



UNIVERSITÀ DEGLI STUDI DI NAPOLI FEDERICO II  
SCUOLA POLITECNICA E DELLE SCIENZE DI BASE

*PhD School in Earth, Environment and Resources Sciences*

*- XXX Cycle -*

**PhD Thesis**

# **Constraining potential field interpretation by geological data: examples from geophysical mapping, inverse and forward modelling**

*Davide Lo Re*

**Advisor**

*Prof. G. Florio*

**PhD Coordinator**

*Prof. M. Fedi*

**2017**



## ABSTRACT

In this thesis three different strategies in potential field data interpretation were implemented and studied. The strategies are related to map transformation, inversion and forward problem. The thesis aims at obtaining geophysical outputs with geological-like features. These kinds of outputs are a significant key to make it easier the geological interpretation of the geophysical data modelling. In particular, the outputs obtained by the different strategies will tend to highlight different units with distinct boundaries and represented by fairly constant field or physical property values.

A map transformation technique (terracing) is first proposed. It is based on the use of a cluster analysis technique applied to a gravity or pole-reduced magnetic map. The centre values of the clusters and the cluster number are selected by a statistical analysis of the data map. The use of cluster technique breaks the continuous function (potential field map) onto different areas characterized by piecewise constant values (terraces). The homogeneity within each area is preserved and this kind of feature allow an easy computation of an apparent physical property horizontal distribution map, directly comparable with a geological map. Tests on synthetic and real data are shown.

The inversion is treated by applying a strategy made up by three steps. The first and the last steps are inversions with different constraints and associated weights, the second one is conducted by clustering the output of the first smooth inversion. The strategy allows obtaining, in the final step, a volume where the retrieved physical property is classified (by clustering technique) in different volumes of relatively constant values, easily relatable to different geological units. The number of the units, as well as the physical property values associated to each unit, it has to be fixed a priori according to the geological knowledge of the area. Tests on synthetic and real data show that the final obtained models are valid in both geophysical (honoring the data) and geological (understandable relationships among clearly-defined geological units) points of view.

A forward problem solver procedure, based on iterative stochastic process is finally proposed. The solution is represented by surfaces that bound different layers having different physical properties. The anomaly field produced by the surfaces is computed by an algorithm working in a

Fourier domain. According to the Markov chain simulation, at each iteration several surfaces are created and the best one is selected to be a starting model in the next iteration. The best model selection is performed according to the value of a goodness coefficient. A synthetic case is shown, and the final model obtained shows a possible shape of different bodies, with homogeneous physical property distribution, able to produce a field that adequately match an observed anomaly field.

## TABLE OF CONTENT

<b>ABSTRACT</b> .....	<b>I</b>
<b>TABLE OF CONTENT</b> .....	<b>III</b>
<b>LIST OF FIGURES</b> .....	<b>IV</b>
<b>INTRODUCTION</b> .....	<b>1</b>
STRUCTURE OF THE THESIS .....	3
<b>CHAPTER 1 – THEORIES AND BASICS</b> .....	<b>5</b>
1.1. POTENTIAL FIELDS THEORY .....	5
1.1.1. Gravity Method .....	6
1.1.2. Magnetic Method .....	7
1.2. ELEMENTARY INVERSION THEORY .....	11
1.3. CLUSTER ANALYSIS.....	13
1.3.1. The k-Means Clustering .....	15
1.3.2. The fuzzy c-Means Clustering.....	17
1.4. BASICS OF MONTE CARLO MARKOV CHAIN .....	18
<b>CHAPTER 2 – SIMPLE GUIDED CLUSTERING</b> .....	<b>20</b>
2.1. METHOD .....	22
2.1.1. Terracing by clustering methods.....	22
2.1.2. Apparent density/magnetization maps from terraced potential fields .....	25
2.2. APPLICATION ON SYNTHETIC DATA .....	26
2.2.1. The gravity field of two prismatic sources .....	26
2.2.2. The total field of the bishop model .....	32
2.3. APPLICATION ON REAL DATA .....	38
2.4. CONCLUSIONS.....	42
<b>CHAPTER 3 – SHARP-EDGED INVERSION</b> .....	<b>44</b>
3.1. LI AND OLDENBURG ALGORITHM .....	45
3.2. GUIDED FUZZY C-MEANS .....	50
3.3. SHARP-EDGED INVERSION STRATEGY .....	52
3.4. A 2-D REAL CASE (PIANO DI PECORE) .....	56
3.4.1. Geological setting .....	56
3.4.2. Data inversion and results .....	58
3.5. 3D REAL CASE (SICILY CHANNEL).....	63
3.5.1. Geological setting .....	64
3.5.2. Data inversion and results .....	65
3.6. CONCLUSIONS.....	71
<b>CHAPTER 4 - A STOCHASTIC FORWARD PROBLEM SOLVER</b> .....	<b>72</b>
4.1. PARKER ALGORITHM.....	73
4.2. THE STOCHASTIC SOLVER.....	74
4.3. A SYNTHETIC 3D CASE .....	76
4.4. CONCLUSIONS.....	80
<b>CONCLUSIONS AND FUTURE PERSPECTIVES</b> .....	<b>82</b>
<b>ACKNOWLEDGEMENTS</b> .....	<b>84</b>
<b>REFERENCES</b> .....	<b>85</b>

## LIST OF FIGURES

<b>Figure 1</b> - Histogram vs. kernel smoothing function. a) Histogram of the frequency distribution of a Gaussian random noise. b) Kernel smoothing function of the same data, computed using different values of the bandwidth parameter (equation 2.1). The bandwidth value, $h=0.16$ , is found by Equation 2.2. ....	25
<b>Figure 2</b> - Estimation of cluster parameters for terracing by SGC. a) Gravity field generated by two prismatic sources having a square base with sides 13 km long, a depth to the top of 2 km and thickness of 10 km. The leftmost prism has a density contrast with the background of $1 \text{ g/cm}^3$ , and the rightmost one has a density contrast with the background of $-1 \text{ g/cm}^3$ . The grid has 121 rows and 121 columns, and a grid step of 1 km. The original data are normalized between 0 and 1, as illustrated by the second color bar. b) Frequency histogram of the equalized data. c) Kernel smoothing function of the equalized data, showing 4 peaks (highlighted by the red circles) at 0.10, 0.34, 0.66 and 0.90 mGal, corresponding, in the original range, to -115.15, -45.18, 45.19 and 115.17 mGal. The kernel smoothing function was computed using equation 2.3 for the choice of the smoothing parameter $h$ . ....	27
<b>Figure 3</b> - Terracing of gravity data of Figure 2a by different approaches. a) SGC using the 4 clusters and respective centroid values as obtained by inspection of Figure 2c. b) SGC using just 3 clusters, namely the two extreme values and the average of the two intermediate values of the previous case. c) k-means clustering assuming 3 clusters whose centers are automatically found. d) Kuwahara filter, using a 3x3 window and 20 iterations. ....	29
<b>Figure 4</b> - Terracing of gravity data of Figure 2a corrupted by a Gaussian noise with a standard deviation of 0.1 mGal, a zero mean and a maximum amplitude equal to about the 3% of the original anomaly amplitude. a) Kernel smoothing function computed by the automatic selection of the smoothing parameter, $h$ , by equation 2.3. Red circles evidence the five peaks found, corresponding to the five clusters of the successive terracing, centered at the respective values of the gravity field. b) SGC terraced data using the parameters determined in a). c) West-East profile of the terraced field (b) at $y=60 \text{ km}$ . The original field and the sign of its curvature are shown for reference. ....	31
<b>Figure 5</b> - The Bishop model magnetic basement. a) Depth to the top of the magnetic basement. b) Magnetization variations in the Bishop basement. c) Total magnetic field generated by the Bishop synthetic basement. The geomagnetic field and magnetization vectors are vertical. The color bar between 0 and 1 is drawn for reference, and illustrates the correspondence between the normalized data and the original ones. The white line indicates the location of the profile shown in Figure 8. ....	32
<b>Figure 6</b> - Terracing of Bishop total field data. a) Kernel smoothing function of the histogram equalized total field intensity of Figure 5c. The 5 peaks of the curve (highlighted by the red circles) and their respective values are used to set up the SGC clustering process. b) SGC terracing. c) k-means terracing assuming 5 clusters whose centers are automatically found. ....	33
<b>Figure 7</b> - Terracing the Bishop total field by the Kuwahara filter: a) using a 3x3 window and 20 iterations; b) using a 9x9 window and 20 iterations. ....	34
<b>Figure 8</b> - Comparison of terracing of the Bishop magnetic field along a S-N profile at $x = 83.2 \text{ km}$ . a) SGC and k-means clustering methods. b) Kuwahara filters of the Bishop magnetic field with different window sizes. c) Kuwahara filtering of the clustered data. ....	35
<b>Figure 9</b> - Kuwahara-filtered terraced maps. a) SGC-terraced map of Figure 5b transformed by a Kuwahara filter with a window of 9x9 points, iterated 20 times. b) k-means terraced map of Figure 5c transformed by a Kuwahara filter with a window of 9x9 points, iterated 20 times. ....	36
<b>Figure 10</b> - Apparent magnetization computation. a) Apparent magnetization obtained from the SGC-terraced map transformed by a Kuwahara filter (Figure 9a). b) Apparent magnetization obtained from the k-means-terraced map transformed by a Kuwahara filter (Figure 9b). c) Total field computed from the apparent magnetization map in a). d) Total field computed from the apparent magnetization map in b). e) Residuals between the computed total field (c) and the observed total field of Figure 5c. f) Residuals between the computed total field (d) and the observed total field of Figure 5c. ....	36
<b>Figure 11</b> - Simplified geological map of the Noranda-Val-d'Or region (after Keating, 1992). ....	39
<b>Figure 12</b> - Analysis of real gravity data of Noranda-Val-d'Or region (Canada). a) Bouguer anomalies as interpolated on a 1 km regular grid. The projection used is UTM zone 17N for the WGS84 ellipsoid. b) Kernel smoothing function of the	

histogram equalized data. The 6 peaks of the curve (highlighted by the red circles) and their respective values are used to set up the SGC clustering process. c) SGC terracing. d) k-means terracing, assuming 6 clusters whose centers are automatically found..... 40

**Figure 13** - Analysis of real gravity data of Noranda-Val-d'Or region (Canada). a) Apparent density map obtained from the SGC terraced map, overlain on a simplified geological map (after Keating, 1992). b) Apparent density map obtained from the k-means terraced map, overlain on a simplified geological map (after Keating, 1992). c) Gravity field computed from the apparent density map in a). d) Gravity field computed from the apparent density map in b). e) Residuals between the computed gravity field (c) and the observed gravity field (Figure 12a); f). Residuals between the computed gravity (d) and the observed gravity field (Figure 12a). ..... 41

**Figure 14** - b) Model cross-section obtained using Li and Oldenburg algorithm and a) comparison of observed, solid line, and calculated data, red asterisks. b) Black lines are the trace of the c) real source extension of two boxes with a positive density contrast of  $0.5 \text{ g/cm}^3$  respect to the background. Colorbar are in  $\text{g/cm}^3$ . ..... 49

**Figure 15** - a) Density distribution model. b) Defuzzified model obtained by the application of guided fuzzy c-means clustering, using 2 classes with defined centre values. c) comparison of observed data (solid line) and data calculated from the model in b) (dashed line). ..... 51

**Figure 16** Sharp-Edged Inversion strategy workflow. .... 52

**Figure 17** - Sources boundary of the defuzzified model in Figure 15c..... 55

**Figure 18** - a) Final density distribution from the third step of the Sharp-Edged Inversion strategy, and b) comparison of observed, solid line, and calculated data, dashed..... 55

**Figure 19** - Complete Bouguer anomaly. .... 56

**Figure 20** - a) The studied area in Southern Italy is the red asterisk. b) A detail of the area and its geological sketch map with gravity station represented by black squares. Green is Mesozoic limestone, orange is debris slope, light blue is colluvium, and red line is the fault trace, solid when exposed and dashed in the basin, where it is marked by a morphological step on the ground (modified after Improta et al., 2003). c) 3D visualization of the detailed DEM (2 x 2 km). Cell size 5 m. Piano di Pecore basin is in the centre. Bold line contours every 50 m. Projection: Rome40 / Italy zone 2 (EPSG 3004). ..... 58

**Figure 21** - Reference model adopted in the inversion, during the first step of the Sharp-Edged Inversion strategy ..... 59

**Figure 22** - a) comparison of observed, blue line, and calculated data, red line, by the b) density distribution model obtained after the smooth inversion, first step of the Sharp-Edged Inversion strategy. Colorbar is in  $\text{g/cm}^3$ . ..... 60

**Figure 23** - Defuzzified output after the clusterization of the model in Figure 22. The cluster centre are 1.8, 2 and  $2.3 \text{ g/cm}^3$ . ..... 61

**Figure 24** - a) comparison of observed data (dots) and calculated data (dashed line) produced by the Sharp-Edged model (b) obtained in the last step of the inversion procedure..... 62

**Figure 25** - a) The studied area in Western Sicily is the red box. b) Gravity data anomaly map (Carrozzo et al., 1991). Colorbar in mGal. c) Residual Bouguer anomalies, obtained removing a planar trend (SW-NE) to the data b). ..... 63

**Figure 26** - a) Topography and bathymetry, in meters, and position of wells (magenta circles) from ViDEPI projects. b) Overlay of the residual Bouguer gravity anomaly map and the main structural features as given in Meccariello (2017). ..... 64

**Figure 27** - Geological sketch map (modified after Catalano et al., 2002) ..... 65

**Figure 28** – a) XZ cross-section of reference model ( $\mathbf{m}_{ref}$ ) obtained by the linear interpolation of the wells data. Colorbar is in  $\text{g/cm}^3$ , axis in meters. Magenta lines are wells projection. b) Logs of converted densities of two wells present in the model cross-section a) ..... 66

**Figure 29** - East-West cross-sections of the density distribution model obtained after the last step of the Sharp-Edged Inversion strategy. Colorbar in  $\text{g/cm}^3$ . ..... 68

**Figure 30** – Computed and input data after the Sharp-Edged Inversion strategy. a) Data computed by the density distribution model obtained from the third stage of the Sharp-Edged Inversion strategy (Figure 29). Axis in meters. Black dashed lines are the traces of the cross-section of the density distribution model shown in Figure 31 and 32. b) Residual Bouguer anomalies. Axis in metres. .... 68

**Figure 31** - Cross-section of the volume solution along A-A' profile in Figure 30a, compared with Oscar Ovest well log. a) Density distribution cross section. Colorbar units are  $\text{g/cm}^3$ , axis in metres. Magenta line is the well projection. Black lines are faults and thrust drawn according to the structural map in Figure 26b. b) Log of converted densities of Oscar Ovest well, with highlighted the three density classes used in the clustering. The three different homogeneous bodies

are related to: Quaternary to Late Pliocene formations (blue), Miocene formations (cyan) and Mesozoic basement (red). ..... 69

**Figure 32** – Cross-section of the volume solution along B-B' profile in Figure 30a, compared with a geological cross-section. a) Density distribution cross section. Colorbar units are  $\text{g/cm}^3$ , axis in metres. The three different homogeneous bodies are related to: Quaternary to Late Pliocene formations (blue), Miocene formations (cyan) and Mesozoic basement (red). Magenta line is the well projection. Black lines are faults and thrust drawn according to b) geological cross-section (modified after Montanari et al., 2017). ..... 70

**Figure 33** – Synthetic data and correct surface with the starting model and its computed data. a) Reduced to Pole magnetic anomalies map used as observed data during the stochastic process. b) correct surface. c) computed field by the starting model, at  $P_{ij}(1)$  shown in d). Axis in km. White dashed line is the trace of a cross-section. .... 77

**Figure 34** – Solutions relative to the first iteration along the profile in figure 33. a) computed fields. b) solution surfaces with a magnetization contrast of 1 A/m. Green line in b) represents the correct surface. Red lines in b) are the  $P_{ij}(1)$  and its computed field (a). Blue lines in b) are the best GC surface value and its calculated data (a). Grey lines are 200 different solutions. X-Axis are n metres. Cross-section at Y-constant of 25 km, white trace in figure 33. .... 78

**Figure 35** - Solutions relative to the iteration number 15 along the profile in figure 33. a) Comparison of the observed field (green line) and the field generated by the best computed surface ( $P_{ij}(16)$ , blue line in b) with a magnetization contrast of 1 A/m. In b) the green line represents the correct surface. X-Axis are in metres. Cross-section at Y-constant of 25 km, white trace in figure 33. .... 79

**Figure 36** - Solutions relative to the iteration number 15, in map. a) Reduced to Pole magnetic anomalies map used as observed data during the stochastic process. b) the correct surface. c) field computed at iteration number 15. d) best surface solution at iteration 15. e) and f) are the difference between the data and the surfaces, real and computed, respectively. Colorbars in a), c) and e) are in nT. Colorbars in b), d) and f) are in metres. Axis are in km. .... 80

## INTRODUCTION

Generally, in the geological interpretation different units are represented as homogeneous bodies with sharp boundaries, although this separation may not be so abrupt in nature. This kind of representation is used for a better understanding of the geological features and the chronological relationship among different units. In these models, the different units are identifiable from petrographic, lithological or paleontological features. Typical representations of geological interpretation are geological maps, geological cross-sections or geological logs, where different colors suggest different units.

In the same way, geophysical interpretation is related to the variation of some investigated parameter within a model. The considered parameters depend on the data type, e.g., density models are computed from gravity data. Typical geophysical models are anomaly maps, volumes of physical parameters distribution, cross-sections and so on. These models can be obtained by solving a forward or inverse problem. Depending on the type of problem solver, the output model often does not show abrupt variations of physical properties. Thus, to extract information on the geological boundaries at depth from geophysical models, often one must rely on the interpreter's experience and external constraints.

In this thesis, it is presented an investigation of different strategies to obtain geophysical models where the physical property distribution at depth could be related in an unambiguous way to geological features, with the result of reducing the subjectivity of their interpretation. The research is focused on the interpretation of potential fields (gravity and magnetic anomalies). Examples involving a map transformation technique, inverse and forward modelling will be presented.

A potential field map transformation aimed at obtaining potential field anomaly maps showing well defined homogeneous geophysical units separated by with sharp boundaries, is known as 'terracing'. In the last thirty years several authors proposed different terracing strategies and methods (URQUHART AND STRANGWAY, 1985; CORDELL AND McCAFFERTY, 1989; COOPER AND COWAN, 2009; SIMPSON ET AL., 2008; LI, 2016). In this thesis a clustering-based strategy to terrace the anomalies map is proposed. The number of clusters and the values of each cluster center can be selected manually or in an automatic way by studying the distribution of the values itself. From such

terraced maps, a simple inverse method is used to recover apparent density/magnetization maps, quantitatively displaying the lithologic variations in an area.

During the last twenty years, several authors developed methods and strategies to find inversion results that can reproduce geological features to help the interpretation stage (PORTNIAGUINE AND ZHDANOV, 1999; CAMACHO ET AL., 2000; LANE ET AL., 2007; BERRINO AND CAMACHO, 2008; FARQUHARSON ET AL., 2008; LELIÈVRE, 2009; SUN AND LI, 2010; PHILLIPS AND SIMPSON, 2015; SUN AND LI, 2015). In this thesis, by application of a clustering-based strategy, it is explored the possibility to obtain layered sharp-edged models, which are not the usual output of potential field inversions. To do that, it is set up an original workflow for the inversion process, integrating a reference model allowing to efficiently taking into account the available a priori information.

The potential field of a set of different units representative of the subsurface geology and characterized by different physical parameters, can be computed by several forward methods, often used in many commercial software (TALWANI ET AL., 1959; TALWANI AND EWING, 1960; PARKER, 1972; BHATTACHARYA, 1978; DIMRI, 1998). In this thesis, a combination of Markov Chain Monte Carlo simulations with a Fourier domain solver (PARKER, 1972) it is used in a relatively fast iterative process. An interesting feature of this kind of modelling is the absence of a voxel size, because the unknown is the depth to a surface separating different geological units. This kind of study is still not fully worked out in this thesis, but from the study done in this PhD project it is possible to identify the path for future developments.

The above-mentioned presented approaches were tested in 2D and 3D cases of synthetic and real Gravity and Magnetic data under different constraints. The obtained results show the good performance of tested methods and their usability in several contexts for different purposes that can be summarized as:

- The Simple Guided Clustering Terracing (SGC Terracing method) is a fast technique to investigate the relation between the gravity or magnetic data and the geological information about the outcropping formations. It produces a transformed geophysical map that can be directly compared to a geological map (Chapter 2);

- The Sharp-Edged Inversion is a strategy to define confined potential fields sources and to retrieve models of layered systems, useful in geophysical identification of faults (section 3.4) and to have a valid geological/geophysical layered model (section 3.5);
- The Stochastic Forward method is a way to obtain continuous surfaces confining different underground bodies such as different layers (Chapter 4).

A common main feature of all these different interpretation approaches consists in the integration of geological constraints, with the possibility to set a degree of closeness to them according to their reliability. This feature allows to counteract the ambiguity and non-uniqueness property of the potential fields data analysis. The constraints may come from surface geology, wells stratigraphy, seismic sections, other geophysical interpretations etc.

## STRUCTURE OF THE THESIS

The **first chapter** is an introduction about the “tools” used to investigate different problems faced up in the other chapters. There are descriptions about: potential fields theory, basics of inversion, forward modelling, cluster analysis definitions and Monte Carlo simulations rudiments.

Potential field data, especially gravity and magnetic method, are one of the oldest geophysical method and although in recent years other geophysical methods are used in a massive way for their better resolution (such as seismic), potential fields are still a good way of investigating the Earth’s interior because of their lighter costs and their passive nature. One of the challenge and aim of this thesis is to find a way to transform the continuous behavior of the potential fields data (anomaly map) in a function showing abrupt variations. Thanks to the use of cluster analysis approach, in the **second chapter** of this thesis will be discussed a new strategy to obtain anomaly maps in which there are shown only a predetermined number of homogeneous units, representing different potential fields sources. Application of k-means clustering ([MACQUEEN, 1967](#)) and the use of a kernel smoothing function ([BOWMAN AND AZZALINI, 1997](#)) are the basis of the Simple Guided Clustering (SGC) technique, presented in this thesis.

Inversion of potential fields data can provide useful information about oil and ore exploration target or for environmental issues like sinkholes, UXO or archeological research. A very timely

research issue in the inversion studies is to develop strategies capable to integrate geological constraints. In this way, the solution model obtained is at the same time valid under both geological and geophysical point of view. The strategy adopted in this thesis, Sharp-Edged Inversion, and discussed in the **third chapter**, is hardly inspired by a recent work ([SUN AND LI, 2015](#)), having the fuzzy c-means clustering ([BEZDEK, 1981](#)) as a fulcrum. Different scales of investigation can provide models able to represent a layered and sharp variation of related physical parameters (density for gravity inversion, susceptibility for magnetic) useful to estimate geological features as faults systems or bedrock oscillations.

In the **fourth chapter** is discussed the use of iterative Monte Carlo Markov Chain simulations as a core of forward problem solver ([PARKER, 1972](#)). The study of this methodology is still in an early stage and not investigated completely. However, the research done demonstrates its promising results and allows individuating future research directions.

## CHAPTER 1 – Theories and basics

In this chapter the theory at the basis of the “tools” developed in the rest of the thesis is exposed. First, the potential fields theory (gravity and magnetic methods) is described, then some information about the cluster analysis, used in the next two chapters, is given. Finally, elements of the inversion theory, the topic of chapter 3, and the basics of Monte Carlo Markov Chain simulation, topic of chapter 4, are given.

### 1.1. POTENTIAL FIELDS THEORY

A unit of mass or a magnetic dipole experiences forces that can be repulsive or attractive, and these kinds of forces are called forces fields and act in the space at a given time (BLAKELY, 1996). Gravitational field and the Magnetic field of the Earth are both force fields.

If the field  $\mathbf{F}$  having a scalar potential  $\phi$  given by  $\mathbf{F} = \nabla\phi$ , is conservative then the field  $\mathbf{F}$  is named potential field and satisfies the Laplace's equation in the region out of the sources:

$$\nabla^2\phi = 0 \tag{1.1}$$

which states that the sum of the rate of change of the field gradient in three orthogonal directions is zero.

In Cartesian coordinates, Laplace's equation is:

$$\frac{\partial^2\phi}{\partial x^2} + \frac{\partial^2\phi}{\partial y^2} + \frac{\partial^2\phi}{\partial z^2} = 0 \tag{1.2}$$

where  $\phi$  refers to a gravitational or magnetic potential.

Any function that satisfy the Laplace's equation, has continuous, single-valued derivatives and has second derivatives (BLAKELY, 1996). If a function is harmonic in a region  $R$  have its maxima and minima on boundaries of the region. Gravity and magnetic fields, are both potential fields and obey all the physical laws mentioned above.

### 1.1.1. Gravity Method

From *Principia Mathematica* of Sir Isaac Newton (1687) it is possible to define the gravitational acceleration due to a point mass  $m$ :

$$\mathbf{g}(P) = -\gamma(m/r^2)\mathbf{r} \quad (1.3)$$

where  $\mathbf{r}$  is a unity vector that point from mass  $m$  to the observation point  $P$ . This gravitational attraction is a conservative field so it can be expressed as the gradient of a scalar potential

$$\mathbf{g}(P) = \nabla U(P) \quad (1.4)$$

where  $U$  is the Newtonian potential and acceleration  $\mathbf{g}$  is a potential field

$$U(P) = \gamma(m/r) \quad (1.5)$$

The gradient of  $U$  represents the gravity  $\mathbf{g}$ , and the first-order directional derivatives of  $U$  are the components of gravity in the corresponding direction (KEAREY ET AL., 2002) and it is defined as:

$$\mathbf{g} = \nabla U = \frac{\partial U}{\partial x}\mathbf{i} + \frac{\partial U}{\partial y}\mathbf{j} + \frac{\partial U}{\partial z}\mathbf{k} \quad (1.6)$$

where  $\mathbf{i}$ ,  $\mathbf{j}$  and  $\mathbf{k}$  are the unit vectors in the positive direction of  $x$ ,  $y$  and  $z$  axes respectively. Being an harmonic function, at all the points outside of the mass,  $\nabla^2 U = 0$ , but in the space occupied by masses:

$$\nabla^2 U = -4\pi\gamma\rho \quad (1.7)$$

where  $\rho$  is the density of the mass distribution at a given point. Equation 1.7 is the Poisson's equation describing the potential at all points of the mass distribution.

In geophysical exploration, gravimeters measure only the vertical component of the gravity, as given by:

$$g_z = \frac{\partial U}{\partial z}$$

(1.8).

Before the results of a gravity survey can be analysed and interpreted it is necessary to correct for all variation in the Earth's gravitational field which do not result from the differences of density in the underlying rocks.

The observed gravity is the sum of the following components (BLAKELY, 1996):

- Theoretical gravity, referred to the reference ellipsoid,
- Free air effect, due to the elevation above the sea level,
- Bouger slab and terrain effects, the normal mass above the sea,
- Tidal and instrumental drift effects, time-dependent variations,
- Eötvös effect, due to moving platform (airborne and shipborne surveys),
- Isostatic effects, accounting for the effects of masses supporting topographic loads,
- Effect of crust and upper mantle density variations.

Isolating the last quantity is the goal of the gravity reductions.

