org/10.3997/2214-4609-pdb.15.P049

104. Zhou, C., Gao, C., Jin, Z., Wu, X., (1992). A Simulated Annealing Approach to Constrained Nonlinear Optimization of Formation Parameters in Quantitative Log Evaluation. SPE Annual Technical Conference and Exhibition, Paper 24723, 1-9.

105. Zimmermann S, Koumoutsakos P, Kinzelbach W (2001). Simulation of pollutant transport using a particle method. Journal of Computational Physics 173(1):322{347, DOI:10.1006/jcph.2001.6879