The mean value of gravity at the Earth's surface is about  $9.8 \text{ ms}^{-1}$ . Variations in gravity caused by density variations in the subsurface are of the order of  $100 \mu\text{m s}^{-1}$ . The *cgs* unit of acceleration due to gravity ( $1 \text{ cm s}^{-1}$ ) is the Gal, in honor of Galileo, and its sub-unit milliGal is common in gravity survey (KEAREY ET AL., 2002).

### 1.1.2. Magnetic Method

The magnetic scalar potential  $V(r)$  of a dipole source whose magnetic moment is  $\mathbf{m}$ , can be written as:

$$V(r) = -\mathbf{m} \cdot \nabla(1/r)$$

(1.9)

where  $r$  is the distance operator.

The magnetic field may also be defined in terms of electric currents. If an electric current  $I$ , is flowing in a loop of radius  $r$ , the magnetic strength at the center of the loop is  $H = I / (2r)$ .

Materials can be magnetized by acquiring the component of magnetization in the presence of an external magnetic field and it is called induced magnetization which is in the same (or reverse) direction of the external magnetic field as:

$$\mathbf{M} = \chi \mathbf{H} \tag{1.10}$$

The constant  $\chi$  in the equation (1.10) is called the magnetic susceptibility.

Susceptibility is a dimensionless quantity but differs in magnitude if expressed in emu or in SI units. Its definition involves the magnetic permeability  $\mu$ :

$$\begin{aligned} \mathbf{B} &= \mu_0(\mathbf{H} + \mathbf{M}) \\ &= \mu_0(\mathbf{H} + \chi \mathbf{H}) \\ &= \mu_0(1 + \chi) \mathbf{H} \\ &= \mu \mathbf{H} \\ \mu &= \mu_0(1 + \chi) \end{aligned} \tag{1.11}$$

The relationship between  $\mathbf{M}$  and  $\mathbf{H}$  is not necessarily linear because the magnetic susceptibility  $\chi$  may vary with the field intensity, may be negative, and may be represented more accurately in some materials as a tensor (BLAKELY, 1996). Susceptibility is a measure of how susceptible a material is to become magnetized (REYNOLDS, 1997). There are many kinds of magnetizations and their understanding is important to understand how the variations of magnetic properties produce the magnetic anomalies (HINZE ET AL., 2013). These properties can be defined as: Diamagnetism, for example, is an inherent property of all matter. In the presence of external magnetic field, the orbital path of the electron rotates in a way that induced magnetization is small and in the opposite sense to the applied field. Consequently, diamagnetic susceptibility is negative.

*Paramagnetism*, is a property of those solids that have atomic magnetic moments because in this substance, the electron shells are incomplete, so the unpaired electrons produce a magnetic field. When it is placed in an external magnetic field, the atomic moments or unpaired electrons

partially align parallel to the applied field thereby producing a net magnetization in the direction of the applied field. This is still, however a relatively weak effect. However, all minerals are diamagnetic, and some are paramagnetic or ferromagnetic but in both cases, their magnetizations do not have significant contributors to the geomagnetic field.

Though, there is a class of magnetism that have great importance on geomagnetic studies. Certain materials not only have atomic moments, but neighboring moments interact strongly with each other. This interaction is a result of a quantum mechanical effect called exchange energy. Suffice is to say that the exchange energy causes a spontaneous magnetization that is many times greater than paramagnetic or diamagnetic effects (BLAKELY, 1996). These types of materials are called ferromagnetic. There are several types of ferromagnetic materials, depending on the alignment of their atomic moments. If the atomic moment aligned parallel to one another, results ferromagnetism; if the atomic moments are aligned antiparallel to one another and total moment is neutralized, results anti-ferromagnetism; and the last is the ferrimagnetism, in which atomic moments are antiparallel but, having different magnitudes, do not cancel. The strength of the magnetization of ferromagnetic and ferrimagnetic materials decreases with temperature and disappears at the Curie temperature (KEAREY ET AL., 2002).

The spontaneous magnetization of ferromagnetic materials can be very large at the scale of individual mineral grain but, due to their random orientation, the net magnetization may be negligible at outcrop scale. Due to the presence of ferromagnetic minerals, rocks will acquire a magnetization  $\mathbf{M}_i$ , called induced magnetization in the direction of applied field  $\mathbf{H}$  can be denoted as:

$$\mathbf{M}_i = \chi \mathbf{H} \tag{1.12}$$

If the rock is placed in a field-free environment, the induced magnetization falls to zero (BLAKELY, 1996). However, ferromagnetic materials have a special ability to retain a permanent magnetization even in the absence of external magnetic fields and it is called remanent magnetization, may be denoted by  $\mathbf{M}_r$ . The remanent magnetization of crustal rock depends not only on their atomic structure, crystallographic and chemical composition, but also on their geological, tectonic and thermal history. In geophysical studies, it is usual to consider the total

magnetization  $\mathbf{M}$  of the rock as the vector summation of induced and remanent magnetization, that is:

$$\mathbf{M} = \mathbf{M}_i + \mathbf{M}_r = \chi\mathbf{H} + \mathbf{M}_r \quad (1.13)$$

The ratio between remanent magnetization and induced magnetization is expressed by the Koenigsberger ratio as the following:

$$Q = \frac{|\mathbf{M}_r|}{|\mathbf{M}_i|} = \frac{|\mathbf{M}_r|}{|\chi\mathbf{H}|} \quad (1.14)$$

These magnetizations may be oriented in different directions and may differ significantly in magnitude. The magnetic effects of a rock arise from the resultant  $\mathbf{M}$  of the two magnetization vectors. Magnetic anomalies caused by the rocks are superposed to the geomagnetic field similar to gravity anomalies which are superposed to the gravitational field. However, the magnetic field is more complex, due to the variation in amplitude and in direction of the geomagnetic field. Consequently, knowledge of the behavior of the magnetic field is necessary both in the reduction of magnetic data to a suitable datum and in the interpretation of the resulting anomalies. The magnetic field is geometrically more complex than the gravity field of the Earth and exhibits irregular variation in both orientation and magnitude with latitude, longitude and time (KEAREY ET AL., 2002). Total-field magnetometers are usually the instrument of choice for airborne and shipborne magnetic surveys. As the name implies, total-field magnetometers measure the magnitude of the total magnetic field without regard to its magnetic direction.

The total field  $\mathbf{T}$  is given by:

$$\mathbf{T} = \mathbf{F} + \Delta\mathbf{F} \quad (1.15)$$

where  $\mathbf{F}$  is the geomagnetic field and  $\Delta\mathbf{F}$  represents the perturbation of  $\mathbf{F}$  due to some crustal magnetic sources. The total-field anomaly is calculated from total-field measurements by subtracting the magnitude of a suitable regional field, usually the IGRF model appropriate for the

date of the survey. If  $T$  represents the total field at any point, and  $F$  is the regional field at the same point, then the total-field anomaly is given by (BLAKELY, 1996):

$$\Delta T = |T| - |F| \tag{1.16}$$

If  $|F| \gg |\Delta F|$ , the total field  $\Delta T$  can be considered as the component of the anomalous field  $\Delta F$  in the direction of  $F$  and thus it can be considered a harmonic function (e.g., BLAKELY, 1996). This condition is usually verified in crustal magnetic studies.

The SI unit of magnetic field strength is the tesla (T). For the magnetic variation due to rock, a sub-unit, the nanotesla (nT), is commonly used;  $1 \text{ nT} = 10^{-9} \text{ T}$ . The strength of  $F$  varies from about 25000 nT in equatorial regions to about 70000 nT at the poles (KEAREY ET AL., 2002).

## 1.2. ELEMENTARY INVERSION THEORY

The starting place in most inverse problems is a description of the data. Since in most inverse problems the data are simply a table of numerical values, a vector provides a convenient means of their representation. If  $N$  measurements are performed in a particular experiment, for instance, one might consider these numbers as the elements of a vector  $d$  of length  $N$ . Similarly, the model parameters can be represented as the elements of a vector  $m$  of length  $M$ .

$$\begin{aligned} \text{data:} & \quad \mathbf{d} = [d_1, \dots, d_N]^T \\ \text{model parameters:} & \quad \mathbf{m} = [m_1, \dots, m_M]^T \end{aligned} \tag{1.17}$$

The basic statement of an inverse problem is that the model parameters and the data are in some way related. This relationship is called the model. Usually the model takes the form of one or more formulas that the data and model parameters are expected to follow.

The simplest and best-understood inverse problems are those that can be represented with the explicit linear equation:

$$A\mathbf{m} = \mathbf{d} \tag{1.18}$$

The matrix  $\mathbf{A}$  (with  $N \times M$  dimension) is called kernel. This is the basic equation of discrete inverse theory. Many important inverse problems that arise in the physical sciences involve precisely this equation. Others, while involving more complicated equations, can often be solved through linear approximations (MENKE, 1989).

When the number of the data ( $N$ ) is less than the number of model parameters ( $M$ ) the problem in equation (1.18) is called underdetermined problem. For these problems it is possible to find more than one solution for which the prediction error  $E$  is zero.

The first kind of a priori assumption we shall consider is the expectation that the solution to the inverse problem is simple, where the notion of simplicity is quantified by some measure of the length of the solution. One such measure is simply the Euclidean length of the solution:

$$L = \mathbf{m}^T \mathbf{m} = \sum m_i^2 \tag{1.19}$$

A solution is therefore defined to be simple if it is small when measured under the L2 norm. We pose the following problem: find the  $\mathbf{m}_{est}$  that minimizes equation (1.19) subject to the constraint that  $\mathbf{e} = \mathbf{d} - \mathbf{A}\mathbf{m} = 0$  and we obtain the **minimum length solution**:

$$\mathbf{m} = \mathbf{A}^T [\mathbf{A}\mathbf{A}^T]^{-1} \mathbf{d} \tag{1.20}$$

There are many instances in which  $L = \mathbf{m}^T \mathbf{m}$  is not a very good measure of solution simplicity. One may not want to find a solution that is smallest in the sense of closest to zero but one that is smallest in the sense that it is closest to some other value. The obvious generalization of equation (1.19) is then (MENKE, 1989):

$$L = (\mathbf{m} - \mathbf{m}_0)^T (\mathbf{m} - \mathbf{m}_0) \tag{1.21}$$

Where  $\mathbf{m}_0$  is the a **priori value** of the model parameters.

Sometimes the whole idea of length as a measure of simplicity is inappropriate and then we can introduce a weighting matrix  $\mathbf{W}_m$  that represents our a priori information. So, the equation (1.21) becomes:

$$L = (\mathbf{m} - \mathbf{m}_0)^T \mathbf{W}_m (\mathbf{m} - \mathbf{m}_0) \quad (1.22)$$

Frequently some observations are made with more accuracy than others. In this case one would like the prediction error  $\mathbf{e}_i$ , of the more accurate observations to have a greater weight in the quantification of the overall error  $E$  than the inaccurate observations. To accomplish this weighting, we define a generalized prediction error:

$$E = \mathbf{e}^T \mathbf{W}_d \mathbf{e} \quad (1.23)$$

where the matrix  $\mathbf{W}_d$  defines the relative contribution of each individual error to the total prediction error (MENKE, 1989).

When the problem is affected by numerical instability, it is called 'ill-posed'. From the numerical point of view, the discretization produces highly ill-conditioned systems of linear equations. This problem needs to be regularized to be solved. The most common and well-known form of regularization is the one known as Tikhonov regularization (TIKHONOV AND ARSENIN, 1977)

$$\mathbf{m}_\mu = \arg \min \{ \|\mathbf{A}\mathbf{m} - \mathbf{d}\|_2^2 + \mu^2 \|\mathbf{m}\|_2^2 \} \quad (1.24)$$

where the regularization parameter,  $\mu$  controls the weight given to minimization of the side constraint relative to minimization of the residual norm. Clearly, a large value of  $\mu$ , (equivalent to a large amount of regularization) favors a small solution semi norm at the cost of a large residual norm, while a small, (i.e., a small amount of regularization) has the opposite effect. This parameter also controls the sensitivity of the regularized solution  $\mathbf{m}$  to perturbations in  $\mathbf{A}$  and  $\mathbf{d}$ , and the perturbation bound is proportional to  $\mu^{-1}$ .

### 1.3. CLUSTER ANALYSIS

Cluster analysis (first used by TRYON, 1939) groups data objects based on information found in the data that describes the objects relationship. The goal is that the objects within a group be similar

and/or related to another and different from the objects in the other groups. There is a countless number of examples in which clustering plays an important role.

Biologists used cluster analysis to define the taxonomy of living things and in recent years to analyze the large amount of data from genetic information. Other applications can be found in psychology, medicine, business, and so on.

Cluster analysis itself is not one specific algorithm, and various algorithms for clustering and cluster validity have proliferated due to their promise for sorting out complex interactions between variables in high dimensional data (DUDA AND HART, 1973, TOU AND GONZALEZ, 1974, HARTIGAN, 1975).

There are several classification procedures (SNEATH AND SOKAL, 1973; GORDON, 1999), most may be grouped into four general types (DAVIS, 2002).

**Partitioning methods** operate in multivariate observations or on projections of the observations onto planes of lower dimension. Basically, these methods cluster by finding regions in the space defined by the variables that are poorly populated regions, and mathematically it is an iterative process that may be extremely time-consuming.

**Arbitrary origin methods** work on the similarity between the observation and a set of arbitrary starting points. The observation closest or most similar to a starting point is combined with it to form a cluster, and the observations are iteratively added to the nearest cluster, whose centroid is then recomputed for the expanded cluster.

**Mutual similarity procedures** group together observations that have a common similarity to other observations. A similarity matrix is computed between all pairs of observations, and the columns representing members of a single cluster will tend to have a precise intercorrelation, while having lower correlations with nonmembers.

**Hierarchical clustering** joins the most similar observations, then successively connects the next most similar observations to these. A  $n \times n$  similarity matrix of similarities between all pairs of observations is computed. Those pairs having the highest similarities are merged and the matrix is recomputed. The process iterates until the similarity matrix is reduced to  $2 \times 2$ .

Hierarchical clustering is most widely applied in Earth sciences, probably because their development has been closely linked with the numerical taxonomy of fossil organisms.

Given a sample  $X_1, \dots, X_n \subset \mathbb{R}^d$ , the population interpretation defines a partition of (a subset) of  $\mathbb{R}^d$ .

Considering four different clustering algorithms:

- *Hierarchical clustering* – the suggested population version is the partition of a low-level set of the density to disconnected components. This interpretation holds in the case of agglomerative clustering.
- *k-Means clustering* – the suggested population version is the minimal variance partition.
- *Density-based clustering* – clusters are defined in terms of the underlying density as regions of high density separated from other such regions by regions of lower density.
- *Tail clustering* – clusters are defined as regions of low density separated from other such regions after removing the regions of high density and regions of zero density.

Clustering algorithms define directly a partition of the sample, but it is possible to start with a population version and find a partition of the sample space, and find the induced partition of the sample. The k-means clustering and hierarchical clustering partition directly the sample and the underlying population interpretation is less important.

### *1.3.1. The k-Means Clustering*

The k-means algorithm belongs to the group of combinatorial clustering algorithms, and search for a partition that minimizes an objective function. In the k-means the objective function is a minimal variance functional. Since the number of all partitions is very large, it is reasonable to use greedy algorithm in solving the optimization problem (KLEMELA, 2009).

The k-means algorithm starts with an initial choice for the k cluster centers, finds for each cluster center its Voronoi neighborhood, and updates the cluster centers to be the average of the observations in the Voronoi neighborhood. The Voronoi neighborhood of a cluster center is the set of those observations that are closer to this cluster center than to the other cluster centers.

Set  $\mathbf{X} = \{x_1, \dots, x_n\} \subset \mathbb{R}^d$ , the k-means algorithm works:

1. Choose k cluster centers  $\mathbf{M} = \{m_1, \dots, m_n\} \subset \mathbb{R}^d$ .
2. Go through the cluster centers  $m \in \mathbf{M}$  until convergence.

- a) Find the Voronoi neighborhood  $X_m$  of the cluster center  $m$  (the set of those observations that are closer to  $m$  than to any other cluster center):

$$X_m = \{x \in \mathbf{X} : \|x - m\| = \min \{\|x - m'\| : m' \in \mathbf{M}\}\}.$$

- b) Update  $m$  to be the average of the observations in  $X_m$  :

$$m = \frac{1}{|X_m|} \sum_{x \in X_m} x$$

(1.25).

The k-means algorithm may be considered as a greedy algorithm to find the sample version of a minimum variance partition, or to find the sample versions of principal points (**MACQUEEN, 1967**; **KLEMELA, 2009**).

1. Find a partition  $P$  of  $\{x_1, \dots, x_n\}$  of size  $k$  such that:

$$\sum_{i=1}^n \|x_i - \mu(x_i; P)\|_2^2$$

(1.26)

is minimized with respect to all partitions of size  $k$ , where  $\mu$  is the arithmetic mean over the cluster where  $x_i$  belongs to

$$\mu(x_i; P) = \sum_{X \in P} \mu_X I_X(x_i) ,$$

$$\mu = \frac{1}{X} \sum_{i=1}^n x_i I_X(x_i)$$

2. Find points  $m_1, \dots, m_k \in \mathbb{R}^d$  such that

$$\sum_{i=1}^n \|x_i - \mu(x; m_1, \dots, m_k)\|_2^2$$

(1.27)

is minimized with respect to all sets of  $k$  distinct points, where

$$\mu(x; m_1, \dots, m_k) = \arg \min_{a \in \{m_1, \dots, m_k\}} \|x - a\|_2$$

### 1.3.2. The fuzzy c-Means Clustering

Fuzzy c-means (FCM) is a clustering method that allows each data point to belong to multiple clusters with varying degrees of membership. FCM (**BEZDEK, 1981**) is based on the minimization of the following objective function:

$$J_m(U, v) = \sum_{k=1}^N \sum_{i=1}^c (u_{ik})^m \|y_k - v_i\|_A^2 \quad (1.28)$$

where  $\mathbf{U}$  is the fuzzy c-partition of the data  $\mathbf{Y} = \{y_1, \dots, y_n\}$ ;  $c$  is the number of clusters in  $\mathbf{Y}$  and  $2 \leq c < n$ ;  $m$  is a weighting exponent and  $1 \leq m < \infty$ ;  $\mathbf{v}$  is the vector of centers. The squared distance between  $y_k$  and  $v_i$  is computed in the A-norm, where A is a positive-definite weight matrix.  $u_{ik}$  is a function interpreted as the grades of membership in the subset of Y and:

$$u_{ik} = \begin{cases} 1; & y_k \in Y_i \\ 0; & otherwise \end{cases}$$

$$\sum_{i=1}^c u_{ik} = 1$$

for all  $k$ .

The weighting exponent  $m$  controls the relative weights placed on each of the squared errors. As  $m \rightarrow 1$  partitions that minimize  $J_m$  become increasingly hard. Increasing  $m$  tends to defocus membership towards the fuzziest state.

FCM algorithm performs the following steps during clustering:

1. Randomly initialize the cluster membership  $u_{jk}$ .
2. Calculate the cluster centers

$$v_i = \frac{\sum_{k=1}^N u_{ik}^m y_k}{\sum_{k=1}^N u_{ik}^m}$$

3. Update  $u_{ik}$  according to the following:

$$u_{ik} = \frac{1}{\sum_{k=1}^C \left( \frac{\|y_i - c_k\|}{\|y_i - c_k\|} \right)^{\frac{2}{m-1}}}$$

4. Calculate the objective function  $J_m$ .
5. Repeat steps 2-4 until  $J_m$  improves by less than a specified minimum threshold or until after a specified maximum number of iterations.

#### 1.4. BASICS OF MONTE CARLO MARKOV CHAIN

In many geological investigations, data sequences may be created that consist of ordered successions of mutually exclusive states (DAVIS, 2002). A sequence in which the state at one point is partially dependent, in a probabilistic sense, on the preceding state is called a Markov chain. A sequence having the Markov property is intermediate between deterministic sequences and completely random sequences.

In a Markov chain process, having a set of states  $N = \{i, j, k, \dots\}$ , the process starts in one of these states and moves successively from one state to another. Each move is called step, if the chain is currently in state  $i$ , then it moves to state  $j$  at the next step with a probability denoted by  $p_{ij}$ , and this probability does not depend upon which states the chain was in before the current state.

A Markov chain is characterized by an  $N \times N$  transition probability matrix  $\mathbf{P}$  each of whose entries is in the interval  $[0,1]$ ; the entries in each row of  $\mathbf{P}$  add up to 1. The Markov chain can be in one of the  $N$  states at any given time-step; then, the entry  $P_{ij}$  tells us the probability that the state at the next time-step is  $j$ , conditioned on the current state being  $i$ . Each entry  $P_{ij}$  is known as a *transition probability* and depends only on the current state  $i$ ; this is known as the Markov property. Thus, by the Markov property:

$$\forall i, j, P_{ij} \in [0,1]$$

and

$$\forall i, \sum_{j=1}^N P_{ij} = 1$$

(1.29).

A matrix with non-negative entries that satisfies equation (1.29) is known as a stochastic matrix. A key property of a stochastic matrix is that it has a principal left eigenvector corresponding to its largest eigenvalue, which is 1.

## CHAPTER 2 – Simple Guided Clustering

*In this chapter a method to retrieve a geophysical anomaly map from potential field data transformation is illustrated. The method is based on clustering techniques, where the constraints that fix the center values are retrieved by an analysis of data distribution. Synthetic and real cases are shown and discussed. (The chapter is an article presently in press in *Geophysics*, Florio and Lo Re 2017).*

Gravity or magnetic maps may be more easily compared with geological data if they could be transformed in maps of the horizontal variation of the related physical property (density or magnetization). Two main approaches are known in literature: a method based on filtering the data in the Fourier domain and another working in the space domain.

In case of data acquired on a flat surface, a simple transformation in the Fourier domain can be used (e.g., [URQUHART AND STRANGWAY, 1985](#)). This method is based on a number of assumptions: the analyzed field could be approximated by the sum of the effects of vertically sided rectangular prisms, the area of their base is equal to one grid cell, the dip of contacts between geological units should be vertical, their depth extent is large (infinite), there is no remanent magnetization (in the magnetic case). The transformation is performed by dividing the anomaly amplitude spectrum for the factors related to the depth to the top of the assumed prismatic sources (resulting in a downward continuation to that depth), to the size of the prisms and, in the magnetic case, to the direction of the magnetization (resulting in a reduction to the pole). By anti-transforming the resultant spectrum, a physical property map can be obtained. The transformed map results sharper than the original data, and displays areas with rather constant values of density or susceptibility, making it easier to locate the boundaries between formations and the comparison with a geological map.

The approach in the space domain is based on the so called ‘terracing’ ([CORDELL AND McCAFFERTY, 1989](#)) a transformation where the original, smooth, gravity or magnetic field values are forced to be constant except across abrupt boundaries that should correspond to the contact between different geological bodies. The method uses a moving window formed by the four neighbor values immediately north, south, east, and west of a data point. The algebraic sign of the second vertical derivative of the field at the central point is evaluated and will determine if the value of the field at that point has to be increased or decreased to the value of its highest or lowest

immediate neighbor, or remain unchanged. The modified map is iteratively subjected to this process, until abrupt boundaries between homogeneous domains are obtained. The terraced field is assumed to be proportional to the physical property within a source confined between a specified depth-range. This gives the opportunity to compute the density (or magnetization) relative to each terrace by a simple least-squares inversion technique. The application of terracing to magnetic data requires their previous transformation in pseudo-gravity or reduced-to-the-pole anomalies.

**Cooper and Cowan (2009)** improved the terracing method by suggesting the use of the profile curvature, computed normally to the maximum gradient. This helps reducing the influence of the orientation of the reference system and the consequent tendency to have distortions such as 'blocky' features in the terraced map.

**Simpson et al. (2008)** used a drainage-divide approach to build a terraced map, and **Phillips and Simpson (2015)** applied the same idea to perform terracing during inversion, aimed at sharpening the boundaries of the inverted density distribution.

**Li (2016)** proposed the application of several edge-preserving smoothing filters, used in image processing, to terrace a potential field map. For example, the Kuwahara filter is based on the estimate of the standard deviation and mean of the data inside four sub-areas of a wider moving window. The central value of the window assumes the value of the mean of the most homogeneous sub-area. This process should preserve the edge position in the transformed map and in presence of noise leads to better results than the original curvature algorithm of **Cordell and McCafferty (1989)**. While the original terracing algorithm by **Cordell and McCafferty (1989)** tends to produce a terraced map too similar to the original data, the Kuwahara filter can produce terraced maps displaying different details by varying the window size and the number of iterations. However, also the application of the Kuwahara filter tends to produce blocky features in the terraced map, the dimension of which is related with the dimension of the used window. The Kuwahara filter, and in general this class of methods, can be modified using the median instead of the mean, to deal more efficiently with the presence of outliers (**Li, 2016**).

In this paper, we present a novel approach to the terracing of the potential field data, leading to the building of apparent density/magnetization maps. We obtain a terraced map through the classification of the original map values by clustering algorithms. Our procedure allows for an

efficient separation of the field values in a number of classes (clusters), whose center values are automatically found. Even the number of clusters can be automatically found. Differently from current terracing algorithms, the terraced map obtained by clustering algorithms as presented in this paper will consist of just a few number of piecewise constant field sections, greatly simplifying the original map pattern. Such a map is suitable to the building of apparent density/magnetization maps. In the next Sections, we will describe our terracing method, our simple inversion technique to build apparent density/magnetization maps, and we will test and apply the procedure on synthetic and real data.

## 2.1. METHOD

### 2.1.1. Terracing by clustering methods

Cluster analysis is a common tool for statistical data investigation. Clustering is referred to the task of grouping a set of objects in different clusters, in such a way that objects in the same cluster are more similar to each other than to those in other clusters.

One of the simplest clustering algorithms is the k-means algorithm (MACQUEEN, 1967). This algorithm aims at minimizing the following objective function:

$$P = \sum_{j=1}^k \sum_{i=1}^n \|d_i - c_j\|_2 \quad (2.1)$$

where  $k$  is the number of clusters,  $n$  is the number of data,  $\mathbf{d}$  represents the data vector,  $c$  the cluster central values and  $\|.\|_2$  stands for the L2 norm. In practice, the algorithm associates each data to the cluster having its center at the minimum Euclidean distance to it. In a successive step, the cluster center values are updated by computing the centroid of the cluster members, and the first step is repeated. After some iterations, the cluster centers no longer change, and the process stops. This algorithm needs the definition of the number of clusters,  $k$ , while the initial cluster centers,  $c$ , are chosen randomly in the sequence of data. Of course, both these choices will have influence on the result.

To apply the k-means clustering to the problem of terracing a potential field map it is necessary to input the number of clusters and, optionally, also their central values that will be updated in the iterative process in order to minimize equation (2.1), converging toward a minimum that can be either a global or more often a local minimum. Thus, the results are not unique but, in our experience, only a few clearly different clustering results are obtainable, and for each of them slightly different versions may be produced. The clustering process may result disappointing in presence of outliers.

For processing a gravity or magnetic dataset in order to get a terraced map, we here explore also a different approach; we propose a simple clustering method based on a classification of the data in a number of clusters centered on definite values. An automated process can define both the number of clusters and their center values, or we can assume the number of clusters and automatically select their center values. A datum will be assigned to a cluster based on a Euclidean distance criterion. Thus, this method is similar to a k-means clustering, but the center values are fixed, and no iteration is performed. For simplicity, in the following we refer to this method as the **Simple Guided Clustering** (SGC) technique. To guide the clustering process, i.e. to define the number of classes and their center values, we exploit the probability density function properties of the gravity/magnetic anomaly data. For example, a histogram of the distribution of the gravity values is useful to have indications about the shape of the underlying probability density function. The analysis of such a histogram can help defining the number of classes and their center values. However, a more flexible tool is a kernel smoothing function, producing a smooth and continuous probability curve (e.g., [BOWMAN AND AZZALINI, 1997](#)):

$$\hat{f}(y) = \frac{1}{nh} \sum_{i=1}^n w\left(\frac{y - y_i}{h}\right) \quad (2.2)$$

where  $n$  is number of samples,  $w$  is itself a probability density, called a 'kernel smoothing function', whose variance is controlled by  $h$  (called the 'smoothing parameter' or 'bandwidth'). Thus, differently from a histogram, which places each sample at the center value of discrete bins, a kernel distribution sums the component smoothing functions  $w$  for each data value to produce a smooth, continuous probability curve.

We use the information coming from the analysis of the kernel smoothing function to select the parameters guiding our clustering. In practice, the number of clusters and their center values will correspond to the number of peaks of the kernel smoothing function and to their corresponding field values. This statistical approach to define the clustering parameters eliminates the need of iterative optimization typical of the k-means method. **Ward et al. (2014)** used a similar approach, by performing fuzzy C-means clustering of resistivity images. Even in that case, the authors did not use an iterative scheme, assuming that the information obtained by the analysis of the kernel smoothing function represent adequate final centroids for their clustering process.

The kernel smoothing function,  $w$ , should be symmetric, with zero mean and standard deviation  $h$ . A natural choice for  $w$  is a normal density function (**BOWMAN AND AZZALINI, 1997**). The choice of the smoothing parameter  $h$  affects the shape of the kernel density function, in the sense that the higher the value of  $h$  and the smoother will be the curve. When the samples have a normal distribution, the optimal value for  $h$  (i.e., the bandwidth that minimizes the mean integrated squared error) can be computed as (**BOWMAN AND AZZALINI, 1997**):

$$h = \left(\frac{4}{3n}\right)^{1/5} \sigma \tag{2.3}$$

where  $\sigma$  is the standard deviation of the distribution. The normal density function is one of the smoothest distributions, so this approach is conservative, in the sense that the optimal value for  $h$  will be large, and the kernel smoothing function will tend to introduce oversmoothing.

Thus, by this method we can automatically select both the number and center values of the clusters. However, in the case of terracing of gravity or magnetic anomalies, it may be desirable that the choice of the number of clusters incorporate in some way the previous knowledge of the outcropping geology in the area. Thus, in some cases, the value of the smoothing parameter  $h$  may be manually changed to modify the resolution of the kernel smoothing function and finally obtain the desired number of peaks (clusters). In this case, even if the number of clusters is manually selected, their center values are still automatically defined as the gravity/magnetic anomaly values corresponding to the peaks of the kernel smoothing function. In Figure 1, a histogram representation of a random data set and its kernel smoothing function are compared, and the effect of the variation of the bandwidth parameter,  $h$ , is shown.

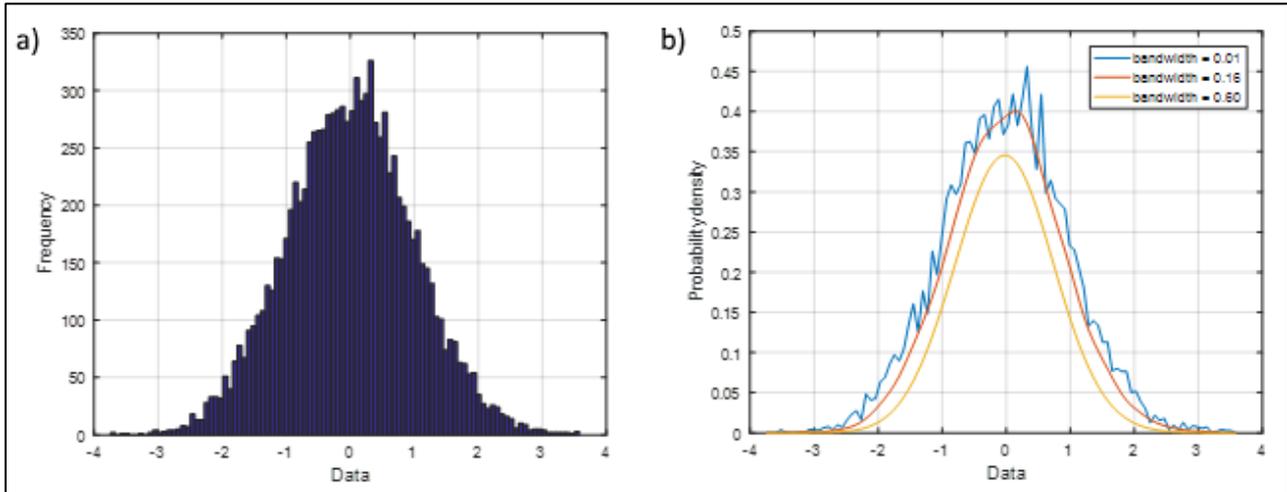


Figure 1- Histogram vs. kernel smoothing function. a) Histogram of the frequency distribution of a Gaussian random noise. b) Kernel smoothing function of the same data, computed using different values of the bandwidth parameter (equation 2.1). The bandwidth value,  $h=0.16$ , is found by Equation 2.2.

We will give more details on the practical application of this approach while describing the terracing of the first synthetic data set.

### 2.1.2. Apparent density/magnetization maps from terraced potential fields

In next Sections, we will show applications of the described clustering methods for terracing synthetic and real data. However, although these maps may be found useful when compared to a geologic map, a final goal may be to compute an apparent density/magnetization map. We start from a terraced map using  $k$  clusters, which has  $k$  distinct areas, each one characterized by a constant value of the original field, and implement an inverse approach based on solving the following overdetermined system of equations:

$$\mathbf{d} = \mathbf{A}\mathbf{m} \tag{2.4}$$

where  $\mathbf{d}$  is the  $n \times 1$  vector of the input data (the map of the gravity anomalies or of the magnetic anomalies reduced to the pole);  $\mathbf{A}$  is a  $n \times k$  matrix containing in each column the gravity or magnetic effect generated in all the map points by a single terraced area with unit density/magnetization;  $\mathbf{m}$  is the  $k \times 1$  unknown vector of apparent density/magnetization of each terraced area.

We proceed to the estimate of the apparent density/magnetization according to the following steps:

For each cluster, a matrix of unit density/magnetization is created, setting to 1 all the points belonging to that cluster and to 0 all other map points.

We compute the elements of  $\mathbf{A}$  (equation 2.4) as the gravity or magnetic effect of each one of such matrices in all the map points. For this purpose, we use the Parker formula ([PARKER, 1972](#)), allowing for horizontal variations of the density/magnetization distribution in a layer. We set the depth to the surfaces defining the top and bottom of the layer based on the available geological information on the area. This formula takes into account the presence of undulated surfaces.

The resulting overdetermined system is finally solved by a least squares method to recover the apparent density/magnetization vector.

## 2.2. APPLICATION ON SYNTHETIC DATA

In this Section, we will explore the performance of the SGC terracing and k-means clustering applied to the problem of terracing gravity or magnetic maps. We will apply these techniques on the gravity field generated by interfering prismatic sources and then to the total field anomalies, contaminated with noise, generated by the Bishop model ([WILLIAMS ET AL., 2005](#)). To appreciate better the performance of our method, its strengths and weaknesses, we will also apply the Kuwahara filter ([LI, 2016](#)). Finally, we will try to improve the results by combining this type of filtering with clustering techniques.

### 2.2.1. The gravity field of two prismatic sources

To study the performance of the SGC method we will use the synthetic gravity field generated by two prismatic sources having a square base with sides 13 km long, a depth to the top of 2 km and thickness of 10 km. In Figure 2a, the leftmost prism has a density contrast with the background of  $1 \text{ g/cm}^3$ , and its sides are rotated of  $45^\circ$  with respect to the coordinate axes. The rightmost one has a density contrast with the background of  $-1 \text{ g/cm}^3$  and has its sides parallel to the coordinate

axes. Their gravity effect is sampled on an area of 121x121 km<sup>2</sup>, with a grid step of 1 km (Figure 2a).

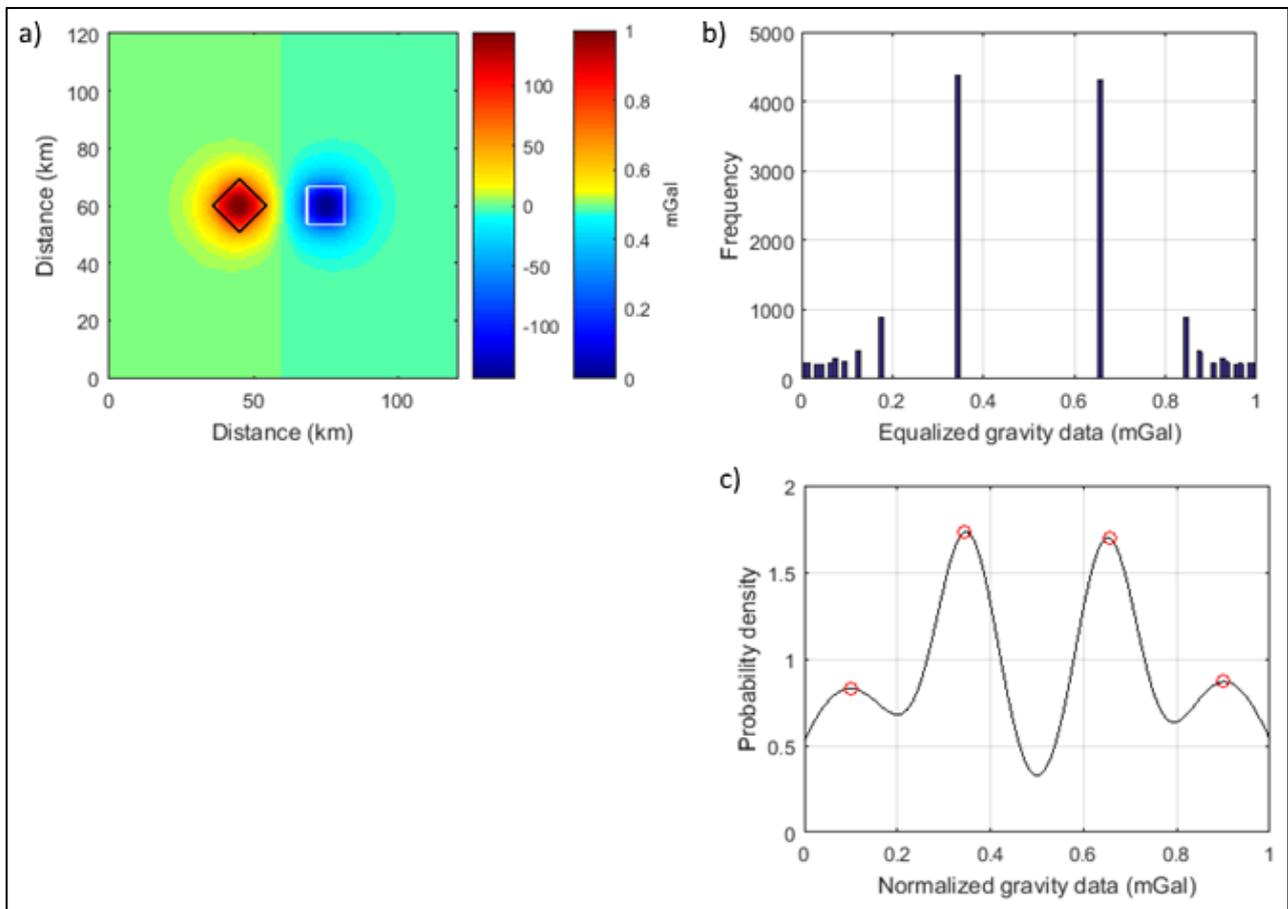


Figure 2 - Estimation of cluster parameters for terracing by SGC. a) Gravity field generated by two prismatic sources having a square base with sides 13 km long, a depth to the top of 2 km and thickness of 10 km. The leftmost prism has a density contrast with the background of 1 g/cm<sup>3</sup>, and the rightmost one has a density contrast with the background of -1 g/cm<sup>3</sup>. The grid has 121 rows and 121 columns, and a grid step of 1 km. The original data are normalized between 0 and 1, as illustrated by the second color bar. b) Frequency histogram of the equalized data. c) Kernel smoothing function of the equalized data, showing 4 peaks (highlighted by the red circles) at 0.10, 0.34, 0.66 and 0.90 mGal, corresponding, in the original range, to -115.15, -45.18, 45.19 and 115.17 mGal. The kernel smoothing function was computed using equation 2.3 for the choice of the smoothing parameter  $h$ .

We use this example also to illustrate a practical feature of our method, i.e. its application to a histogram-equalized version of the gravity/magnetic anomaly map. In fact, in its original form, the histogram of the gravity/magnetic map may be dominated by the most common values, failing to display clearly all the key features of the data. Histogram equalization is a commonly used image-processing tool that transforms the original histogram so that it approximately matches a uniform distribution, enhancing the presence of small populations. Thus, the SGC procedure involves the following steps, also illustrated in Figure 2:

- a. the original gravity data are normalized between 0 and 1 (Figure 2a); this normalization is required for the successive step b);
- b. the normalized data are subjected to histogram equalization (Figure 2b);
- c. the kernel smoothing function of the transformed data is computed (Figure 2c);
- d. the number of peaks of the kernel smoothing function is taken to define the number of clusters;
- e. their corresponding values, transformed back to the original data range, provide the centroids for the clusters;
- f. the gravity data are clustered using these parameters by a simple, non-iterative, minimum distance criterion (SGC), producing a terraced map.

For the case of Figure 2a, the histogram of the original data is dominated by the background values and do not show any peaks related to the gravity anomalies of the two sources. However, from the kernel smoothing function of the equalized data (Figure 2c) there appear 4 peaks at 0.10, 0.34, 0.66 and 0.90 mGal, corresponding, in the original data range, to -115.15, -45.18, 45.19 and 115.17 mGal. The two extreme values (-115.15 and 115.17 mGal) are related to the main gravity low and high in Figure 2a, while the two intermediate values are related to the background values, in fact slightly changing from positive to negative values going from W to E. Thus, we apply SGC using four clusters centered on the found values. The corresponding terraced gravity map is displayed in Figure 3a.

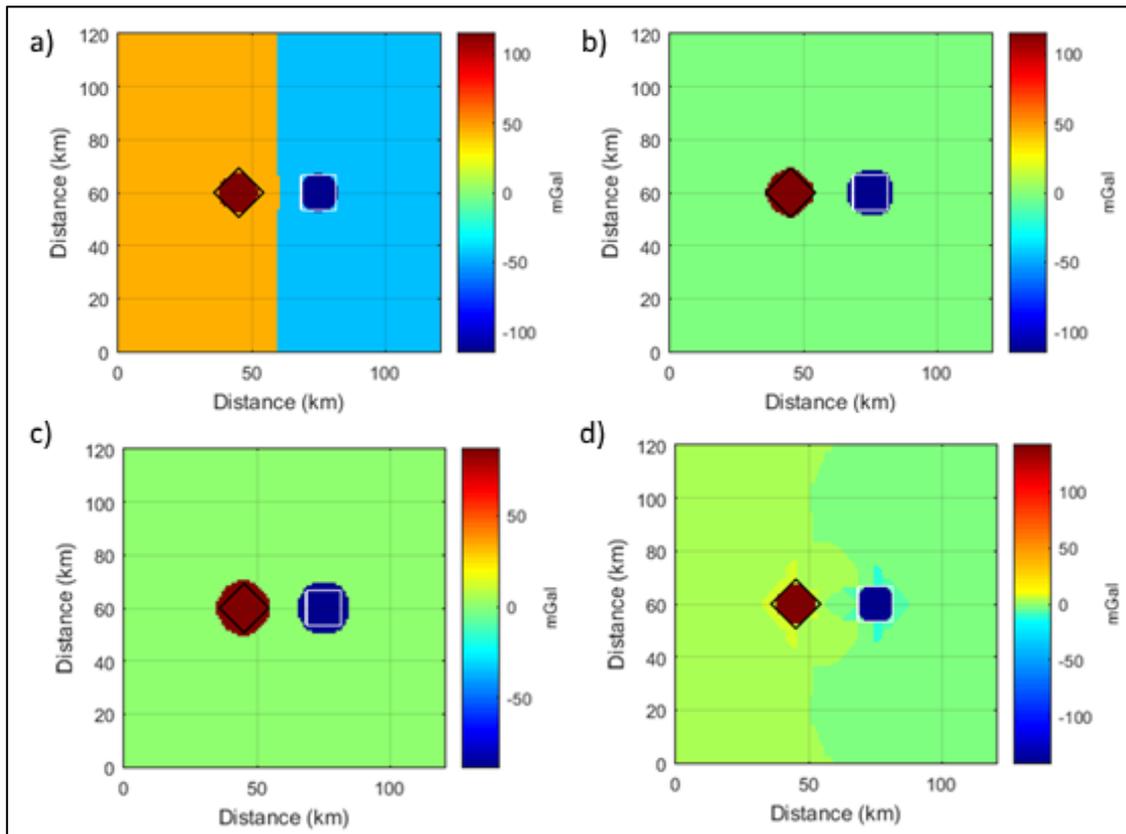


Figure 3 - Terracing of gravity data of Figure 2a by different approaches. a) SGC using the 4 clusters and respective centroid values as obtained by inspection of Figure 2c. b) SGC using just 3 clusters, namely the two extreme values and the average of the two intermediate values of the previous case. c) k-means clustering assuming 3 clusters whose centers are automatically found. d) Kuwahara filter, using a 3x3 window and 20 iterations.

The terraced field data correspond strictly to the found four values for the clusters centroids, and the area of the two sources is identified very well. On the other hand, in Figure 3a the background area is characterized by positive and negative gravity values, while the original background gravity field is almost zero everywhere (Figure 2a). In this case, we determined automatically both the number (by using equation 2.3 for the choice of the smoothing parameter  $h$ ) and centroid values of the clusters. However, looking at the original data (Figure 2a) one would perhaps choose just 3 clusters, one for the background and the others for the two anomalies. Thus, we may perform the SGC using just 3 clusters derived from Figure 2c, two of them related to the main gravity low and high in Figure 2a (the two extreme values: -115.15 and 115.17 mGal), and another one representative to the background, taken as the average value of the two intermediate peaks (0 mGal; Figure 2c). The result is shown in Figure 3b, where now the background has a single value about 0 and the area of the two sources is still nicely identified, although with less resolution than before. In Figure 3c, the terraced map using k-means clustering and 3 clusters is shown. The

automatically found cluster centers are -87.4, 0 and 87.7 mGal. Also in this case, the field results perfectly terraced on these three field values, without any intermediate value, and the sources' area is nicely represented, although with less resolution with respect to the SGC results. In Figure 3d is shown the terraced field using the Kuwahara filter (window of 3x3 points and 20 iterations, the best parameters we found after some trials for this case). In this map, the two sources' areas are successfully identified, with an optimal resolution, although there is still the presence of intermediate values. All these maps are well suited for subsequent calculations of apparent density, but the clustered ones have the advantage of being perfectly terraced on just three (or four) field values and the area corresponding to the apparent sources is well estimated.

The same case was repeated after adding a Gaussian noise to the data of Figure 2a. The noise has a standard deviation of 0.1 mGal, a zero mean and a maximum amplitude equal to about the 3% of the original anomaly amplitude. We computed the kernel smoothing function (Figure 4a) by using equation 3 for the automatic choice of the smoothing parameter and we find five peaks corresponding, in the original data range, to -81.53, -39.88, -7.16, 34.49 and 96.97 mGal. The terraced field (Figure 4b) presents a constant value for the background and two classes for each of the two anomalies.

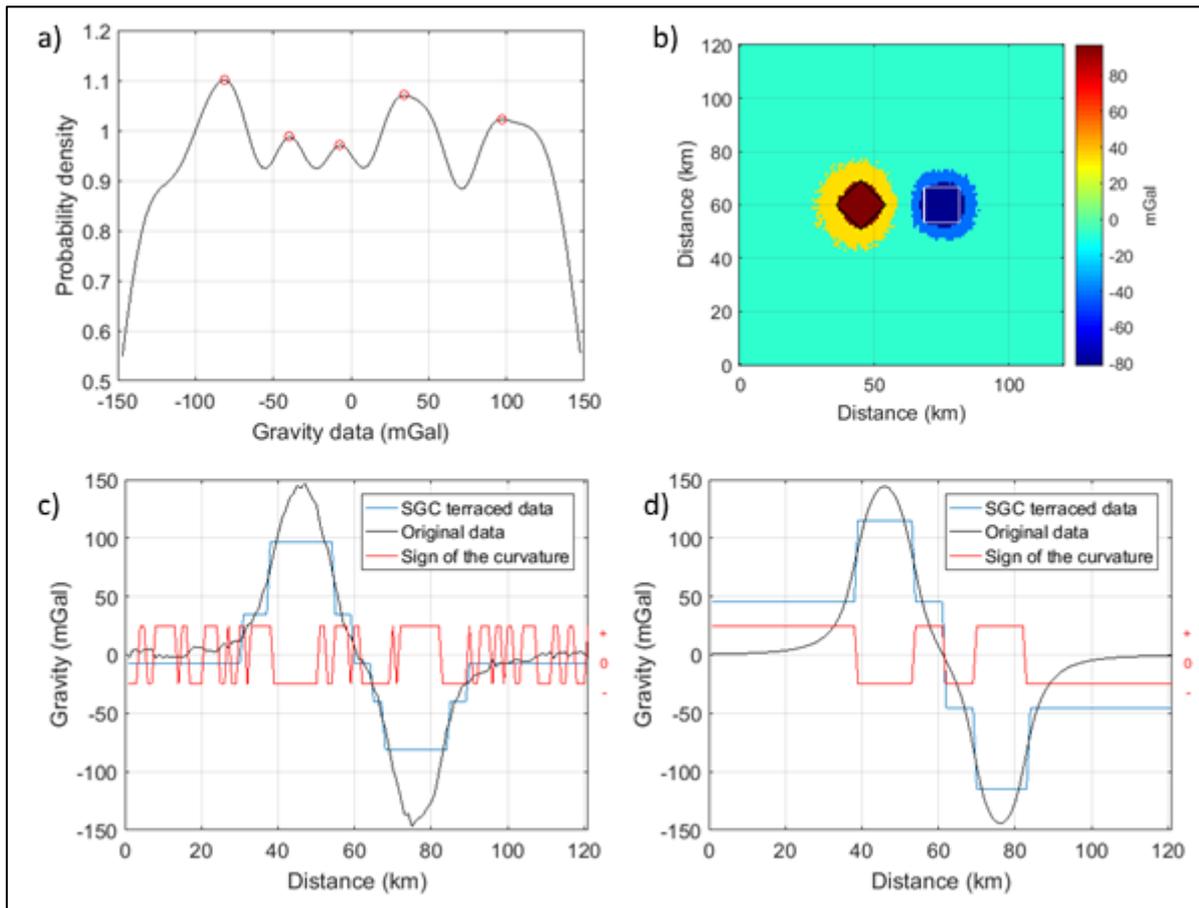


Figure 4 - Terracing of gravity data of Figure 2a corrupted by a Gaussian noise with a standard deviation of 0.1 mGal, a zero mean and a maximum amplitude equal to about the 3% of the original anomaly amplitude. a) Kernel smoothing function computed by the automatic selection of the smoothing parameter,  $h$ , by equation 2.3. Red circles evidence the five peaks found, corresponding to the five clusters of the successive terracing, centered at the respective values of the gravity field. b) SGC terraced data using the parameters determined in a). c) West-East profile of the terraced field (b) at  $y=60$  km. The original field and the sign of its curvature are shown for reference.

In Figure 4c and 4d the results of terracing of the gravity field with and without noise are compared along a West-East profile at  $y=60$  km. In these figures, we show also the variation of the sign of the input gravity field curvature ( $\approx \partial^2 g / \partial x^2$ ) along the profile, so that an approximate comparison with the results from the standard terracing method can be made. We notice that in the case without noise (Figure 4d, relative to a profile extracted from Figure 3a) there is a perfect spatial correspondence between the change of the sign of the curvature and the change of the cluster value in the terraced field. This correspondence is less clear in the case of noisy field (Figure 4c) because the noise gets strongly enhanced in the curvature, while the terraced field remains rather stable.

### 2.2.2. The total field of the bishop model

In this Section, we will apply SGC and k-means terracing techniques to the magnetic field generated by the synthetic model of magnetic basement known as the Bishop model (e.g., WILLIAMS ET AL., 2005). The Bishop model represents an articulated topography of the top of a magnetic basement (Figure 5a) overlain by non-magnetic sediments and intruded by four isolated highly magnetized bodies. The basement extends down to 20 km depth and has a variable magnetization representing three different geologic units (Figure 5b). Besides these lateral magnetization contrasts, the total field intensity is modulated by the basement's top undulations (Figure 5a). The geomagnetic field strength is 50000 nT. We added a Gaussian random noise with 2 nT standard deviation and zero mean. As required by every terracing algorithm, we consider the total field data due to a vertical total magnetization of the sources (Figure 5c). The synthetic field grid has a sampling step of 1.6 km and 252 rows x 238 columns.

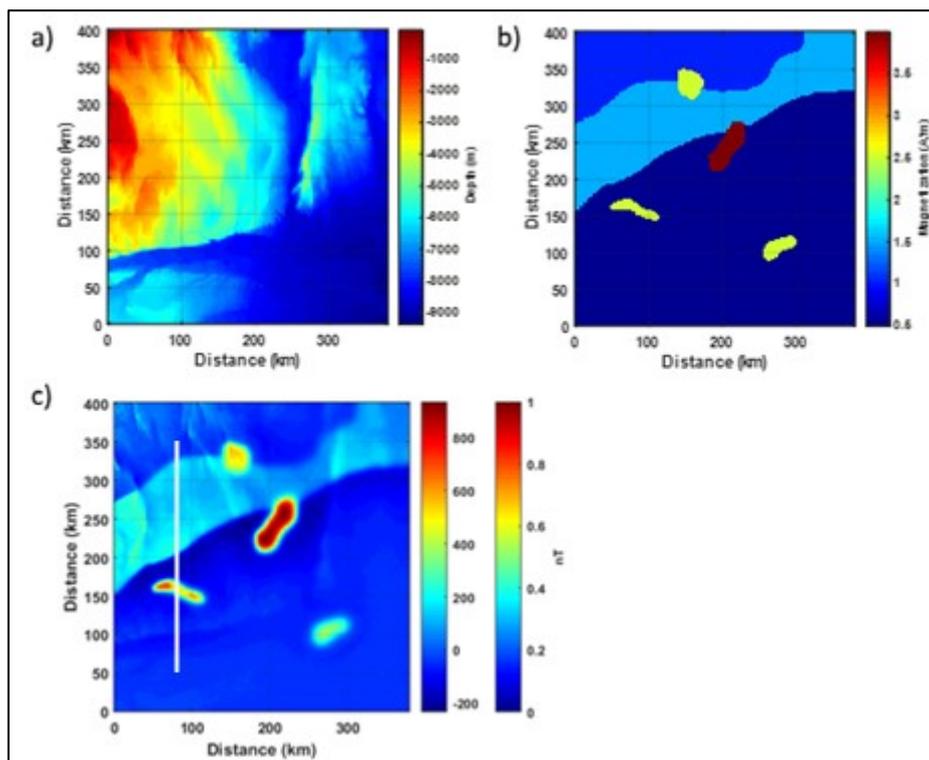


Figure 5 - The Bishop model magnetic basement. a) Depth to the top of the magnetic basement. b) Magnetization variations in the Bishop basement. c) Total magnetic field generated by the Bishop synthetic basement. The geomagnetic field and magnetization vectors are vertical. The color bar between 0 and 1 is drawn for reference, and illustrates the correspondence between the normalized data and the original ones. The white line indicates the location of the profile shown in Figure 8.

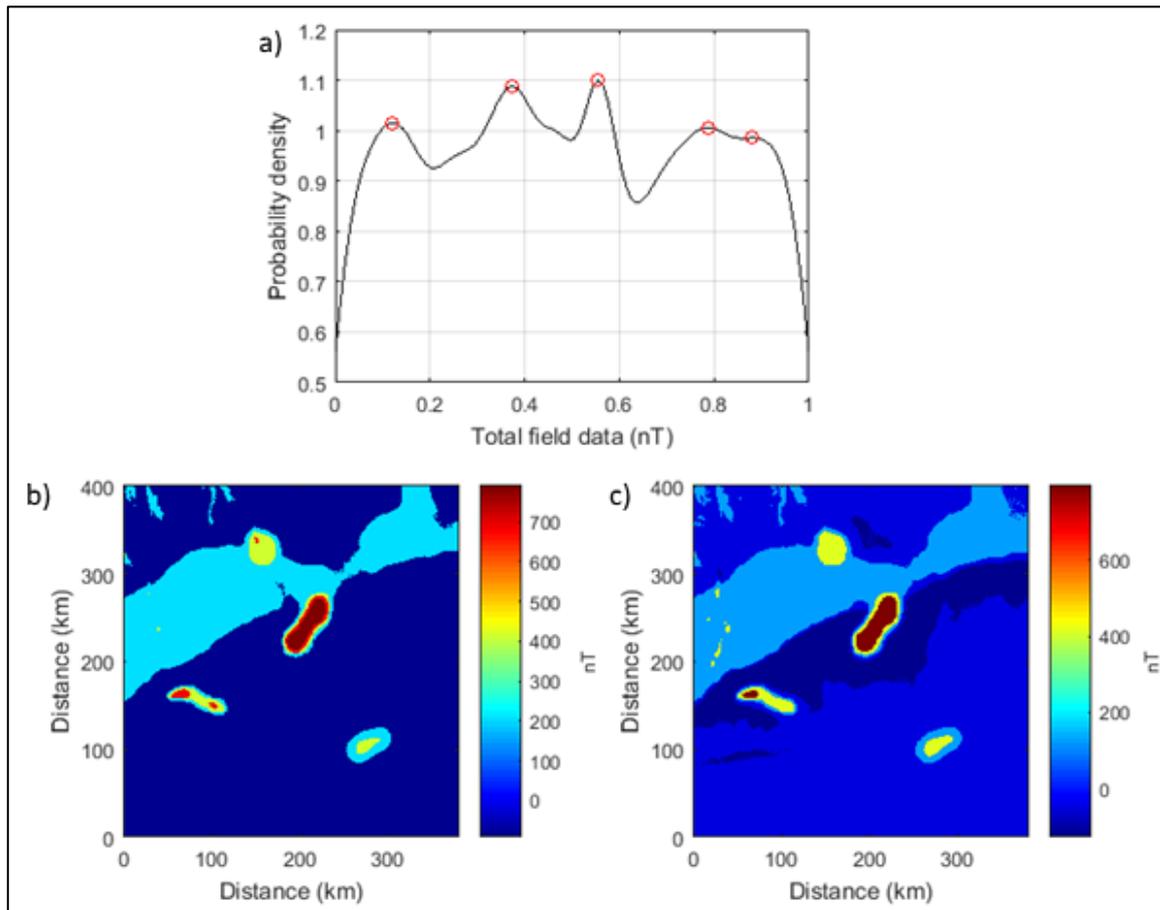
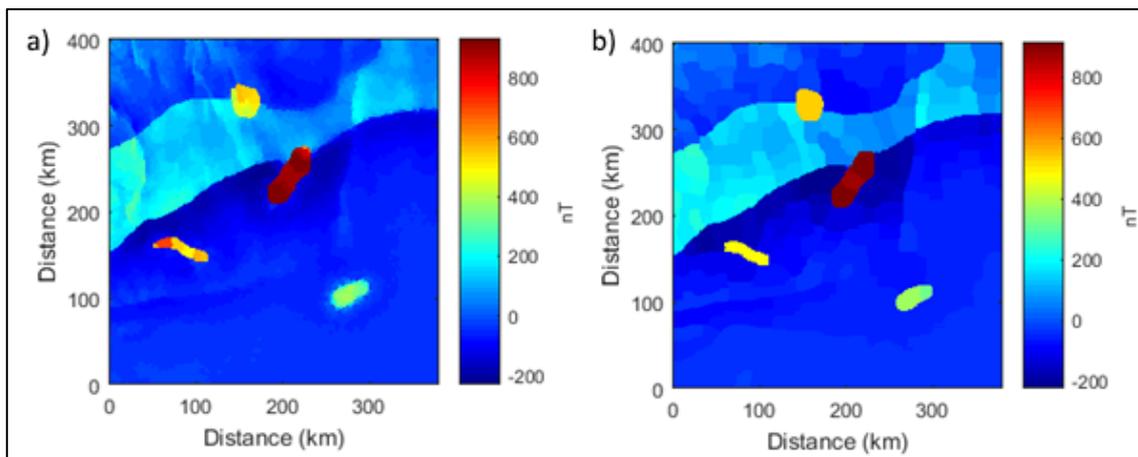


Figure 6 - Terracing of Bishop total field data. a) Kernel smoothing function of the histogram equalized total field intensity of Figure 5c. The 5 peaks of the curve (highlighted by the red circles) and their respective values are used to set up the SGC clustering process. b) SGC terracing. c) k-means terracing assuming 5 clusters whose centers are automatically found.

To use the SGC terracing, we follow the processing described in the previous section. From the histogram equalized version of the data we computed the kernel smoothing function (Figure 6a) from which we determined the key parameters for the clustering process, i.e. the number of clusters (5) and their center values (-92.14, 202.94, 415.40, 686.88 and 793.11 nT). We show the resulting SGC terraced map in Figure 6b. We performed also the k-means clustering of the magnetic data, using 5 clusters, and the automatic centers were found to be -64.30, 50.53, 166.92, 443.60 and 808.40 nT. The relative terraced map is shown in Figure 6c. The differences in the SGC and k-means terraced maps are due to the different cluster centers found by the two procedures. In the case of the k-means, the clustering process starts from a random set of centroids that are iteratively updated, trying to minimize equation 2.1, up to convergence. Because of the initial random choice of the centroids, the result is not unique, but a few differently clustered maps may

be produced. However, they are similar enough as to be considered equivalent and, after a visual inspection, one of the clustered map is finally chosen.

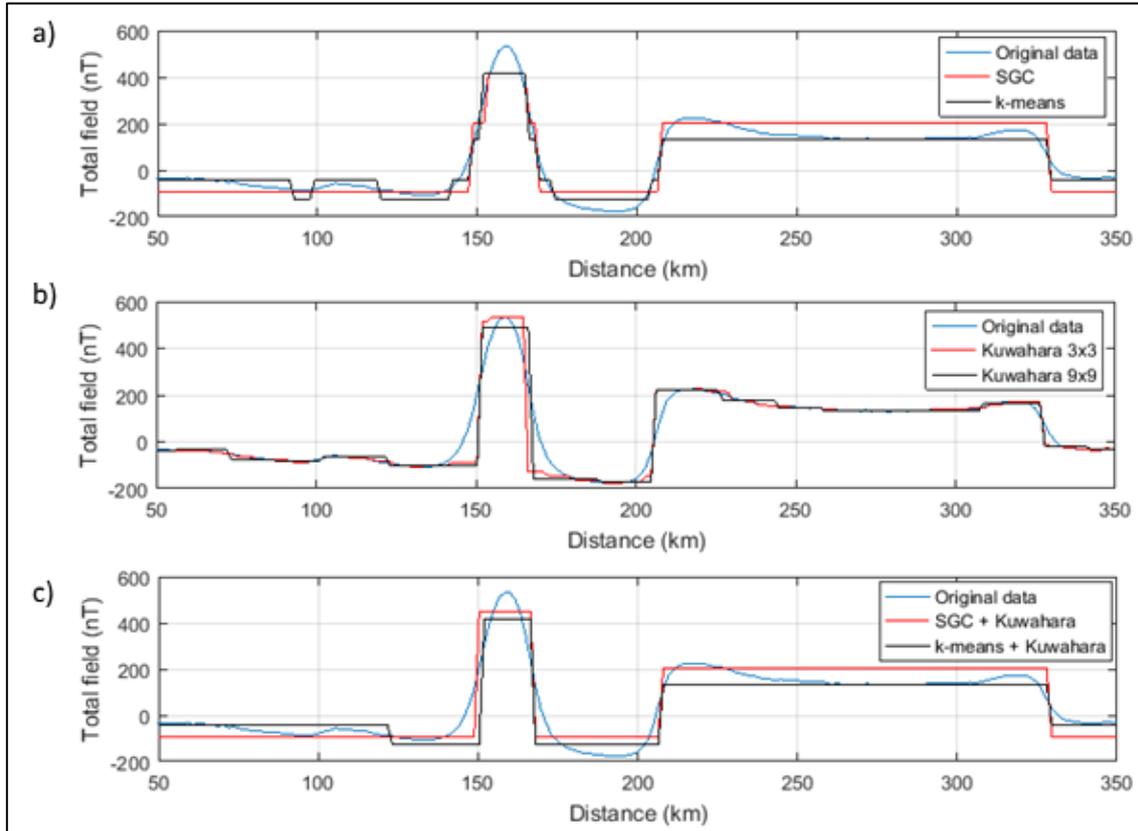
The main characteristic of both SGC and k-means terraced maps is to have a desirable quality for a terraced map, namely to present abrupt field changes separating a number of areas having different, constant, field values. We can spot some differences between the two maps. For example, in the southern area the background field is constant in the SGC–terraced map (Figure 6b), while it varies in the k-means case (Figure 6c), with minimum values toward North. In general, both terracing procedures successfully simplify the original map, highlighting the main horizontal geology variations. On the other hand, in presence of strong horizontal gradients of the original field (e.g., the effect of the intrusion at  $x=200$  km,  $y=250$  km, Figure 5c), the clustering–terraced maps present a number of discrete steps (Figure 6), while the correspondent geology is instead abruptly changing. This may be disturbing when recovering the apparent magnetization from the terraced map.



*Figure 7 - Terracing the Bishop total field by the Kuwahara filter: a) using a 3x3 window and 20 iterations; b) using a 9x9 window and 20 iterations.*

In Figure 7 we present the terracing of the Bishop magnetic data by the Kuwahara filter using different window sizes, namely a small one (3x3, Figure 7a) and a bigger one (9x9, Figure 7b). These maps are characterized by a successful identification of different areas corresponding to different magnetic properties, and the boundaries between these areas are very sharp, with no intermediate values between them. However, the terraced field is not constant in these areas. The window size determines the ‘texture’ of these variations and, when this size is big, blocky features

appear in the terraced map (Figure 7b). One way to overcome this effect is to use a circular window (Li, 2016), or adaptive implementations of the filter shape depending on the local structure of an input image (KYPRIANIDIS ET AL., 2009).



*Figure 8 - Comparison of terracing of the Bishop magnetic field along a S-N profile at  $x = 83.2$  km. a) SGC and k-means clustering methods. b) Kuwahara filters of the Bishop magnetic field with different window sizes. c) Kuwahara filtering of the clustered data.*

We combine the clustering techniques and the Kuwahara filtering to try to obtain a terraced map having the best features of the two transformations:

- a. a number of terraces having different, constant, field values (Figure 8a);
- b. very sharp boundaries between terraces, with no intermediate values between them (Figure 8b).

We show in Figure 8c and 9a,b the results of Kuwahara filtering (using a window size of 9x9 data points and 20 iterations) of the SGC and k-means terraced maps of Figure 6b and 6c. The resulting maps have now the two desirable characteristics outlined above, describe clearly the geologic variations in the area and are simple enough to be suitable for the apparent magnetization computation.

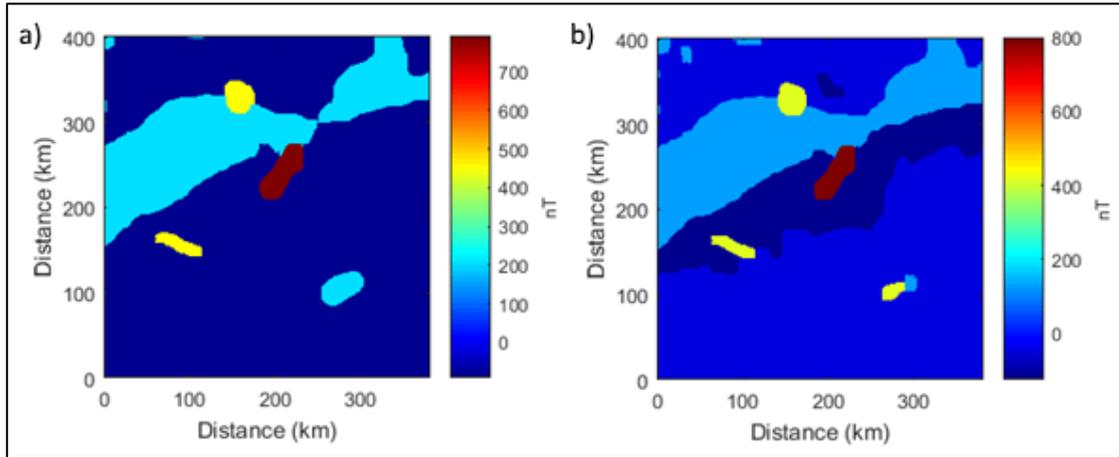


Figure 9 - Kuwahara-filtered terraced maps. a) SGC-terraced map of Figure 5b transformed by a Kuwahara filter with a window of 9x9 points, iterated 20 times. b) k-means terraced map of Figure 5c transformed by a Kuwahara filter with a window of 9x9 points, iterated 20 times.

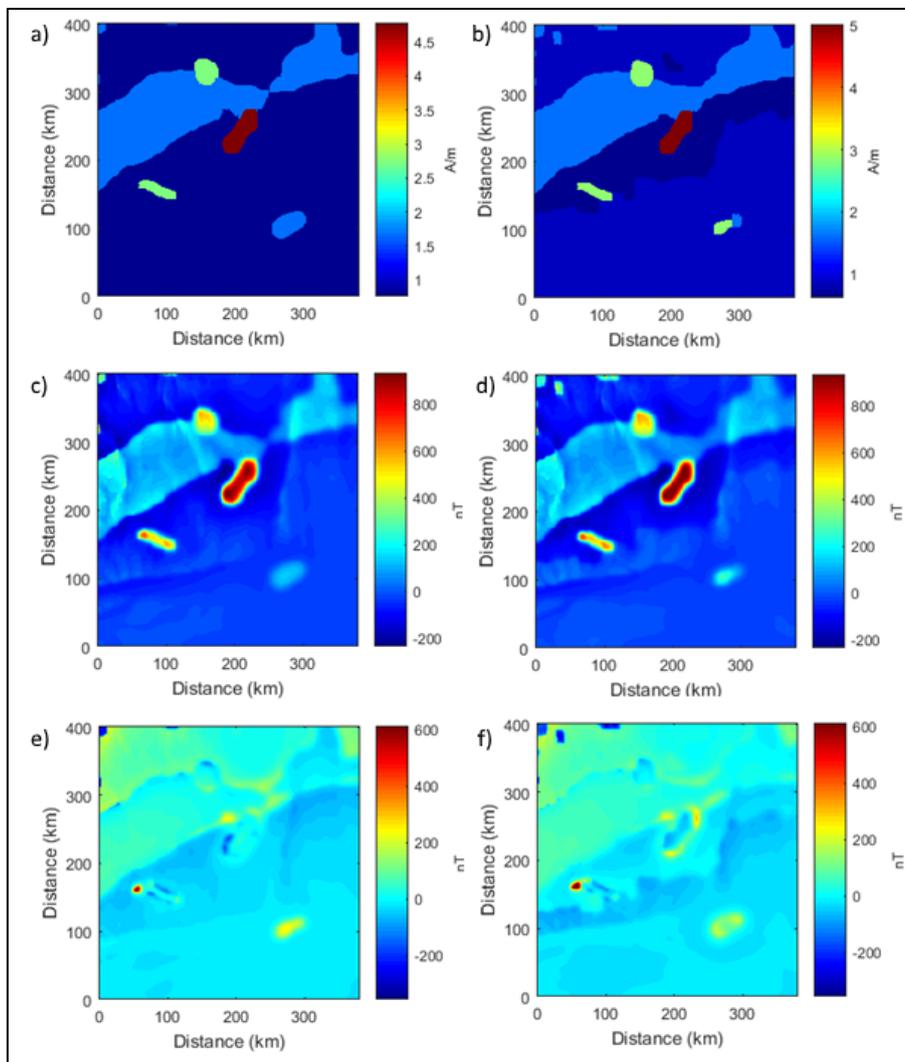


Figure 10 - Apparent magnetization computation. a) Apparent magnetization obtained from the SGC-terraced map transformed by a Kuwahara filter (Figure 9a). b) Apparent magnetization obtained from the k-means-terraced map transformed by a Kuwahara filter (Figure 9b). c) Total field computed from the apparent magnetization map in a). d) Total field computed from the apparent

*magnetization map in b). e) Residuals between the computed total field (c) and the observed total field of Figure 5c. f) Residuals between the computed total field (d) and the observed total field of Figure 5c.*

To perform this calculation, we assumed the true basement's top and thickness and followed the procedure described in the Method section. The apparent magnetization maps (figures 10a,b) have clear similarities with the true magnetization distribution (Figure 5b), showing the magnetization difference among the intrusive bodies and among different sectors of the basement. However, these maps fail in highlighting the difference in magnetization between the northern and southern sectors of the basement, and in the k-means apparent magnetization map (Figure 10b) a fictitious change of magnetization appears in the basement southern sector. In general, the found magnetization values are higher than the true ones. Despite these errors in reproducing the true magnetization distribution, the difference between the true total field and that produced by the apparent magnetization maps of figures 10a and 10b has, respectively, a rms value of 50.0 nT (SGC, Figure 10a) and 51.2 nT (k-means, Figure 10b). These rms values are quite good for this approximate procedure, considering that the original total field amplitude range exceeds 1000 nT. These rms values are some units smaller than those obtainable from the analysis of the simple clustered maps (not Kuwahara-filtered; not shown). Besides this improvement, the Kuwahara filtering of the clustering-terraced maps provide us with an apparent density/magnetization map cleaner than that produced by the only clustering techniques, because of the partial or total fixing of the stepped pattern in the high gradient areas.

As shown before, by setting an appropriate value for the bandwidth parameter, we may increase the number of peaks of the kernel smoothing function and reclassify our data using a greater number of clusters. In the Bishop case, for example, we might have chosen a higher number of clusters to better represent the observed field. However, we think that the cluster number should be kept to a minimum to allow for a substantial simplification of the original map and thus obtaining an effective terracing. Thus, the choice of the clusters number depends on the degree of simplification of the original field that is wanted. In any case, the terraced map computed by clustering methods will have a discrete number of perfectly flat terraces separated by abrupt boundaries. Stepped structures may form in case of high horizontal gradients in the original field. The rms value we computed to express the overall similarity between the observed and computed

data will also depend on the number of clusters used, and will decrease when the number of clusters is increased, as the terraced field tends to become more similar to the observed field.

Finally, we highlight the fact that in real cases, differently from the Bishop case, the information about the depth to basement and its thickness is not available, so the resulting apparent magnetization map will suffer for a much greater approximation than in this synthetic application.

A final consideration regards the stability of the proposed technique with respect to the noise. The added noise in this case of the Bishop magnetic field had a rather low amplitude. **Li (2016)** showed how in the maps terraced by the method of **Cordell and McCafferty (1989)** the noise is significantly enhanced, as this method is based on the computation of the second vertical derivative of the data. However, our procedure does not involve any differentiation of the field and, as the above Bishop synthetic example demonstrated, in the clustering-terraced map the noise is not enhanced.

### **2.3. APPLICATION ON REAL DATA**

We apply now clustering techniques to terracing real gravity data in a mining area, and will use the terraced map to compute an apparent density map. To give the reader the opportunity to compare the performance of different terracing algorithms, we analyzed the same gravity dataset used in previous papers (**KEATING, 1992; LI, 2016**), relative to the Noranda-Val-d'Or region of the Abitibi Greenstone Belt, Canada.

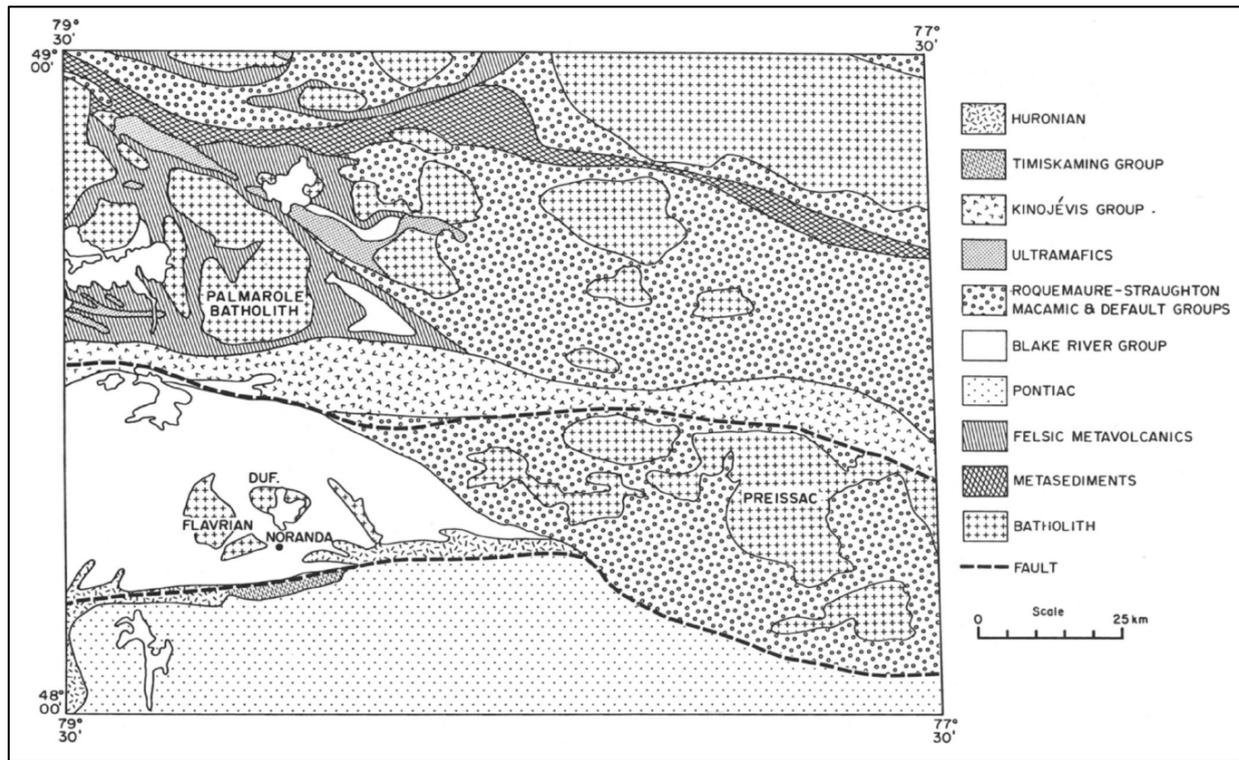


Figure 11 - Simplified geological map of the Noranda-Val-d'Or region (after Keating, 1992).

The area is extensively mined since the discovery of gold and other base metals in outcropping volcanic rocks (Blake River Group, the Kinojévis and the Straughton-Rocquemaure Groups, Figure 11) as massive sulphide deposits. In the area, also some granitic intrusions outcrop (KEATING, 1992). The gravity point-data are available for download from the Natural Resources Canada web site. The gravity data are irregularly distributed in the study area, with the highest station density in the central-western area and only sparse sampling in the NE. We interpolated the data on a 1 km regular grid in an area from 48°N to 49°N and from 79.5°W to 77.5°W (Figure 12a).

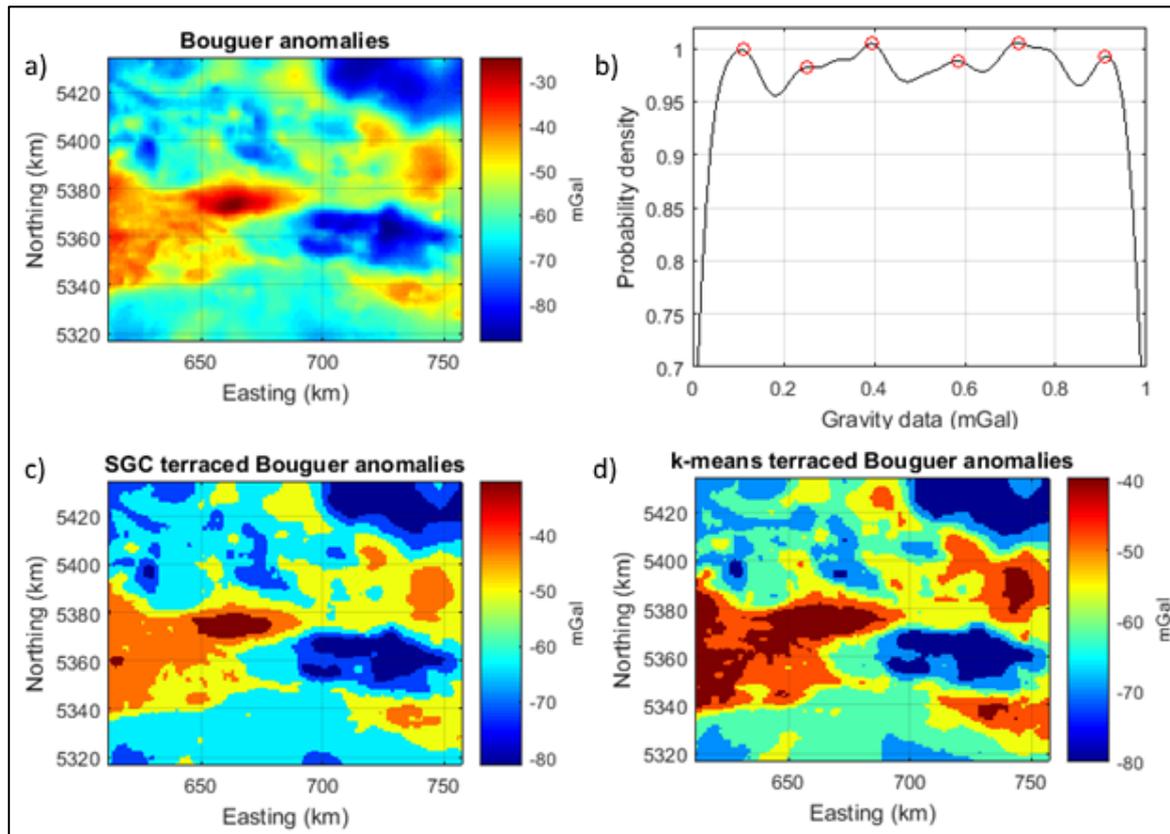


Figure 12 - Analysis of real gravity data of Noranda-Val-d'Or region (Canada). a) Bouguer anomalies as interpolated on a 1 km regular grid. The projection used is UTM zone 17N for the WGS84 ellipsoid. b) Kernel smoothing function of the histogram equalized data. The 6 peaks of the curve (highlighted by the red circles) and their respective values are used to set up the SGC clustering process. c) SGC terracing. d) k-means terracing, assuming 6 clusters whose centers are automatically found.

To terrace the gravity map of Figure 12a by the SGC method, we first computed the kernel smoothing function of the histogram equalized data. Using the bandwidth given by equation 2.3, we obtained a function with just 3 peaks, that should correspond to the number of clusters. We considered this number of clusters too low to represent adequately the complexity of the observed gravity field. Thus, we manually adjusted the bandwidth parameter  $h$  in equation 2.2 to increase the resolution of the kernel smoothing function, up to obtain a curve with 6 peaks (Figure 12b). According to our method, this number corresponds to the number of clusters and the corresponding values in the original map are taken as their centers (-81.46, -72.40, -63.35, -51.04, -42.63, -30.34 mGal). The resulting SGC terraced map, as well as the k-means terraced map assuming 6 clusters, are shown in Figure 12b,c. The k-means map has the following cluster centers: -80.09, -69.41, -62.38, -55.68, -48.10 and -39.66 mGal.

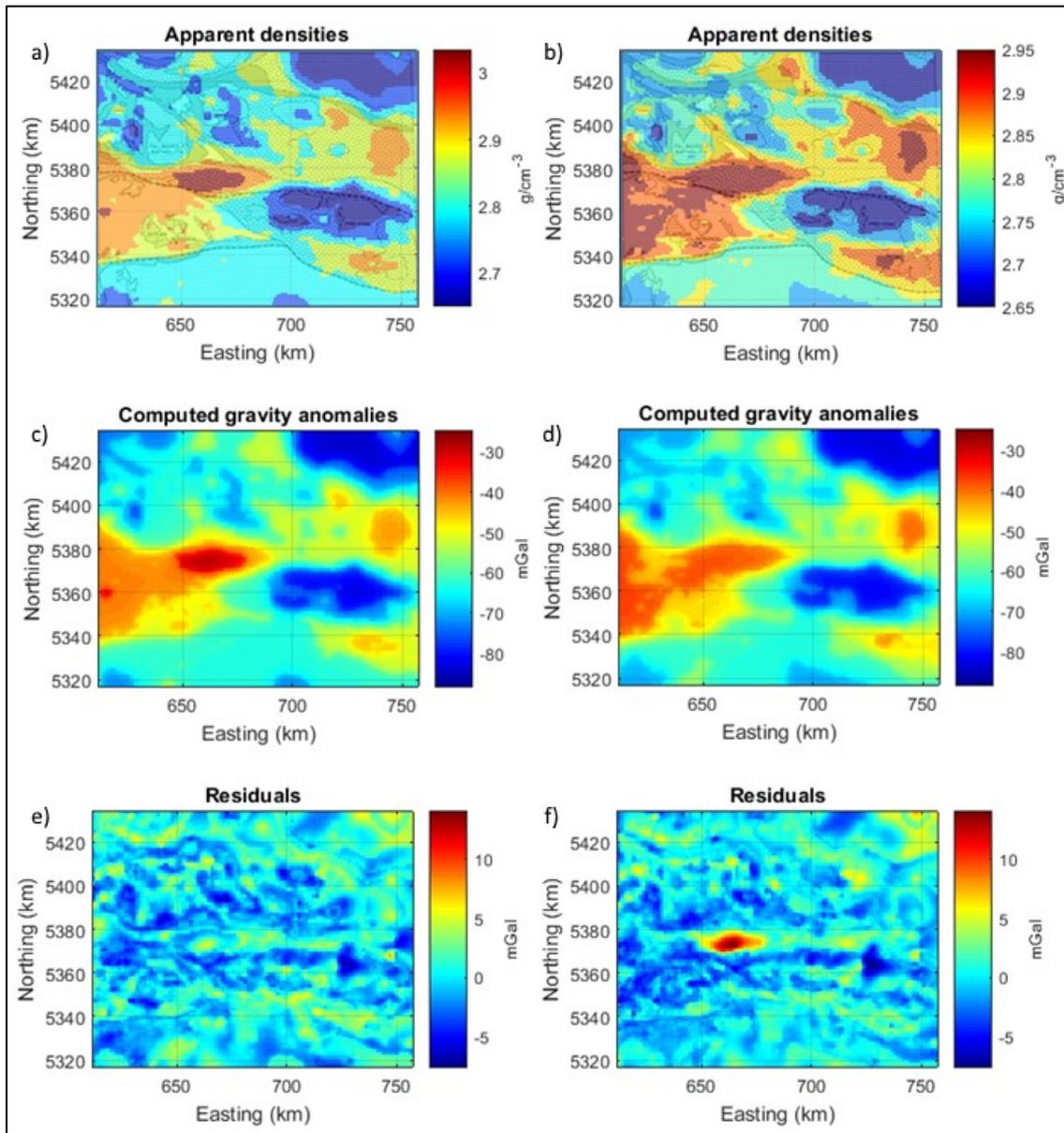


Figure 13 - Analysis of real gravity data of Noranda-Val-d'Or region (Canada). a) Apparent density map obtained from the SGC terraced map, overlain on a simplified geological map (after Keating, 1992). b) Apparent density map obtained from the k-means terraced map, overlain on a simplified geological map (after Keating, 1992). c) Gravity field computed from the apparent density map in a). d) Gravity field computed from the apparent density map in b). e) Residuals between the computed gravity field (c) and the observed gravity field (Figure 12a); f). Residuals between the computed gravity (d) and the observed gravity field (Figure 12a).

The study area is a plain with only small topographical undulations, so the density layer we used has a flat top (i.e., we assume outcropping sources) and a thickness of 4 km, consistently with the assumption of Keating (1992). As the gravity map presents only negative values, the obtained apparent densities are consistently negative. To obtain meaningful density values, we rescaled the densities by setting the minimum estimated value to zero and then summing to this rescaled density vector the minimum expected density value for the study area ( $2.65 \text{ g/cm}^3$ ; KEATING, 1992).

The resulting apparent densities derived from the SGC terraced map (Figure 13a) vary between 2.65 to 3.04 g/cm<sup>3</sup>, while those deriving from the k-means terraced map (Figure 13b) vary from 2.65 to 2.95 g/cm<sup>3</sup>. In both cases, the apparent density range is very similar to that found by **Keating (1992)**. It is also possible to appreciate as the geological map and the apparent density maps are similar: the lower densities are in correspondence to the granitic outcrops, while the highest values are correlated with volcanic rocks. Small variations in density not correlated with geology changes may be explained as the effect of the true depth extent of the geological body or as the effect of a depth to the top not corresponding to the surface, both situations different from the assumptions of the existence of a flat layer in the density modelling. The occurrence of vertical density variations unpredictable from the geologic map might also explain the presence of apparent density changes not correlated with geology changes.

To compare the gravity field obtained from these density distributions with the observed gravity field, we first need to compensate the constant shift between these two anomaly maps. To determine this value, we compute a least-squares regression between the computed and observed data. The slope of the best-fit line is 1, but the constant term results -530.6 mGal in the case of the SGC-derived data, and -529.23 mGal in the case of the k-means derived data. Optimally shifted by these constants, the gravity anomalies computed by the inverted density distributions (Figure 13c,d) compare well with the original field, with an rms value of 2.06 mGal in the SGC case and 2.22 mGal in the k-means case. The maps of the residuals (Figure 13e,f) shows some correlations with the original anomalies (e.g., the arcuate feature at the extreme NW of the maps) and, more generally, shows that many high wavenumber features of the observed field cannot be adequately modelled by the simple apparent density map.

Also in this case, we tried to combine the benefits of the clustering and Kuwahara filtering, but the results did not improve the final misfit between the observed and computed data.

## **2.4. CONCLUSIONS**

The aim of terracing potential fields is to obtain a simplification of the original map that should ideally present only a few areas of constant values, better relatable with geological variations than the original, smoother, map. We introduced clustering techniques in this kind of potential field

map transformations. We showed as a simple reclassification of the field values based on the minimum Euclidean distance from a set of cluster centers, whose number and center value can be automatically determined, (Simple Guided Clustering, SGC) or the classical k-means method can be used to produce optimally terraced maps. The number of clusters can be manually selected to obtain a final map with the desired level of simplification (the lower the number of clusters and the greater will be the simplification). This feature of clustering methods offers the possibility of obtaining a terraced map having different, constant, field values in each terrace and without smooth transitions between terraces. In addition, blocky features in the terraced map, as produced by other algorithms, are totally absent using our method. Starting from a terraced map, an apparent physical property horizontal distribution map can be easily computed. This map will show abrupt variations resembling what is generally displayed in a geological map. We tried to further enhance the abrupt character of the found density variations by combining the clustering and the Kuwahara filtering. The application of clustering algorithms to synthetic as well as real data from a mining area in Canada, proved their effectiveness.

## CHAPTER 3 – Sharp-Edged Inversion

*In this chapter a strategy to retrieve a geological-like output from potential field inversion is described. The strategy is built on an application of guided fuzzy c-means clustering in an iterative process, where the constraints are related to the petrological characteristics of the media investigated. Synthetic and real cases are shown and discussed.*

The interpretation of potential fields often involves their data inversion to obtain models describing the distribution of physical properties in the subsurface. However, a careful analysis of the output models is indispensable to extract some geological information from the geophysical model. This analysis is often made in cascade as a separate stage, especially when the inversion is carried out by minimization of an objective function promoting structural smoothness in each spatial direction, as most algorithms in use today. The output is often a model where the inverted parameter smoothly varies across the volume and the recognition of geological features is related to external information as well as the experience of the interpreter. This way to extract some geological information may be too subjective and ambiguous.

By application of some mathematical filters, the output model might be characterized by abrupt variations of the investigated parameter but, in general, the resulting model produces a dataset that do not adequately fit the observed one, thus it has a limited geophysical significance. To overcome this problem, during the last twenty years several authors developed methods and strategies to find inversion models realistically reproducing the expected geological features (PORTNIAGUINE AND ZHDANOV, 1999; CAMACHO ET AL., 2000; LANE ET AL., 2007; BERRINO AND CAMACHO, 2008; FARQUHARSON ET AL., 2008; LELIÈVRE, 2009; SUN AND LI, 2010; PHILLIPS AND SIMPSON, 2015; SUN AND LI, 2015). All these methods try to find a solution where the sources are well defined in their volumes, so that the geological interpretation of the model should result easy, when not straightforward.

In 1999, Portniaguine and Zhdanov proposed an iterative inversion technique, called focusing, that favors the compactness of solution. This method is often used for isolated sources, like in mining and ore exploration. Camacho et al. (2000) proposed the method of growing bodies, where a starting elementary volume, having an a-priori assumed density, grows in the 3D space until the produced anomaly is as close as possible to the observed one. Also this method finds most of

applications in the definition of isolated sources. In 2008, **Berrino and Camacho** studied a way to produce a layered model as extension of growing bodies method, where a distribution of preselected layers of assigned density is deformed to produce a field approximating the observed one. In the inversion proposed by **Phillips and Simpson (2015)**, at each iteration is computed the “terraced” version of the solution. With the terracing procedure (e.g. **CORDELL AND McCAFFERTY, 1989**), the model is simplified presenting only a few areas of constant values. In this way, it should result easier to relate the physical property variations in the subsurface with the geological variations. The multidomain petrophysically constrained inversion (**SUN AND Li, 2015**), integrates in the algorithm a clustering technique to constrain the solution toward a prefixed number of rock property values corresponding to distinct geological domains.

This last method inspired a new strategy, proposed in this thesis, to produce geophysical models easily relatable to the subsurface geological and structural setting. This inversion process is aimed at producing geophysical models characterized by the presence of a few layered homogeneous volumes corresponding to different geological bodies. In practice, the aim is to carry out a geophysical inversion integrating as much as geological information as possible with the aim to obtain valid solution in both geophysical and geological point of view.

The tools used in this strategy are essentially the **Li and Oldenburg** algorithms (2003) as solver of inversion problem and the guided fuzzy c-means clustering to incorporate the geological constraints. This inversion strategy will be discussed in the subchapter 3.3.

### 3.1. LI AND OLDENBURG ALGORITHM

The **Li and Oldenburg (2003)** algorithm allows to solve underdetermined problems using an objective function that has the flexibility to construct many different models, applying appropriate weighting functions whose parameters are empirically selected.

The solution is obtained by the following minimization problem (**THIKONOV AND ARSENIN, 1977**):

$$\text{minimize } \phi = \phi_d + \mu\phi_m \tag{3.1}$$

where  $\Phi_d$  is the data misfit, and measures the misfit between the observed data and data predicted by a candidate model,  $\Phi_m$  is a regularization term that measures the structural complexity of an inverted model and  $\mu$  (regularization parameter) determines the relative importance of these 2 terms. When  $\mu$  approaches zero the function to be minimized consists only of the data misfit ( $\Phi_d$ ), and the model constructed will fit the data as well as possible, irrespective of the amount of structure that is built into the model.

The data misfit can be defined by (LI AND OLDENBURG, 1998):

$$\phi_d = \|\mathbf{W}_d(\mathbf{d} - \mathbf{d}_{obs})\|^2 \quad (3.2)$$

where  $\mathbf{d}$  are the predicted data-vectors,  $\mathbf{d}_{obs}$  are the observed data-vectors and  $\mathbf{W}_d$  is the inverse data covariance matrix, a diagonal matrix containing the standard deviations related to each data.

The structural complexity can be defined by (LI AND OLDENBURG, 1998):

$$\phi_m = \alpha_s \|\mathbf{W}_s(\mathbf{m} - \mathbf{m}_{ref})\|_p^p + \alpha_m \|(\mathbf{W}_m \mathbf{m})\|_p^p \quad (3.3)$$

$\alpha_s$  and  $\alpha_m$  are coefficients that ensure relative importance among the different terms of the functions,  $\mathbf{m}_{ref}$  is a reference model that can include several constraints related to the distribution of investigated physical property and  $\mathbf{m}$  is the solution model. The  $\mathbf{W}_s$  term control the closeness of the solution to the reference model, and the term  $\mathbf{W}_m$  is a regularization matrix that allows increasing or reducing the gradient of the modelled quantity in the 3 directions. This matrix results useful to constrains the solution when a priori knowledge about distribution of investigated parameter are known.

The extended form of the equation 3.3 is (OLDENBURG AND LI, 2005):

$$\begin{aligned} \phi_m(m) = & \alpha_s \int_V w^2(z)(\mathbf{m} - \mathbf{m}_{ref})^2 dv + \\ & \alpha_x \int_V \left( \frac{\partial w(z)(\mathbf{m} - \mathbf{m}_{ref})}{\partial x} \right)^2 dv + \\ & \alpha_y \int_V \left( \frac{\partial w(z)(\mathbf{m} - \mathbf{m}_{ref})}{\partial y} \right)^2 dv + \end{aligned}$$

$$\alpha_z \int_V \left( \frac{\partial w(z)(\mathbf{m} - \mathbf{m}_{ref})}{\partial z} \right)^2 dv \quad (3.4)$$

where  $\alpha_m$  is different for the three directions ( $\alpha_x, \alpha_y, \alpha_z$ ) and  $w(z)$  is the depth weighting function (LI AND OLDENBURG, 1996):

$$w(z) = \frac{1}{(z + z_0)^{\beta/2}} \quad (3.5)$$

where  $z$  is the depth of each layer in the 3D model and the value of  $z_0$  depends upon the observation height and cell size. Li and Oldenburg propose to use for  $\beta$  a value equal to 3 in the magnetic case (LI AND OLDENBURG, 1996) and equal to 2 in the gravity case (LI AND OLDENBURG, 1998), assuming 3 and 2 as the rate decay of the magnetic or gravity field of a single, small, cubic cell. Oldenburg and Li (2005) later suggested that the exponent value used in a particular inversion could be chosen, by finding the “best performance” of different exponent values applied to trial inversions of synthetic data from forward models similar to the expected solution.

Cella and Fedi (2012) showed instead that the appropriate value of  $\beta$  must be related to  $N$ , the structural index of the source (Table 3.1), rather than to the power-law decay of the field generated by a single cell. Estimation of the structural index can be made with standard methods such as Euler Deconvolution or the study of the scaling function (FEDI, 2007; FLORIO ET AL., 2009).

SOURCE SHAPE	N (GRAV)	N (MAG)
Sphere	2	3
Cylinder	1	2
Sill / Dike	0	1
Contact	-1	0

Table 1- Structural index.

**Ialongo (2013)** tested that the introduction of constraints, related to the 3D boundaries of the source, makes the information relative to depth-weighting ( $\beta$ ) not so crucial for correctly reconstructing the source features.

The role of  $\alpha_s$  is provide greater emphasis on closeness to a reference model ( $m_{ref}$ ), while  $\alpha_x, \alpha_y, \alpha_z$  penalizes first order roughness in the x, y and z directions and its role is to keep the constructed model as flat as possible. Small value of  $\alpha_x, \alpha_y, \alpha_z$  drive the solution toward the reference model ( $m_{ref}$ ).

A good setting of parameters can guarantee, at the end of the iterative process, a smooth model producing a field that fits adequately the observed data.

In Figure 14 the results of an inversion of synthetic gravity data obtained by two square prismatic sources are shown (Figure 14c). The sources have sides of 400 m in X and Z (depth) and are infinite in Y direction, with a centre placed at 300 m depth and a positive density contrast of 0.5 g/cm<sup>3</sup> with respect to the background. The data are calculated (Figure 14a) in 96 stations, equally spaced each 25 m, from the retrieved model (Figure 14b) and shows a perfect fit with the observed data, although the model is not properly fitting the original one.

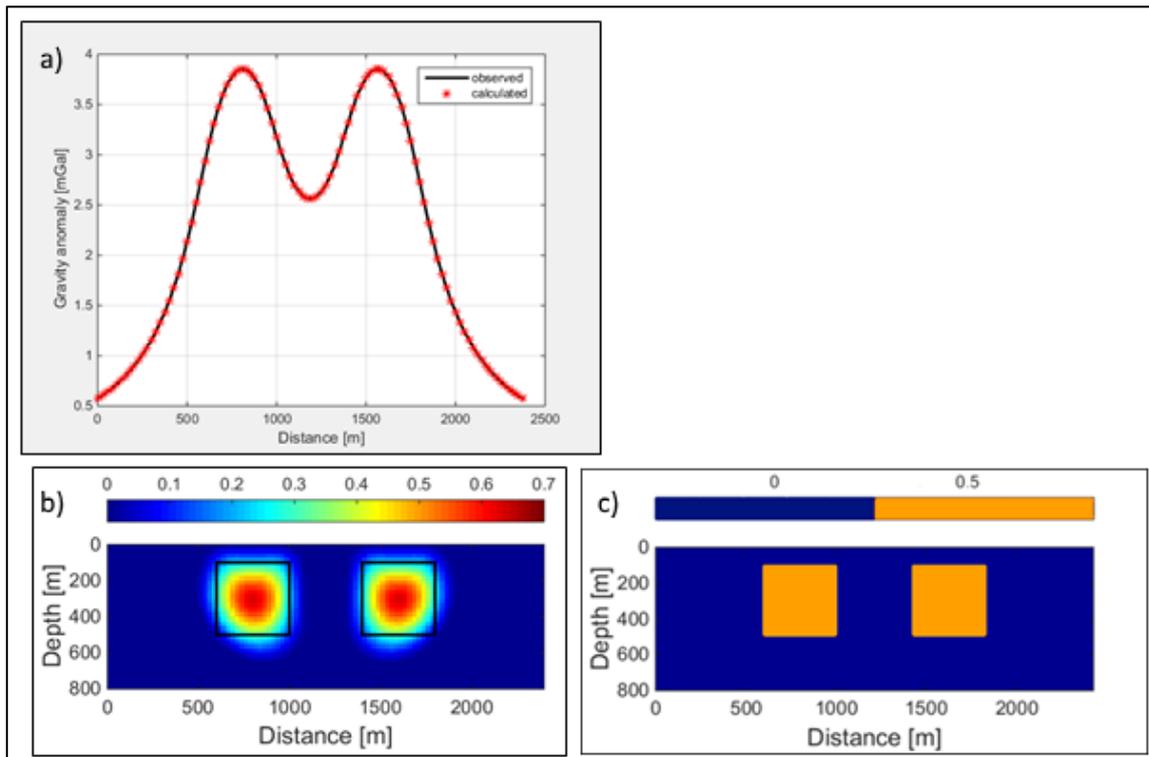


Figure 14 - b) Model cross-section obtained using Li and Oldenburg algorithm and a) comparison of observed, solid line, and calculated data, red asterisks. b) Black lines are the trace of the c) real source extension of two boxes with a positive density contrast of  $0.5 \text{ g/cm}^3$  respect to the background. Colorbar are in  $\text{g/cm}^3$ .

In the case shown, the inversion was run by setting the weights without forcing any particular directional gradient ( $\alpha_x = \alpha_y = \alpha_z = 1$ ) with a constant reference model ( $\mathbf{m}_{ref}$ ) made by zeros,  $\alpha_s = 1e-4$ ,  $\beta = 2$  and  $\mu = 1e-9$ .

Within the obtained model, the sources are compact and although well-defined under geophysical point of view, their density slowly fades away from a dense center. This implies the need for an interpretation stage to identify the boundary position of each source and making the model well-defined under a geometrical and geological point of view.

The choice to use Li and Oldenburg algorithm is related to its flexibility, especially in the constraints definition, ensuring the possibility to be used in an extensive list of geological problems.

To overcome the ambiguous definition of source boundaries (Figure 14b) and to defining them with an objective procedure, we will adopt a three-stages inversion procedure, using clustering

algorithms to classify the model solution in different units according to petrological and petrophysical knowledges.

### 3.2. GUIDED FUZZY C-MEANS

Given a certain number of values, as the parameter distribution within the inversion solution  $m$ , by the minimizing of the following function it is possible to classify the values in to a certain number of groups (SUN AND LI, 2015):

$$\phi_{FCM} = \sum_{j=1}^M \sum_{k=1}^C u_{ij}^q \|\mathbf{m}_j - \mathbf{v}_k\|_2^2 + \eta \sum_{k=1}^C \|\mathbf{v}_k - \mathbf{t}_k\|_2^2 \quad (3.6)$$

where the first term represents the classic fuzzy c-means cluster technique (Bezdek, 1981),  $C$  represents the number of geological units, the parameter  $u_{jk}$ , called membership, quantifies the degree to which each cells ( $M$ ) of the model ( $\mathbf{m}_j$ ) belongs to the geological units, represented by cluster centres  $\mathbf{v}_k$ . The second term guides the choice of the cluster centres (SUN AND LI, 2015) toward the a priori values  $\mathbf{t}_k$  according to a weight  $\eta$ . A small value to  $\eta$  should be assigned for poorly known a priori  $\mathbf{t}_k$  values, and large values for highly reliable a priori information. A value of 0 should be assigned when no a priori cluster centre value is known at all. This equation can be applied to the problem of the geological differentiation of an inverted geophysical model.

The membership value is constrained to be nonnegative and to satisfy the relation:

$$\sum_{k=1}^C u_{jk} = 1 \quad (3.7)$$

The parameter  $q$ , in equation 3.6, controls the degree of fuzziness of the resulting membership functions and satisfies  $q \geq 1$ . If  $q$  is set to 1 each object can only be assigned to a single cluster, as opposed to fuzzy clustering in which each object can belong to multiple clusters with differing membership values. In many applications  $q = 2$  is a good compromise when the total number of clusters  $C$  is known (HATHAWAY AND BEZDEK, 2001).

By assigning each model cell to the cluster with the highest probability, the fuzzy partition is reduced to a hard classification. This process is sometimes referred to as defuzzification (PHAM, 2001).

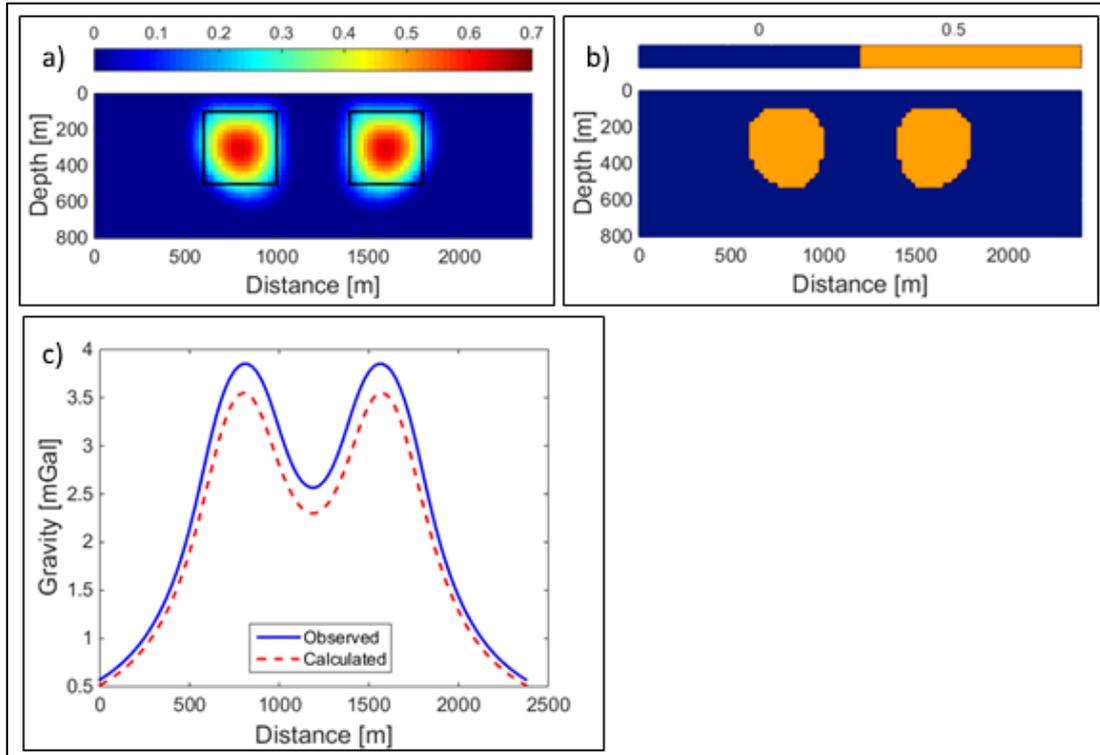


Figure 15 - a) Density distribution model. b) Defuzzified model obtained by the application of guided fuzzy c-means clustering, using 2 classes with defined centre values. c) comparison of observed data (solid line) and data calculated from the model in b) (dashed line).

An example of application of guided fuzzy c-means clustering and its defuzzification is shown in Figure 15. The density distribution (Figure 15a) is clustered into 2 ( $C$ ) classes defined by values  $t_k = [0, 0.5]$ , considered as highly reliable ( $\eta = 1e5$ ). As an effect of the clustering process, each cell ( $M = 32 \times 96$ ) of the model has a value of membership  $u_{jk}$  for each cluster considered. The defuzzified model (Figure 15b) is the result of assigning to each model cell the cluster with highest membership.

The case shown above represents a solution to the problem of classifying subsurface regions into different geologic units (i.e., geology differentiation), which is normally performed after the geophysical inversion.

The resulting model (Figure 15b), where the sources are well separated from the background, can be seen as a good geological model, but as shown by the red dashed line in Figure 15c the data produced is not satisfactorily fitting the data produced by the smooth model Figure 15a, which perfectly reproduces the observed data.

**Sun and Li (2015)** proposed a method to integrate the guided fuzzy c-means clustering into the objective function of the inverse problem:

$$\Phi(m; u_{jk}, v_k) = \Phi_d(m) + \mu\Phi_m(m) + \lambda \sum_{j=1}^M \sum_{k=1}^C u_{jk}^q \|m_j - t_k\|_2^2 \quad (3.8).$$

In the strategy adopted in this thesis the inverse problem is not constrained by the clustering minimizing a single objective function, but as explained in the following subchapter (3.3) the final output is obtained after three different phases, to ensure a better control of the whole inversion process.

### 3.3. SHARP-EDGED INVERSION STRATEGY

In this Section a procedure to constrain the geophysical inversion aiming at obtaining a valid solution from both the geophysical and geological points of view will be described.

The strategy, that we call Sharp-Edged Inversion, has three steps (Figure 16). The first one is a smooth inversion, by the minimization of the objective function described in equation 3.1. Then, the obtained solution is clustered by equation 3.6 to define weights and the regularization matrix in equations 3.3 and 3.4. The last step is a constrained weighted-inversion according to the previously defined weights.

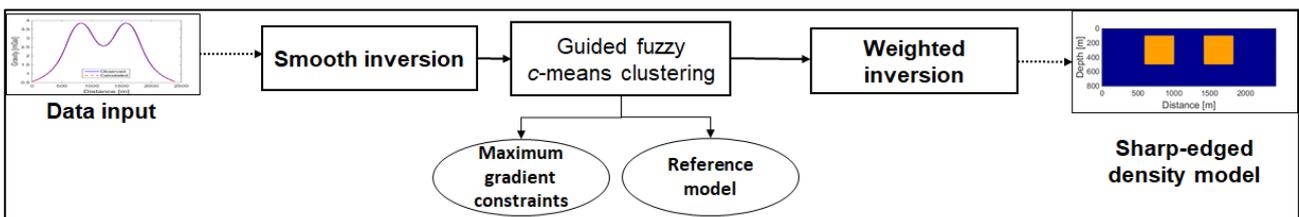


Figure 16 Sharp-Edged Inversion strategy workflow.

The smooth inversion step is based on the expectation that the solution to the inverse problem is “simple”, where the notion of simplicity is quantified by the Euclidean length of the solution. This assumption tends to promote a structural smoothness in each spatial direction while fitting the observed data. In this stage all the information related to a-priori knowledge can be set, such as particular distribution of the investigated parameter (e.g. from wells), source’s boundaries (derived from the data analysis) or a reference model that include geological knowledges. To promote the structural smoothness in this stage, the external information is added as a reference model accompanied with an optimal regularization parameter ( $\mu$ ) that produce a data misfit ( $\Phi_d$ ) consistent with the error and noise in data (often with a feature of high frequencies). The reference model can be built by integrating available data constraining some parts of the volume, like wells or geophysical and/or geological cross-section.

The values of the model obtained in the first step are clustered by the fuzzy c-means method. This clustered model is then used in a second inversion step. The clustering helps to define constraints representing volumes of the model where to force strong density variations. The clusters are created starting from external knowledge about the physical properties in the subsurface geology. The choice of the clusters number has to be selected according to the geological features considered as important to investigate and taking into account the potential field data used and its limitation. In fact, if under a geological point of view two different formations could be defined because of a change of fossils content inside the same sedimentological and mineralogical texture, the density of these two formations can be similar, with a very low geophysical contrast, which makes the gravity inversion unable to detect them as two distinct units. In addition, the reliability of the external information related to the cluster values ( $\mathbf{t}_k$ ) has to be evaluated, and associated with a correct weight ( $\eta$ ). If there is no information about the values of the density at depth, it is possible to choose the number of desired clusters ( $C$ ) and set  $\eta = 0$ , so that cluster centres ( $\mathbf{v}_k$ ) are automatically computed based on the representative values inside the data distribution according to the Euclidean distance. Moreover, when the uncertainty in a priori knowledge about density is not similar for each class,  $\eta$  can be different for each value of  $\mathbf{t}_k$ .

The second step of the Sharp-Edged Inversion strategy is based on the guided fuzzy c-means clustering of the density distribution. Of course, any other cluster method can be used but, thanks

to the membership values ( $u_{jk}$ ) associated to each cell, guided fuzzy c-means clustering allow us a control in the classification of each cell in a cluster. In fact, clustering methods (see Chapter 2) often are based on distance between centers and data. Membership must satisfy equation 3.7 and it means that in some cases the membership value associated to a data can be the same for more than one cluster. The possibility to investigate in a separate stage the distribution of membership across the volume can be a good way to assign different threshold to move some cell to another cluster when the membership is equal between two clusters.

In the third step of the inversion strategy, is used the defuzzified model obtained after the clustering of the smooth inversion model. As shown in subchapter 3.4 and 3.6, this model can be used to constrain and weight the inversion in different ways. The simplest way is to use it as a reference model, as done for the 3D application later described (in subchapter 3.6) by setting appropriate values for  $\alpha_s$  and  $\alpha_m$  parameters. Another way, tested in the 2D case (subchapter 3.4) is to use the sources boundaries (Figure 17), easily obtained by the clustered model, to create a regularization matrix ( $W_m$ ; **OLDENBURG AND LI, 1994**):

$$W_m^T W_m = \alpha_s (S_s D_s Z)^T (S_s D_s Z) + \alpha_x (S_x D_x Z)^T (S_x D_x Z) + \alpha_y (S_y D_y Z)^T (S_y D_y Z) + \alpha_z (S_z D_z Z)^T (S_z D_z Z) \quad (3.9)$$

where  $S_i$  ( $i=s,x,y,z$ ) are diagonal matrices whose elements  $w_i$  ( $i=s,x,y,z$ ) are spatially dependent weighting functions to input additional prior information about density or susceptibility model. The function  $w_s$  controls the relative closeness of the final model to the reference model at any location. In fact, different parts of the reference model can be more or less constrained by external data. The weights  $w_i$  ( $i= x,y,z$ ), with or without a reference model, control the degree of smoothness of the solution along the three directions (**OLDENBURG AND LI, 2005**). In equation 3.9  $D_s$  has elements  $\sqrt{\Delta x \Delta y \Delta z}$  on its diagonal, where  $\Delta x$ ,  $\Delta y$  and  $\Delta z$  are the unit cell dimensions in each direction.  $D_i$  ( $i= x,y,z$ ) are matrices representing the finite difference operator along the three spatial directions.  $Z$  is a diagonal matrix that represent the depth weighting function described in equation 3.5.

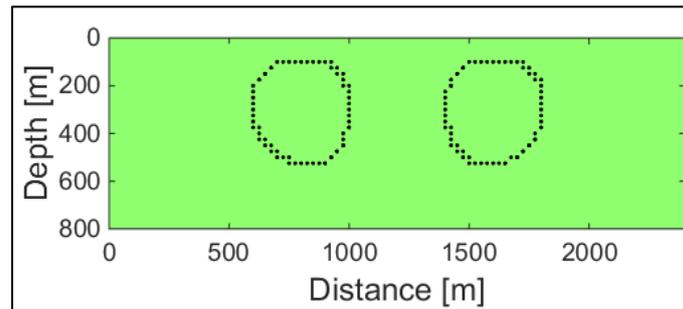


Figure 17 - Sources boundary of the defuzzified model in Figure 15c.

Once the weight parameters are defined, the inversion can be run (third phase of the strategy) and the final output solution is obtained. In this thesis, the aim of the inversion is to obtain a valid representation of the physical property (density or magnetization) distribution from both the geophysical and geological points of view, i.e. a model with sharp boundaries, able to fit the observed data (Figure 18b).

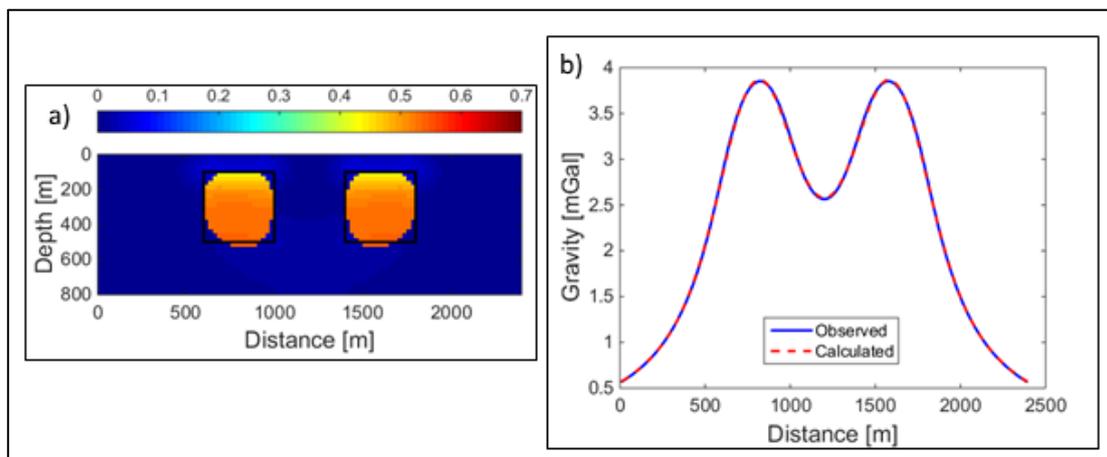


Figure 18 - a) Final density distribution from the third step of the Sharp-Edged Inversion strategy, and b) comparison of observed, solid line, and calculated data, dashed.

The inversion strategy explained above is now applied in different real cases and results are discussed in the following subchapter.

### 3.4. A 2-D REAL CASE (PIANO DI PECORE)

A microgravity survey was completed to precisely locate and better characterize the near-surface geometry of a recent fault with small throw (LO RE ET AL., 2016) in a mountainous area in the Southern Apennines (Italy). The site is on a segment of the Irpinia fault, which is the source of the  $M_w$  6.9 1980 earthquake. This fault cuts a few meters of Mesozoic carbonate bedrock and its younger, mostly Holocene continental deposits cover. The amplitude of the complete Bouguer anomaly along a 150m profile is about 40  $\mu\text{Gal}$  (Figure 19). Following the Sharp-Edged Inversion strategy, the fault has been clearly identified and localized in its horizontal position and depth.

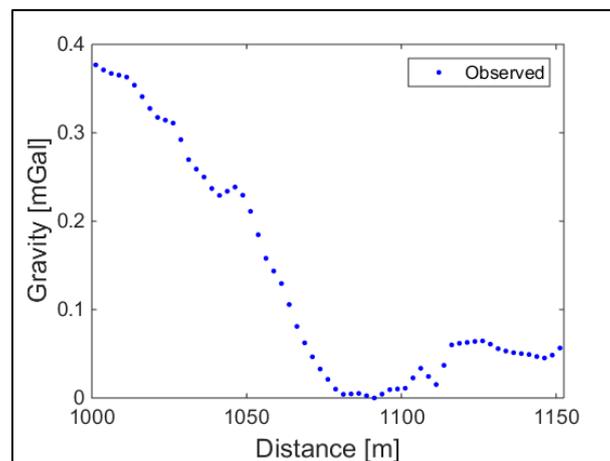


Figure 19 - Complete Bouguer anomaly.

#### 3.4.1. Geological setting

The studied area is a small karst basin (named “Piano di Pecore”) partly filled-in with colluvial and fluvio-palustrine deposits (ASCIONE ET AL., 2003) nested on top of Monte Marzano, a mountainous ridge of the axial part of the Southern Apennines, Italy (Figure 20a and c). The outcropping lithologies in the area (Figure 20b) are colluvial and fluvio-palustrine deposits, debris slope deposit, and weathered and fractured Mesozoic limestone. The basin is underlain by a segment of the Irpinia fault, which slipped during the devastating  $M_w=6.9$ , 1980 earthquake causing the longest (nearly 40 km) albeit discontinuous co-seismic surface break so far observed in Italy (PANTOSTI AND VALENSISE, 1990). Co-seismic rupture observations showed that the fault is segmented, and paleo-seismological investigations have suggested a 0.4–0.6 mm/yr extension rate (DISS WORKING GROUP, 2010).

In Piano di Pecore basin, the morphological expression of the displacement generated by the 1980 earthquake has been recognized as a sharp flexure in the basin fill, which outlines the presence of a blind segment of the fault immediately beneath (CINQUE ET AL., 1981; PANTOSTI ET AL., 1993). Previous studies, like paleoseismic trenches (D'ADDEZIO ET AL., 1991; PANTOSTI ET AL., 1993) and other geophysical surveys (IMPROTA ET AL., 2003; GALLI ET AL., 2014), showed the top of the fault surface at a depth of few tens of meters within the Mesozoic carbonate bedrock.

The geological setting of the area is relatively complex because the Irpinia fault crosses the Piano di Pecore basin with a bend relative to its regional trend (CINQUE ET AL., 1981). Being the fault buried by more recent sediments, its surficial expression is related to sediments morphology, deformed to accommodate the stress generated by the 1980 earthquake reactivation of the fault (PANTOSTI ET AL., 1993). The study of the stratigraphy observed in trenches excavated normal to the fault plane shows that the fault is younger than 700 ky (D'ADDEZIO ET AL., 1991; PANTOSTI ET AL., 1993). Moreover, the deformed sediments record at least 5 earthquake events in the last 9000 years, with a vertical slip rate per event between 40 and 70 cm (PANTOSTI ET AL., 1993).

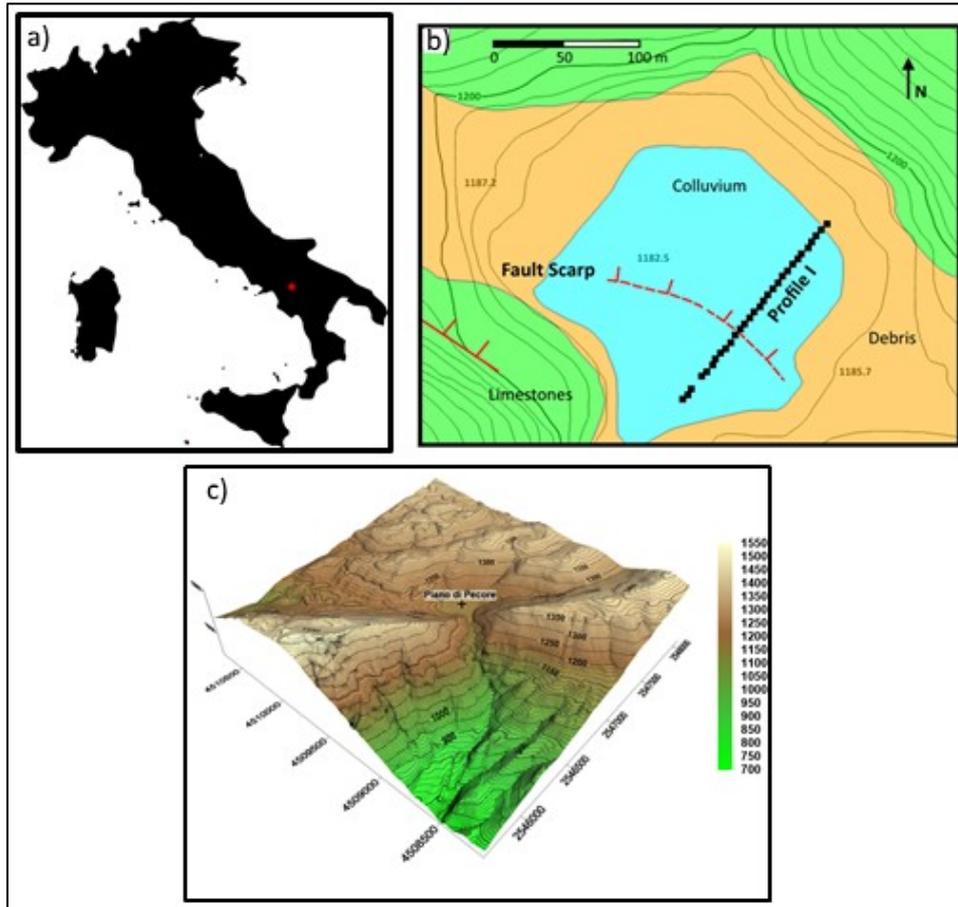


Figure 20 - a) The studied area in Southern Italy is the red asterisk. b) A detail of the area and its geological sketch map with gravity station represented by black squares. Green is Mesozoic limestone, orange is debris slope, light blue is colluvium, and red line is the fault trace, solid when exposed and dashed in the basin, where it is marked by a morphological step on the ground (modified after Improta et al., 2003). c) 3D visualization of the detailed DEM (2 x 2 km). Cell size 5 m. Piano di Pecore basin is in the centre. Bold line contours every 50 m. Projection: Rome40 / Italy zone 2 (EPSG 3004).

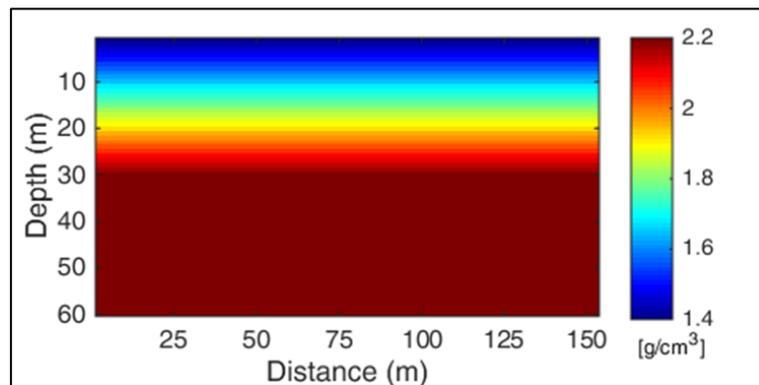
### 3.4.2. Data inversion and results

The subsurface has been discretized in the horizontal direction into 63 cells. The cells have a size of 2.5 m in the horizontal direction and are 2 m thick. The model space is extended up to 60 m depth, with 30 cell layers of the same dimensions. The model is 2-dimensional, so that cells are infinitely extended in the direction perpendicular to the profiles. In order to reduce distortions generated by the finite horizontal dimension of the model at the edges of the profiles, 2 cells having horizontal dimension of 100 m have been included outside each edge of the profile.

Scalfun analysis (FEDI AND FLORIO, 2006; FLORIO ET AL., 2009) was performed on the data along the profile. The scaling function becomes constant if a depth to source of 30 m is assumed, and the estimated structural index  $N=1$ . This value is in the expected range for a limited throw fault,

indicating a very small throw, so that the depth estimate should correspond to the top of the fault footwall (LO RE ET AL., 2016).

The information obtained by the Scalfun analysis, as well as that coming from previous studies (ASCIONE ET AL., 2003; PANTOSTI ET AL., 1993; GALLI ET AL., 2014) have allowed the definition of a reference model ( $m_{ref}$ ) of the density distribution at depth to be used during the inversion process (equation 3.3) during the first step of the Sharp-Edged Inversion strategy. To recover the density values related to the subsurface, the Gardner's relation (GARDNER ET AL., 1974) was applied to the seismic velocities found in a seismic tomography model in the area (IMPROTA ET AL., 2003). Thus, the reference model (Figure 21) consists in a linear density increase with depth, from the surface to 30 m depth, having a minimum value at surface ( $\rho_{min} = 1.4 \text{ g/cm}^3$ ) and a maximum value at 30 m ( $\rho_{max} = 2.2 \text{ g/cm}^3$ ). For depths greater than 30 m the density is constant, equal to  $2.2 \text{ g/cm}^3$ , as assumption of the depth estimate from Scalfun analysis relative to the top of the footwall.

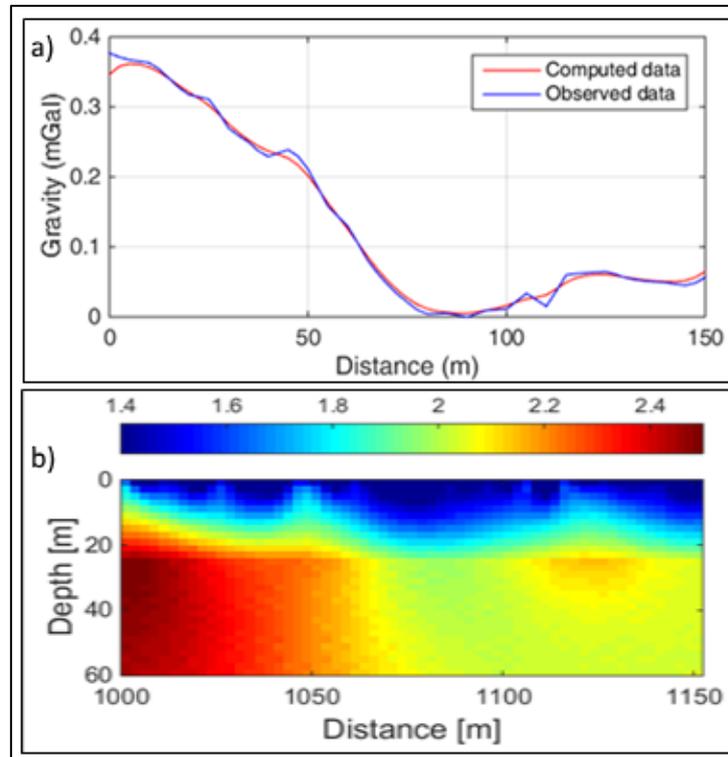


*Figure 21 - Reference model adopted in the inversion, during the first step of the Sharp-Edged Inversion strategy*

The data show minor oscillations at  $x = 1020, 1045$  and  $1110 \text{ m}$  (Figure 19). With the aim to focus the model on the fault structure, we would like to obtain a model of the density distribution where these high wavenumber signals, presumably coming from 1) errors in the measurement process; 2) the data treatment to compute the complete Bouguer anomalies, and 3) very shallow structures, not sufficiently well sampled (what we could refer to as 'geologic noise'). Thus, analysis of data may probably benefit of some sort of low-pass filtering process.

However, instead of performing some low-pass filtering to the data, in this case was preferred to opportunely set the regularization parameter  $\mu$ . In fact, the value of this parameter influences the way the predicted data fit the observed data (equation 3.1). After some tests, was selected a value

of  $\mu = 1e-7$ , a value that deliberately produces a density distribution that do not account for the high-frequency oscillations of the analysed gravity field, which in this way are considered as “noise”. In practice, the regularization parameter  $\mu$  was used as a “low-pass filter” applied to the model of density distribution.



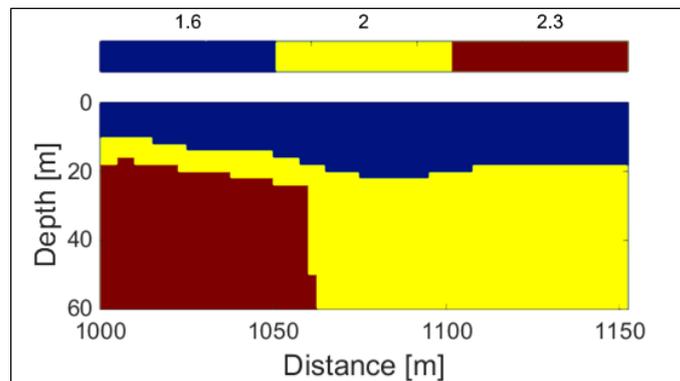
*Figure 22 - a) comparison of observed, blue line, and calculated data, red line, by the b) density distribution model obtained after the smooth inversion, first step of the Sharp-Edged Inversion strategy. Colorbar is in g/cm<sup>3</sup>.*

The result of the first inversion step (Figure 22) reflects a good geophysical response of the model, detectable on the data misfit (Figure 22a), but the density distribution within the cross-section does not allow a simple geological interpretation. In fact, there are no abrupt variations in density in the two dimensions (depth and distance), useful to recognize the formation’s limits and to place a fault in an unambiguous way.

The second step consisted in the classification of the model into three different units. The number and the value of the different classes derived from the analysis of different a-priori knowledges. The outcropping geology (Figure 20b) was taken into account to define the number of classes (3), identifiable essentially in alluvium-colluvium deposits, a fractured limestone unit and a debris unit derived from the limestone itself. To recover the density values related to these classes, the

Gardner's relation ([GARDNER ET AL., 1974](#)) was applied to the seismic velocities related to a seismic tomography model in the area ([IMPROTA ET AL., 2003](#)).

The defuzzified output is shown in Figure 23. Starting from the surface, the cluster centres are associated to different geological "units", a first low-density body ( $1.6 \text{ g/cm}^3$ ), attributable to Holocene alluvium-colluvium and palustrine sediments. The first unit mantle a second one, of higher density (cluster centre  $2 \text{ g/cm}^3$ ), which is attributable either to extremely karstified and faulted bedrock limestone, and/or to cemented carbonate breccias. These breccias are often found at the foot of cliffs in carbonate massifs of the southern Apennines ([BRANCACCIO, 1968](#)) and are common at Monte Marzano too. Finally, the deeper unit, with density of about  $2.3 \text{ g/cm}^3$ , corresponds to the fractured carbonate bedrock.



*Figure 23 - Defuzzified output after the clusterization of the model in Figure 22. The cluster centre are 1.8, 2 and 2.3  $\text{g/cm}^3$ .*

The interfaces created between the three different reconstructed units are used to build the  $S_i$  matrices in equation 3.9 to force strong density contrast in x and z directions along these surfaces. The last step of the Sharp-Edged Inversion strategy was launched, and the final model was obtained (Figure 24).

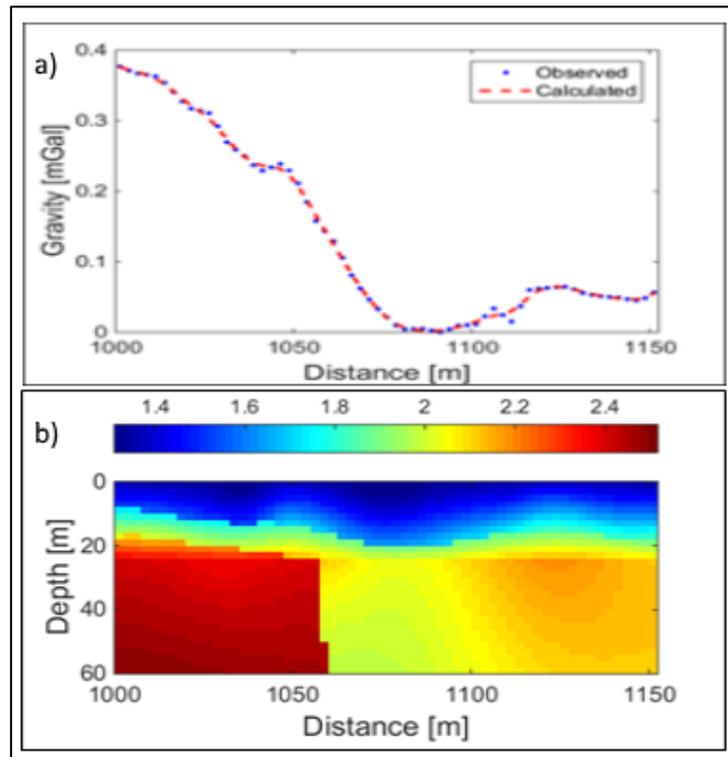


Figure 24 - a) comparison of observed data (dots) and calculated data (dashed line) produced by the Sharp-Edged model (b) obtained in the last step of the inversion procedure.

In the last step a smaller value of  $\mu$  was chosen ( $\mu=1e-10$ ), to ensure the fitting with some, but not all, high frequencies, and according to its effect of low-pass filter, the data calculated by the final output (red dashed line in Figure 24a) fit properly the observed data. Thus, the final model has a considerable significance from geophysical point of view, meanwhile is strictly connected with some geological features. In the depth direction (Figure 24b), is it possible to detect a first discontinuity between two homogeneous bodies. The top one (blue) is representative of the Holocene alluvium-colluvium and palustrine sediments, while the second one (yellow) is referred to the cemented carbonate breccias. The formation's limit between this two units results not very distinct where the data misfit is bad ( $1040 < x < 1050$  meters and  $1100 < x < 1115$  meters) due to the high frequencies not retrieved by the computed data. The red body in the model corresponds to the fractured carbonate bedrock and it is well confined. Its spatial relationship with the yellow body expresses the presence of a fault, segment of the Irpinia Fault system.

The quite homogeneous different bodies represented (in Figure 24b) make the geological interpretation easier, fulfilling the aim of the Sharp-Edged Inversion strategy.

### 3.5. 3D REAL CASE (SICILY CHANNEL)

The Sharp-Edged Inversion strategy is finally tested on a gravity dataset (CARROZZO ET AL., 1991, Figure 25b) relative to an area of the south-westernmost part of Italy, in a zone including the on-land side of Sicily and its offshore (Figure 25a). Tectonically, the area, that includes a portion of the African-European plate boundary, is complex, where different unrelated geodynamic processes such as rifting and mountain building interact at the same time. The amplitude of the residual Bouguer anomalies, obtained removing a linear planar trend, along a 65 x 58 km map is about 30 mGal (figure 25c). The main maxima of the gravity anomaly map can be associated to different known geological structures in a layered density distribution model, as described by the interpretation of seismic data and a set of wells. The gravity data inversion aims at defining the structure and the spatial continuity of the main geological horizons up to 6 km depth.

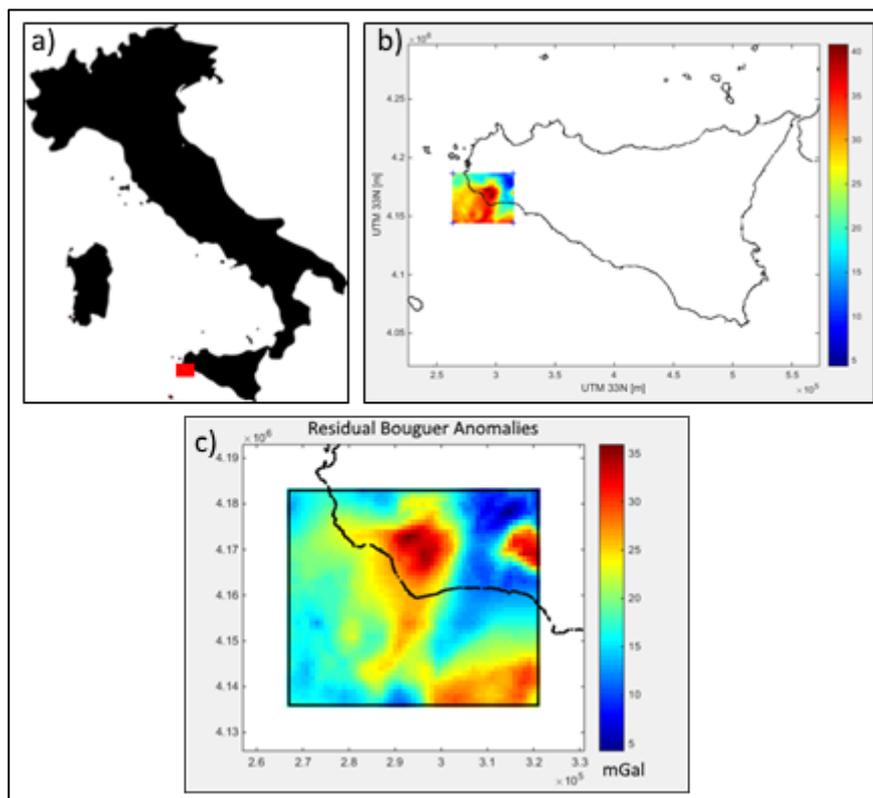


Figure 25 - a) The studied area in Western Sicily is the red box. b) Gravity data anomaly map (Carrozzo et al., 1991). Colorbar in mGal. c) Residual Bouguer anomalies, obtained removing a planar trend (SW-NE) to the data b).

### 3.5.1. Geological setting

The geological setting of the area is quite complex. The western Sicily fold and thrust belt links central and eastern Sicily to the Late Cenozoic Maghrebian submerged chain, developing along the Africa-Europe plate boundary. This chain involves a complex architecture of thrust systems resulting from a deformation history which has been active since Oligocene-Early Miocene times (e.g. [BARRECA AND MAESANO, 2012](#); [CATALANO ET AL., 2013](#); [MONACO ET AL., 1996](#); [MONTANARI ET AL., 2017](#)). Western Sicily is constituted at depth by Mesozoic carbonate platforms and their clastic covers, which are Cretaceous to Miocene in age, represented by several terrigenous formations. From a structural point of view, in the northern portion of the chain a tectonic stack of units up to 12 km thick is detected ([CATALANO ET AL., 2002](#)), while the southern sector is characterized by a less deformed southern sector, still affected by shortening ([BARRECA ET AL., 2014](#)). According to the outcropping geology ([CATALANO ET AL., 2002](#)), GPS data ([BARRECA ET AL., 2014](#); [MECCARIELLO ET AL., 2017](#)) the seismic studies ([CATALANO ET AL., 2000](#); [GHISSETTI ET AL., 2009](#); [MECCARIELLO ET AL., 2017](#)) and well data ([VIDEPI PROJECT](#)) drilled in the area for oil and gas exploration purpose (Figure 26a), in the north portion of the area the carbonate platform units and relative sedimentary cover are involved in a regional structural high, while in the southeast, these units are encountered only in deep wells drilled for oil exploration ([MONTANARI ET AL., 2017](#)).

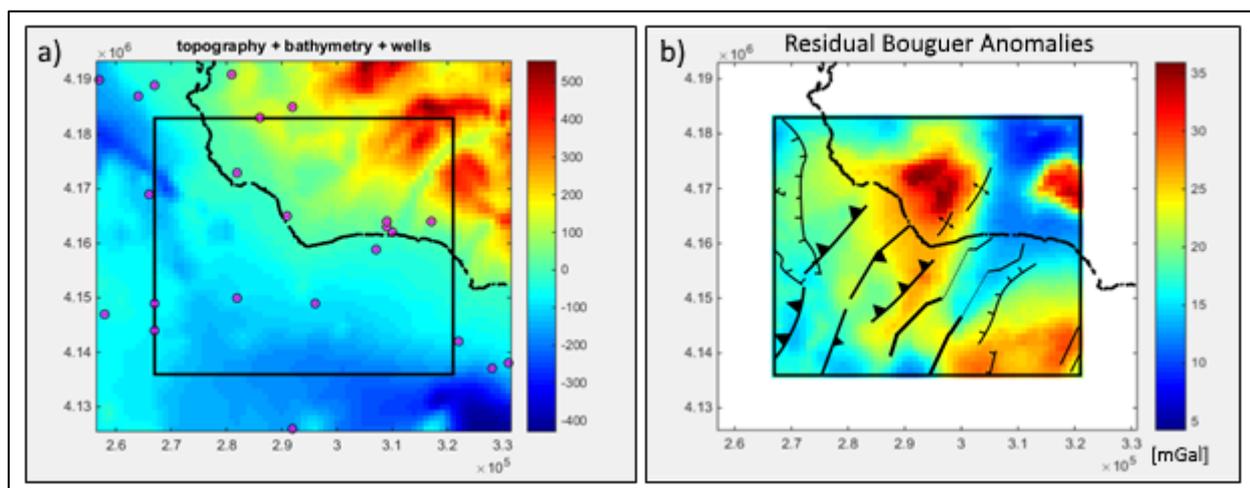
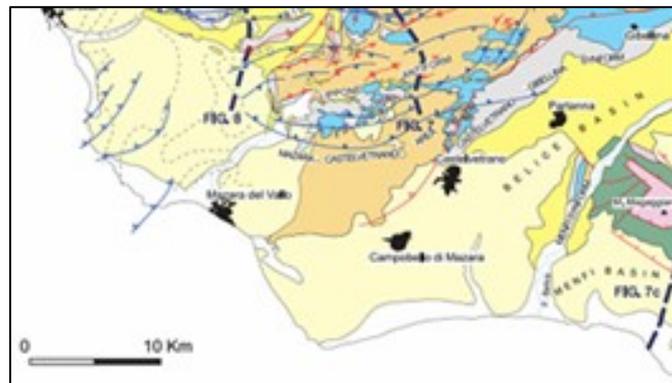


Figure 26 - a) Topography and bathymetry, in meters, and position of wells (magenta circles) from ViDEPI projects. b) Overlay of the residual Bouguer gravity anomaly map and the main structural features as given in Meccariello (2017).

In Figure 26b the main structural features (MECCARIELLO, 2017) of the area overlay the gravity map, demonstrating that the residual Bouguer anomalies largely depend on the geological setting relative to the first kilometers depth. The reverse faults (bold line with triangles) interpreted as reactivated during the Plio-Pleistocene from inherited extensional structures and probably related to the bending of the foreland plate which subsided into the foredeep ahead of the migrating orogen, are essentially aligned in the NW-SE direction (MECCARIELLO ET AL., 2017; MONTANARI ET AL., 2017). The gravity high located in the NE sector of the residual Bouguer anomalies, is well connected to the outcropping of the high-density bodies (Figure 27) related to the Mesozoic carbonate (CATALANO ET AL., 2002), in a mountainous area known as Monte Magaggiaro.



*Figure 27 - Geological sketch map (modified after Catalano et al., 2002)*

The maxima in the northern central section of the residual Bouguer anomalies is not related to any outcropping high-density bodies, but its geometry can be associated with the growth of distinct folds bounding fault (thrust) segments (BARRRECA ET AL. 2014), especially in the inland zone along the Castelvetro - Campobello di Mazara alignment. The high value of the residual Bouguer anomalies in the SE sector of the map cannot be associated to the outcropping geology, because located in the off-shore zone, but can be associated with a portion of Graham Bank, where the thrust is recent and is involving pre-existing extensional fault (MECCARIELLO ET AL., 2017).

### *3.5.2. Data inversion and results*

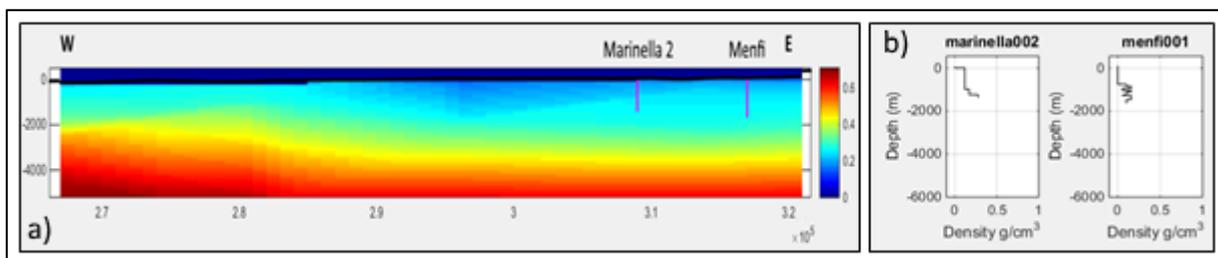
The subsurface has been discretized in the horizontal directions into 65 x 58 square cells. The cells have a size of 1 km in the horizontal direction and are 150 m thick. The model space extends up to

6000 m, with 40 cell layers of the same dimensions. This maximum depth value of the model was determined based on different factors, involving the geological information as well as computational problems:

- I. the deformation history of the thrust took place from Oligocene-Early Miocene and the top of the Mesozoic platform formations should not be deeper than 4 km (in the investigated area);
- II. the inversion process is related to the density contrast and below the chosen depth the structure are build up on similar density geological units (the study of the deeper crystalline basement units may be better conducted using a different kind of data, for instance aeromagnetics);
- III. to investigate a thicker volume, there would be two possible ways: a) to use of more cell layers, but this would reveal prohibitive from the point of view of the computational time; b) to use the same number of cell layers but using thicker cells, but this choice would deteriorate the model resolution.

To reduce distortions generated by the finite horizontal dimension of the model at the edges of the map, an external portion of the map, having horizontal dimension of 10 x 10 km, have been included outside each edge of the map.

The reference model (Figure 28a), adopted in the first stage of the Sharp-Edged Inversion strategy, consists in a smooth 3D linear interpolation of density in the model volume, recovered by the description of available wells (**VIDEPI PROJECT**, <http://unmig.sviluppoeconomico.gov.it/videpi>), and vary in a range of 0-0.65 g/cm<sup>3</sup> (densities are scaled by the minimum value).



*Figure 28 – a) XZ cross-section of reference model ( $m_{ref}$ ) obtained by the linear interpolation of the wells data. Colorbar is in g/cm<sup>3</sup>, axis in meters. Magenta lines are wells projection. b) Logs of converted densities of two wells present in the model cross-section a)*

As described in section 3.1, a small value of  $\alpha_m$  and big for  $\alpha_s$  guide the inversion solution toward a reference model; for that purpose, the first inversion is run according to  $\alpha_s = 1$  and  $\alpha_m = 0.0001$ . In this way the smooth model (Figure 28a) used as reference model leads a smooth final model fitting the observed data.

In order to ensure a reasonable number of geological units (clusters) distinguishable in terms of density contrast, the descriptions of oil exploration wells were taken into account. Additionally, the cluster centers (listed in Table 2) were associated to densities obtained from the Gardner's relationship (**GARDNER, 1974**) applied on seismic velocities assumed, in the same area, by different authors (**CATALANO ET AL., 2000; GHISSETTI ET AL., 2009; MECCARIELLO ET AL., 2017**). Thus, the following three classes of density contrasts ( $\text{g/cm}^3$ ): 0-0.17, 0.22-0.3,  $\geq 0.4$  (values scaled by the minimum) were selected, and used during the second phase of the inversion strategy. They can be geologically relatable to:

- from Quaternary to Late Pliocene formations (0-0.17  $\text{g/cm}^3$ ), terrigenous and clastic deposits, essentially clay and marl;
- Miocene formations (0.22-0.3  $\text{g/cm}^3$ ), clastic and evaporitic sequences that fill the Sicilian foredeep, basically clay, marl, calcarenite and gypsum;
- Mesozoic basement ( $\geq 0.4 \text{ g/cm}^3$ ), consisting in carbonate and siliceous-carbonate units, principally claystone, mudstone, cherty mudstone, limestone and dolostone.

Period/Epoch	Formation	Lithology	a)	b)	c)	d)
Middle - Upper Pleistocene		clay/silt/sand	2000-2600	2000	2000-2400	2,073
Lower Pleistocene - Upper Pliocene	Narbone	clay	2600-2800		2550	2,192
Lower Pliocene	Trubi	clay/marl	2800-3200		2750	2,244
Upper Miocene	Gessoso	gypsum/limestone/clay	3200-3400	3000	3300	2,349
Upper Miocene	Terravecchia	clay/sand	3400-3500		3000	2,294
Middle Miocene	Castellana	clay/silt			3000	2,294
Middle Miocene	S.Cipirello	marl/clay			3000	2,349
Lower Pliocene	Corleone	calcarenite/clay/marl marly limestone with chert		3500-4000	3300	2,367
Lower Miocene – Oligocene	Ragusa	chert			3400	2,384
Lower Cretaceous	Hybla	mudstone marl	3500-6000	4500	3600-4000	2,465
Lower Cretaceous - Upper Jurassic	Lattimusa	cherty mudstone/chalk clay mudstone- dolostone			4700	2,566
Middle - Upper Jurassic	Buccheri	dolostone			4700	2,566
Lower Jurassic	Inici	limestone-dolomite		5000	5400	2,657
Upper Triassic	Sciacca	dolomite			5800	2,705

Table 2 - Geological formation listed in well logs of the ViDEPI projects related to the studied area. Seismic velocity for the recognized by different authors: a) Ghisetti et al. 2009, b) Catalano et al. 2000, c) Meccariello 2017. Gardner's conversions in column d) are expressed in  $\text{g/cm}^3$

To decrease the computational cost, instead of creating a  $\mathbf{W}_m$  matrix to constrain the maximum gradient position in the 3D volume, the defuzzified model obtained from the second stage of the strategy is used as the reference model ( $\mathbf{m}_{ref}$ ), and appropriate values for  $\alpha_s$ , and  $\alpha_m$  were selected in equation 3.3. In this case, the inversion is aimed at constraining the final solution model to the

clustered one, easily interpretable from the geological point of view. Accordingly, the last step inversion has been run to  $\alpha_s = 10$  and  $\alpha_m = 0.0001$ .

In the resulting final output (figure 29) the density distribution consists of three homogeneous units (as the number of selected clusters) fitting the observed data (Figure 30).

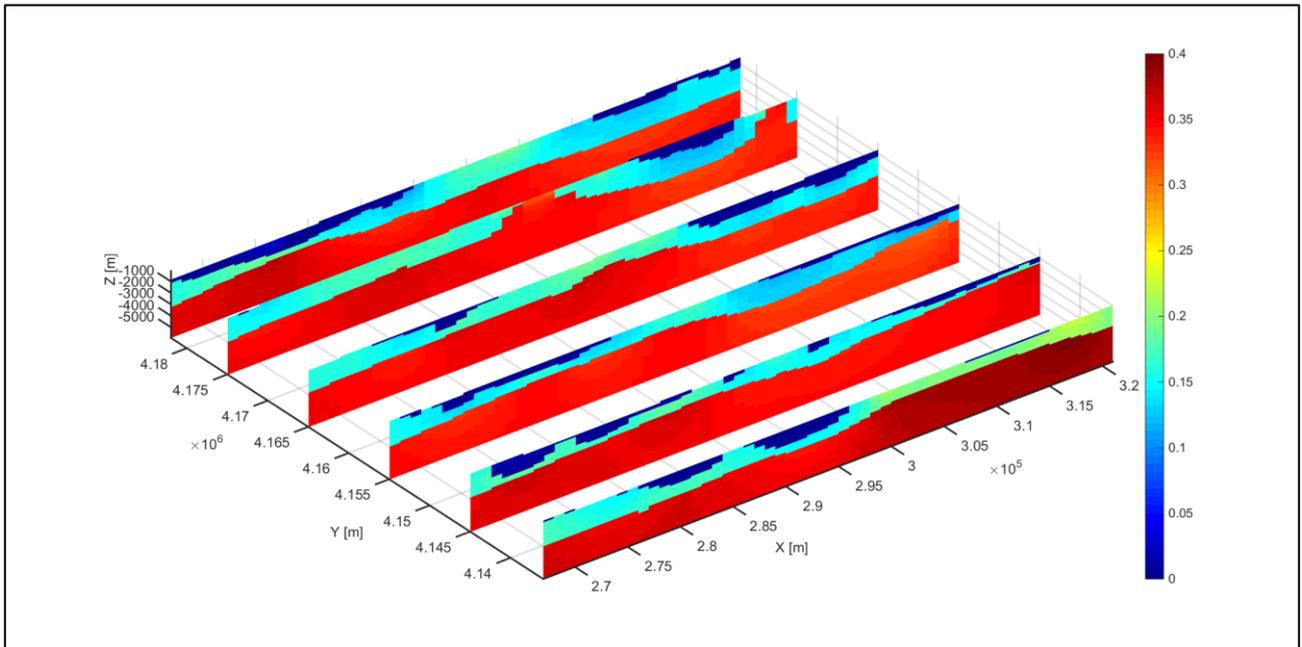


Figure 29 - East-West cross-sections of the density distribution model obtained after the last step of the Sharp-Edged Inversion strategy. Colorbar in  $\text{g/cm}^3$ .

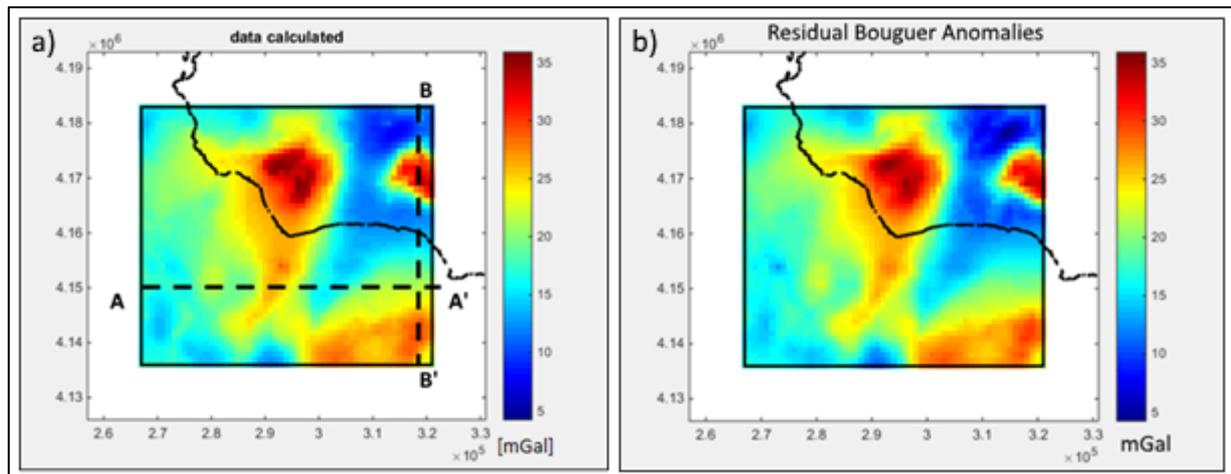
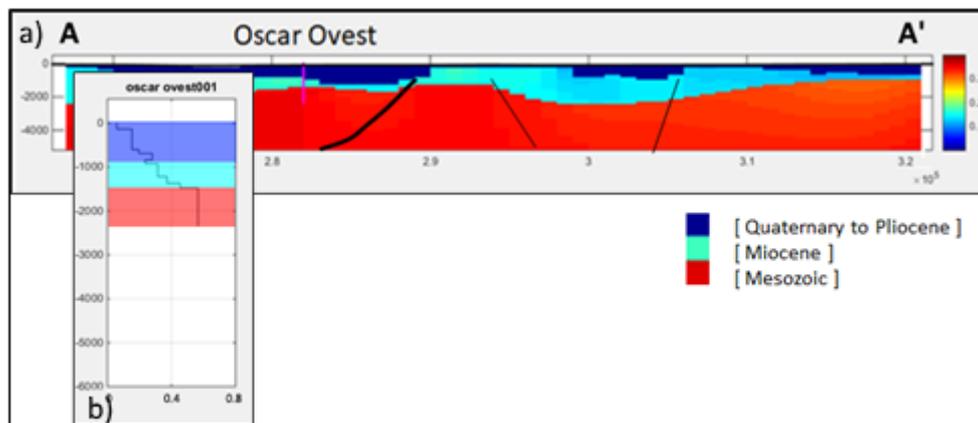


Figure 30 – Computed and input data after the Sharp-Edged Inversion strategy. a) Data computed by the density distribution model obtained from the third stage of the Sharp-Edged Inversion strategy (Figure 29). Axis in meters. Black dashed lines are the traces of the cross-section of the density distribution model shown in Figure 31 and 32. b) Residual Bouguer anomalies. Axis in metres.

The overall calculated anomaly map (Figure 30a) shows a good fit with the residual Bouguer anomalies, used as data input in the Sharp-Edged inversion strategy from the first step. The value of  $\mu$  used ( $1e-7$ ) ensure that the output do not present isolated high or low-density cells within the model, able to generate the high frequencies existing in the input map and not well reproduced.

In Figure 31a is shown the A-A' density model cross-section oriented East-West traced in Figure 30a, accompanied by the log of converted density (according to the Table 2) of "Oscar Ovest" well (Figure 31b) overlayers by the classification obtained during the second step of the Sharp-Edged inversion strategy. The density distribution clustered in the well is respected along its projection within the cross-section. A sort of structural interpretation is made by adding faults according to the structural map presented in Figure 26b (MECCARIELLO, 2017).



*Figure 31 - Cross-section of the volume solution along A-A' profile in Figure 30a, compared with Oscar Ovest well log. a) Density distribution cross section. Colorbar units are  $g/cm^3$ , axis in metres. Magenta line is the well projection. Black lines are faults and thrust drawn according to the structural map in Figure 26b. b) Log of converted densities of Oscar Ovest well, with highlighted the three density classes used in the clustering. The three different homogeneous bodies are related to: Quaternary to Late Pliocene formations (blue), Miocene formations (cyan) and Mesozoic basement (red).*

In Figure 32a is shown the B-B' density model cross-section oriented North-South traced in Figure 30a, and the black lines in the density model are a projection of the thrust and faults related to a geological cross-section (Figure 32b) from Montanari et al. (2017). In the two examples of figures 31 and 32, It is clear how the model is in agreement with the geological and structural information.

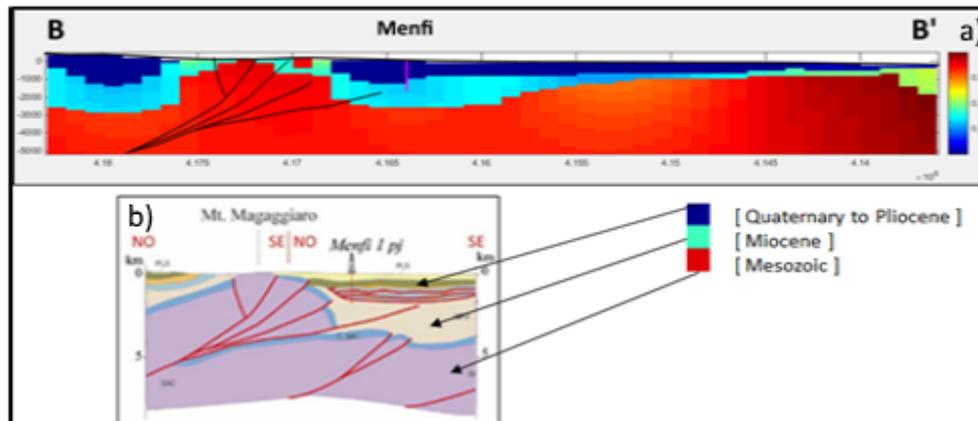


Figure 32 – Cross-section of the volume solution along B-B' profile in Figure 30a, compared with a geological cross-section. a) Density distribution cross section. Colorbar units are g/cm<sup>3</sup>, axis in metres. The three different homogeneous bodies are related to: Quaternary to Late Pliocene formations (blue), Miocene formations (cyan) and Mesozoic basement (red). Magenta line is the well projection. Black lines are faults and thrust drawn according to b) geological cross-section (modified after Montanari et al., 2017).

The density model (Figures 29, 31a and 32a) shows three different homogeneous bodies related to: Quaternary to Late Pliocene formations (blue), Miocene formations (cyan) and Mesozoic basement (red). Generally, within the model is possible to define the Mesozoic basement top and relate it to the geological structures. In the north-eastern part of the map the maximum gravity anomaly is relatable to a shallower Mesozoic basement inside the solution model. The surface geology (Figure 27) and its geological model across this maximum (Figure 30b) show an outcropping of the Mesozoic basement in the Mt. Magaggiaro area. The eastern side of the A-A' and the southern part of B-B' (Figures 31a and 32a) shows a shallow Mesozoic basement that can be related to a portion of the carbonate platform (northeaster part of the Graham Bank). This portion of the basement is interested by an extensional phase, as shown in the structural map (Figure 26b). The central part of the section A-A' shows another shallow portion of the Mesozoic basement producing the gravity high in the central area of the anomaly map. This basement high should represent the area deformed by the thrust system, visible in the structural map (Figure 26b).

The good fit of the observed and computed data and the general agreement between the Sharp-Edged density model and the available geological and structural information demonstrate the satisfactory performance of the Sharp-Edged Inversion strategy in producing a geophysical model of the variation of the density distribution depth ready for the geological interpretation.

### **3.6. CONCLUSIONS**

The target of the inversion of potential field is to retrieve a volume of physical property distribution representative of the subsoil geology and at the same time able to produce a field matching the observed data. In this thesis, is introduced an Inversion strategy able to calculate a solution model where the investigated physical property is classified in different homogeneous units. The goal of the proposed strategy is to obtain a model where the distribution at depth could be related in an unambiguous way to geological features. It is shown a three steps strategy involving two inversions of the same data input but constrained in a different way. In the first inversion a set of a priori constraints is used to soft constrain the inversion, meaning that the process to obtain the solution involves the minimization of a function promoting the spatial smoothness in the spatial domain. The obtained solution is clustered in a specific number of units, according to the fuzzy c-means clustering technique. The clustered solution can be viewed as a good model from the geological point of view, but loses its geophysical significance; in fact, the produced field mismatch the observed data. Thus, in the second inversion, the clustered model is used as hard constraint. The final output preserves the layered feature of the clustered version of the solution, and within the layer the physical property are only slightly varied until the model solution is able to produce a field that adequately fit the observed data. The application of the inversion strategy to synthetic as well as real data from two different geological complicated areas in Southern Italy, proved its effectiveness.

## CHAPTER 4 - A Stochastic Forward Problem Solver

*In this chapter is discussed a strategy to retrieve geophysical output, defined by homogeneous bodies, able to reproduce an observed potential field. The strategy is based on a stochastic method associated to a solver that works in Fourier domain. An application on a synthetic case is discussed.*

The interpretation of potential field has to face up with the uncertainty estimation due of its inherent nonuniqueness. There are infinite models whose theoretical response could fit the observed data, but the number of possible models is reduced if a priori guess about the source model is known (DIMRI, 1992). Chosen a set of different units representative of the subsurface, and associated the relative physical parameters, a layered model of the subsurface can be created and several forward methods can be used compute the data produced (such as TALWANI ET AL., 1959; TALWANI AND EWING, 1960; PARKER, 1972; BHATTACHARYA, 1978; DIMRI, 1998). Generally, this methods are based on a trial-and-error approach involving the computation of the effects from a simplified mathematical model (i.e. the forward model) of the presumed subsurface conditions with iterative re-computation of the effects based on alterations to the parameters of the forward model until an “acceptable” correlation with the residual anomaly is achieved (HINZE ET AL., 2013). The procedure is used for either profile or map analysis, where map calculations works on 3D sources which are more memory and time-consuming in order to achieve an acceptable match. The acceptance of the fit between the residual anomaly and calculated effect is very subjective and generally fixed a-priori according to the data errors, to the amount of geologic and geophysical information and also involving the interpreter experience.

In the previous chapters (2 and 3) two different strategies to obtain models where distinct geophysical units directly relate to the geology are discussed. In this chapter is discussed a strategy to obtain a volume of layered physical distribution able to generate a potential field fitting the observed one. In this method that the magnetization (or density) is assumed known, and several series of different solutions (surfaces shape and position) are generated in a stochastic way in an iterative process. For each iteration, a fast algorithm (PARKER, 1972) is used to compute the field produced by every solution. According to the best value of a coefficient evaluated on data misfit with respect to the observed data, a single solution is used to constrain the next iteration as a starting model.

The method shown in this chapter is related to solving an unknown surface between two bodies with different physical property. The process can be adopted in solving more than one unknown surface, and it is open to force the solutions to some constraints related to the position of certain portion of the volume. However, this study is still at an early stage and not yet fully developed.

#### 4.1. PARKER ALGORITHM

The Parker algorithm (PARKER, 1972) allows computing, in a fast way, the sum of the contribution of the gravity (or magnetic) effect due to the prisms in which a solution volume is discretised. The algorithm involves the Fourier transform for its advantage in speed computation. The algorithm proposed is able to solve problem for a non-uniform and uneven layer.

The two-dimensional Fourier transform of a function  $f(r)$  is defined by:

$$\mathcal{F}[f(\vec{r})] = \int_X dS f(\vec{r}) \exp(i\vec{k} \cdot \vec{r}) \quad (4.1)$$

where  $\mathbf{k}$  is the wave vector of the transformed function and  $X$  is taken to be the whole x-y plane.

Let's assume that the gravitational potential at a position  $r_0$  due to the layer is:

$$\begin{aligned} U(r_0) &= G\rho \int_V \frac{dV}{|r_0 - r|} \\ &= G\rho \int_D dS \int_0^{h(r)} dz / |r_0 - r| \end{aligned} \quad (4.2)$$

where  $G$  is Newton's gravitational constant and  $\rho$  is the density. If the observation point is confined to the plane  $z = z_0$ ,  $U$  is only depending on  $r_0$ . This acquisition plan must lie above all the topography, condition familiar in air and ship-borne acquisitions. Is it possible to relate the two above equations (see PARKER, 1972 for a full description) and obtain the following equation generalized for compute the vertical attraction due to a single layer with a not-flat lower boundary:

$$\mathcal{F}[\Delta g] = -2\pi G \exp(-|\vec{k}|z_0) \sum_{n=1}^{\infty} \frac{|\vec{k}|^{n-1}}{n!} \mathcal{F}[\rho(\vec{r})\{h^n(\vec{r}) - g^n(\vec{r})\}] \quad (4.3)$$

where  $\Delta g$  is the vertical attraction and  $h$  the thickness of the layer.

The equivalent magnetic problem can be solved according to a magnetized layer. It is assumed that the direction of magnetization is constant (but the intensity may vary), according to a simplification from the fact that perturbations to the observed field due to the magnetized material are always small and that the measurements are made of the total field  $|\mathbf{B}|$ . With these conditions the equivalent results to equation 4.3 is:

$$\begin{aligned} \mathcal{F}[\Delta|\mathbf{B}|] &= \frac{1}{2}\mu_0 \exp(-|\vec{k}|z_0) \widehat{B}_0 \cdot (i\vec{k}, |\vec{k}|) \widehat{M}_0 \cdot (i\vec{k}, |\vec{k}|) \\ &\quad \sum_{n=1}^{\infty} \frac{|\vec{k}|^{n-2}}{n!} \mathcal{F}[M(\vec{r})\{h^n(\vec{r}) - g^n(\vec{r})\}] \end{aligned} \quad (4.4)$$

where  $\Delta|\mathbf{B}|$  is the magnetic anomaly.

Using the appropriate Fourier transforms (equations 4.3 in gravity case, 4.4 in magnetic) it is possible to recover the field by using the inverse transform on the resultant function.

## 4.2. THE STOCHASTIC SOLVER

A Markov Chain is a stochastic process in which the conditional distribution at any future time  $t+1$  for a given past state and a present state depends only on the present state:

$$\begin{aligned} P_{ij}(t+1) &= P\{X(t+1) = j \mid X(t) = i, X(t-1) = i_{t-1}, \dots, X(0) = i_0\} \\ &= P\{X(t+1) = j \mid X(t) = i\} \end{aligned} \quad (4.5)$$

where  $i$  is the outcome of the  $k$ th trial of the process  $X(k)$  when  $k=t$  and  $P_{ij}(t+1)$  is the probability that the outcome of the stochastic process will be  $j$  at the trial  $k=t+1$  (SEN AND STOFFA, 2013). From

equation (4.5) is easily understandable that the probability of the outcome of a given trial depends only on the outcome of the previous trial. This probabilistic approach is used in optimization problems in several research fields.

In this thesis a Markov Chain simulation scheme is adopted to generate a stochastic process aimed at retrieving the correct depth to some surfaces able to generate a field that adequately fit the observed data. A starting surface (or a curve for 2D problem) is discretized in a number of points equal to the measurement points. Using a fast forward algorithm (PARKER, 1972) the field produced by the surface (or the curve) is computed and a goodness coefficient (GC) is evaluated by:

$$GC = \frac{\|X - Y\|_2^2}{\rho_{X,Y}} \quad (4.6)$$

where  $X$  is the potential field input (observed data),  $Y$  is the computed anomaly after the stochastic assignment. The denominator  $\rho_{X,Y}$  is the Pearson correlation coefficient (GIBBONS, 1985) that shows the linear relationship between two sets of data, and is described by:

$$\rho_{X,Y} = \frac{cov(X, Y)}{\sigma_X \sigma_Y} \quad (4.7)$$

where  $cov$  is the covariance,  $\sigma_X$  and  $\sigma_Y$  are the standard deviation of  $X$  and  $Y$  respectively. The Pearson correlation coefficient can have a value between -1 and 1. The closer the value of  $\rho_{X,Y}$  to zero, the greater the variation the data points around the line of best fit. Generally the sign of this parameter determine the correlation between the two  $X$  and  $Y$  dataset, when  $\rho_{X,Y} < 0$  the dataset are inverse correlate,  $\rho_{X,Y} > 0$  are directly correlated,  $\rho_{X,Y} = 0$  are not correlate. Moreover, according to the value of  $\rho_{X,Y}$  the degree of correlation can be: low (0 to 0.3), moderate (0.3 to 0.7) and high (0.7 to 1.0).

To choose the  $P_{ij}(t+1)$  in the stochastic process, a large series of models are produced. According to a random assignment of value to sum to the depth to each point of the surface, different models will produce different fields. According to a previous idea about the roughness of the real surface we want to retrieve, an average moving window can be used as filter to smooth the spikes

created summing the random numbers. It is selected the model (surface) producing a field with the smallest positive GC coefficient (equation 4.6).

The presented method has the advantage, with respect to the voxel-based inversion, that the boundaries between the different bodies present smooth variations and that the various layers are perfectly homogeneous.

By this way, the interpreted anomalous field is caused only by layers shape sources, and the contribution by deeper sources or side effect are not taken into account, neither in terms of trend.

In this thesis, the study of this approach is still in an embryonic stage and needs some improvement to solve some ambiguities. However, the good performance in a synthetic case can be shown.

### **4.3. A SYNTHETIC 3D CASE**

A magnetic synthetic model of two bodies separated by a rough surface is tested (Figure 33).

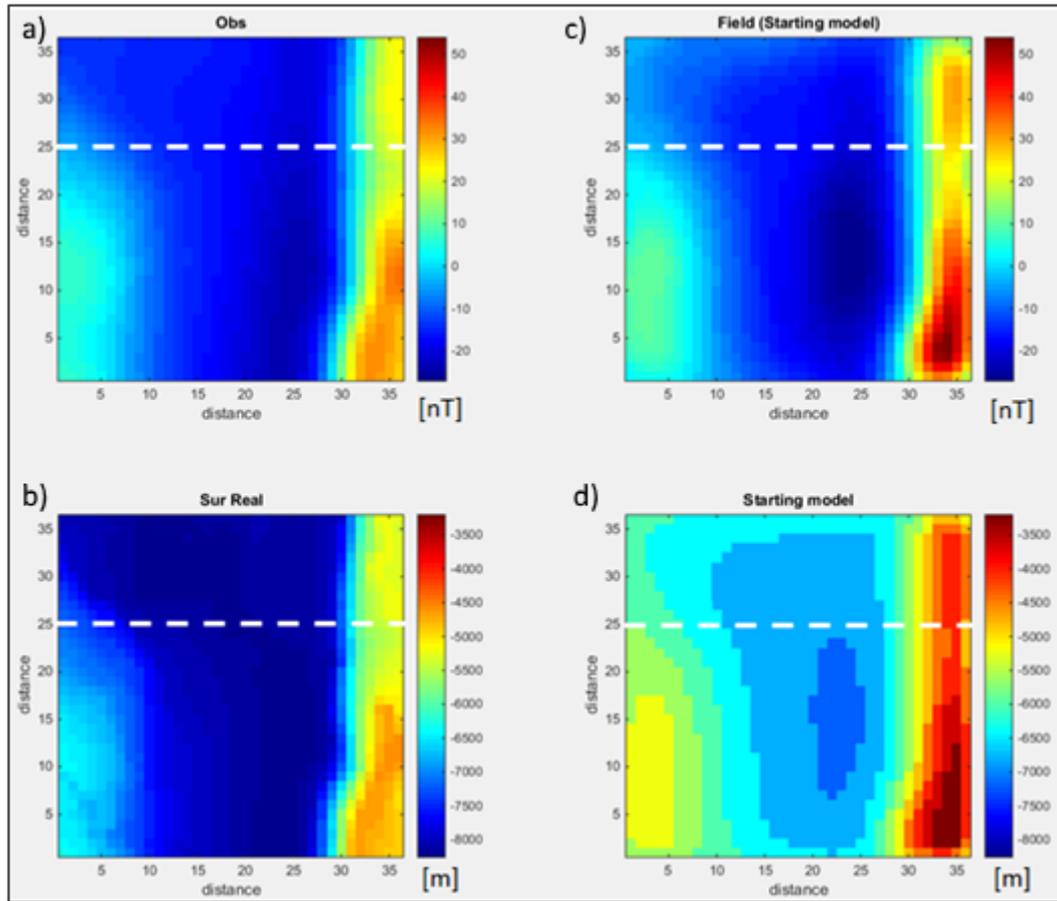


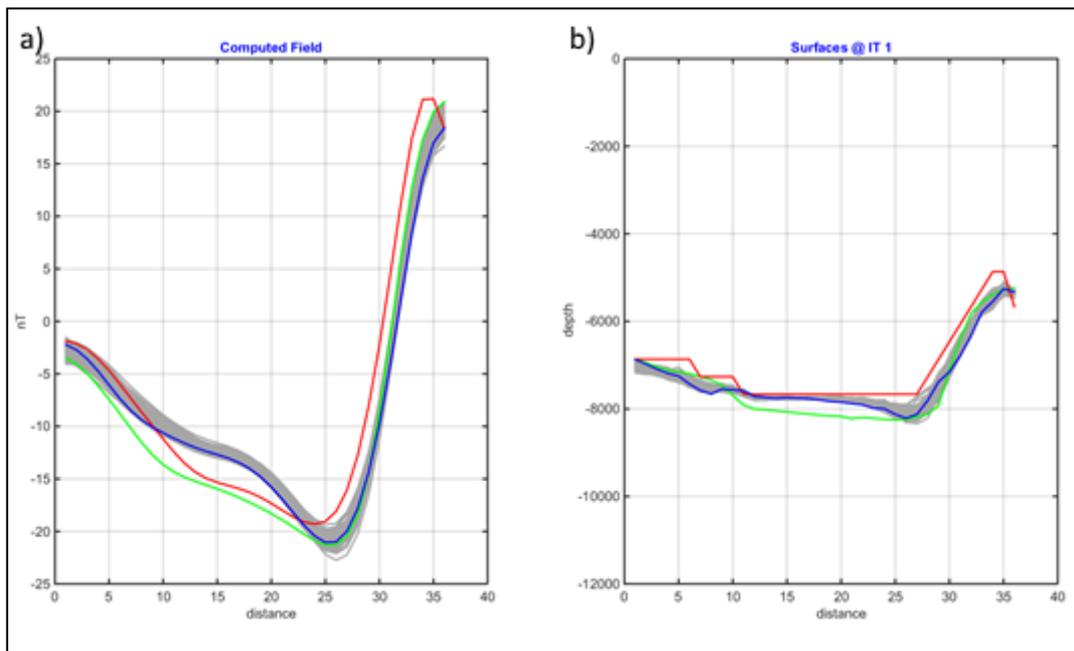
Figure 33 – Synthetic data and correct surface with the starting model and its computed data. a) Reduced to Pole magnetic anomalies map used as observed data during the stochastic process. b) correct surface. c) computed field by the starting model, at  $P_{ij}(1)$  shown in d). Axis in km. White dashed line is the trace of a cross-section.

The reduced to pole magnetic anomalies due to a rough surface (Figure 33b) separating a magnetized deeper body (1 A/m) from a cover of non-magnetic sediments in a volume extending from 0 to 12000 meters in depth (Figure 33a) was generated by the Parker algorithm (PARKER, 1972). The observation grid is formed by equally spaced points (1 km step) in the X-Y plane at 90 meters above the surface.

The  $P_{ij}(1)$  surface (Figure 33d), used to start the iterative stochastic process, comes from the interpretation of a solution of previous inversion of the observed data (Figure 33a) using a volume discretized by a relatively small number of cells. Thus the shape of the  $P_{ij}(1)$  surface is controlled by the cell size. The field computed by the  $P_{ij}(1)$  is shown in Figure 33c. Although the difference in shape and depth between the original and the  $P_{ij}(1)$  surface, the observed field and the one computed by  $P_{ij}(1)$  shows a good match. To observe the iterative and stochastic process a cross-

section in the model volume was selected (white dashed line in figure 33) for both field and surface solution (Figure 34).

The first iteration consists on the generation of 200 models obtained by the sum of random numbers to the starting model surface. Figure 34 shows the observed data produced by the correct surface position (green line in Figure 34a and b, respectively), the  $P_{ij}$  (1) input surface (red line in Figure 34b) and its produced field (red line in Figure 34a). The grey lines represent the 200 different solutions (in terms of surface in Figure 34b and the data computed in Figure 34b) obtained during the first iteration. The obtained surfaces were filtered by an average moving window (size 5x5) before computing the field. The blue line in Figure 34b represent the best solution in terms of the GC value and is used as the  $P_{ij}$  (2) for the second iteration of the Markov Chain simulation.



*Figure 34 – Solutions relative to the first iteration along the profile in figure 33. a) computed fields. b) solution surfaces with a magnetization contrast of 1 A/m. Green line in b) represents the correct surface. Red lines in b) are the  $P_{ij}$  (1) and its computed field (a). Blue lines in b) are the best GC surface value and its calculated data (a). Grey lines are 200 different solutions. X-Axis are n metres. Cross-section at Y-constant of 25 km, white trace in figure 33.*

The stochastic process continues up to a prefixed number of iterations or when the GC value reach a chosen target. In Figure 35b blue line is the  $P_{ij}$  (16), considered as a good solution in terms of the data misfit between the observed input data (green line in Figure 35b) and the one computed by  $P_{ij}$  (blue line in Figure 35b) after 15 iterations.

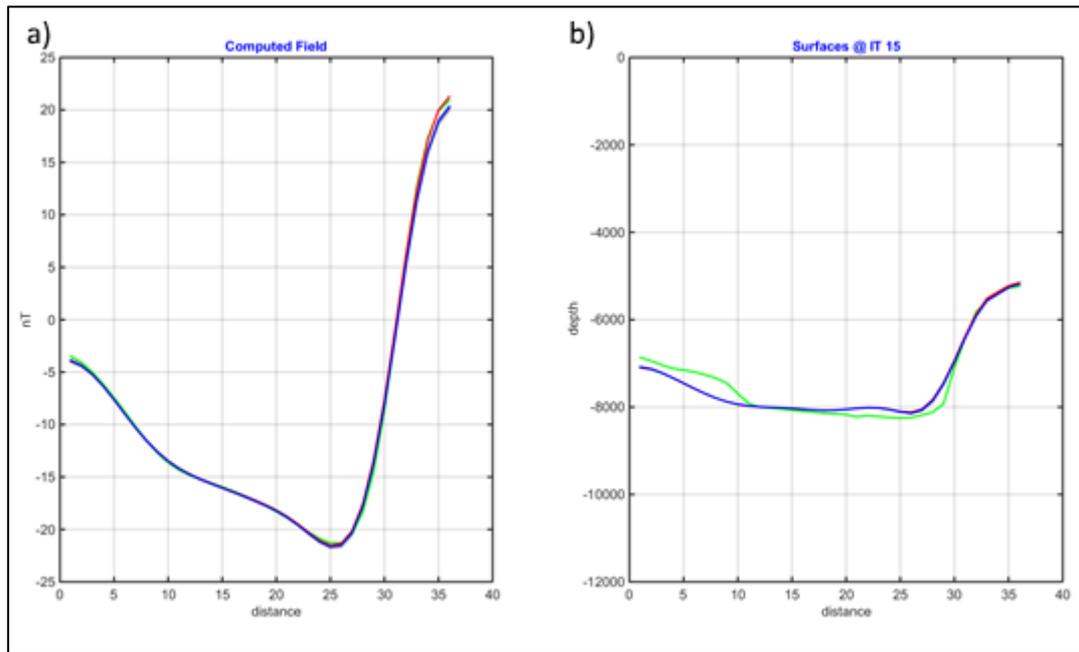


Figure 35 - Solutions relative to the iteration number 15 along the profile in figure 33. a) Comparison of the observed field (green line) and the field generated by the best computed surface ( $P_{ij}(16)$ , blue line in b) with a magnetization contrast of 1 A/m. In b) the green line represents the correct surface. X-Axis are in metres. Cross-section at Y-constant of 25 km, white trace in figure 33.

In figure 36 is shown the result of the stochastic forward solver strategy after 15 iterations in the 2D (X-Y plane), for both best surface and the relative computed field, and its shown the arithmetic difference with the reduced to pole magnetic anomaly map used as input and the generated surface. The best solution obtained  $P_{ij}(16)$  (Figure 36d) is characterized by a smooth and regular surface, especially if compared with the starting surface (Figure 31d). In Figure 36f is possible to see that the main differences between the surface  $P_{ij}(16)$  and the real surface (Figure 36b) are located in the area where the real one present the highest horizontal gradient. These differences are also dependent on the choice of the size of the averaging window (5x5 in this case). The terraced aspect of the first starting model (Figure 33d) is totally removed in the computed surface (Figure 36f). The computed surface separates now two bodies with two different homogeneous magnetization (0 and 1 A/m), the generated field fits the observed data and the model has a clear geological significance.

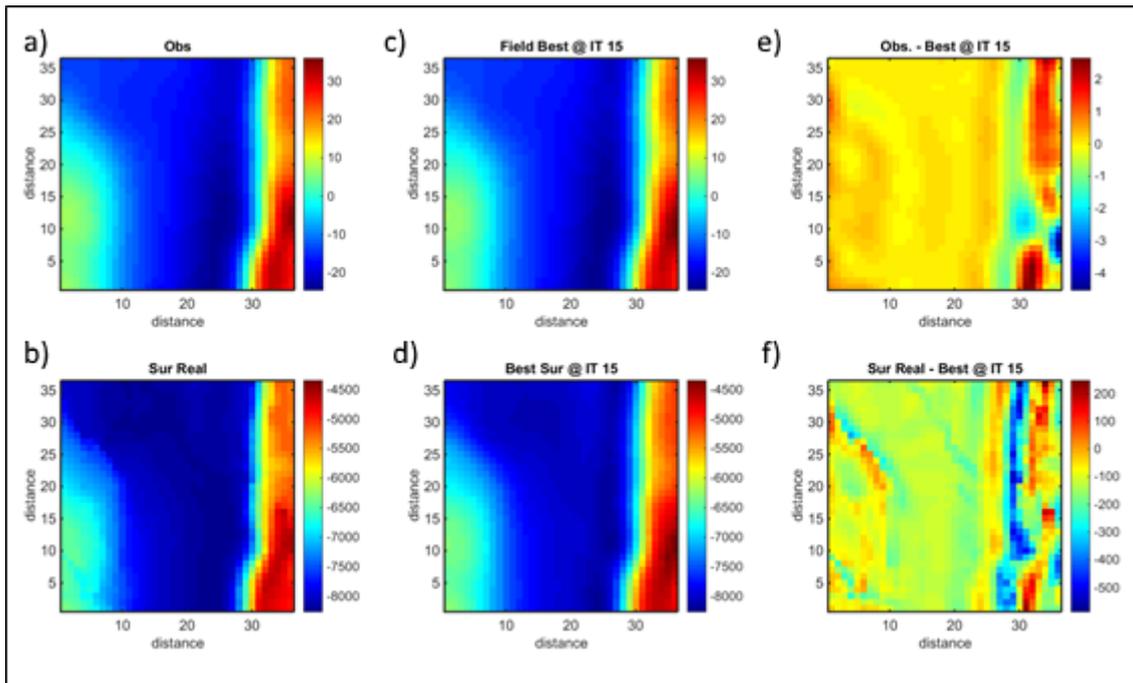


Figure 36 - Solutions relative to the iteration number 15, in map. a) Reduced to Pole magnetic anomalies map used as observed data during the stochastic process. b) the correct surface. c) field computed at iteration number 15. d) best surface solution at iteration 15. e) and f) are the difference between the data and the surfaces, real and computed, respectively. Colorbars in a), c) and e) are in nT. Colorbars in b), d) and f) are in metres. Axis are in km.

This methodology is open for solving the depth for more than one layer. It is possible to integrate constraints about the depth position of the layers, for instance derived from well or other external information.

#### 4.4. CONCLUSIONS

Forward modelling is one separate part of potential fields evaluation and can be used in various positions during the interpretation. According to a function that compute the field produced by different source bodies, one of possible forward problem is to retrieve the geometry and position of sources with homogeneous physical property distribution. The shape and location of the sources can vary until an acceptable correlation with the observed anomaly is achieved. It is introduced a stochastic forward problem solver based on a Markov chain iterative simulation, where the field is computed using a fast computational algorithm that works in Fourier domain. Once it is set the value of the physical property of two different bodies, several random surfaces are generated during each Markov chain iteration. The goodness coefficient (GC) for the field

generated by each surface is evaluated and the one with the best GC value is selected as starting model for the next Markov chain simulation. Once prefixed iterations are conducted or a selected value of GC is reached, the iterative process ends. The use of the Parker algorithm guarantees the computation of a several numbers of models in a fast way. The final model obtained shows a possible shape of different bodies with homogeneous distribution able to produce a field that adequately match an observed anomaly field. The application on a synthetic 3D model proved its effectiveness for solving the shape and position of a single unknown surface.

## CONCLUSIONS AND FUTURE PERSPECTIVES

In this thesis, three novel strategies are proposed to obtain geophysical output where the physical property distribution could be related in an unambiguous way to geological features. The three strategies are applied on classical methods used for potential field interpretation. According to the order in which the strategies are presented in the thesis: a) the first one is related to the interpretation of the potential field maps; b) the second one is connected to the potential field data inversion; c) the third one is focused on the forward problem solution.

Under the point of view of this thesis, the feature we want to obtain from the models, easily related to geology, consist in different homogeneous units with distinct boundaries separating each other, such as the interpreting geological models show.

About the forward problem solver, an iterative stochastic procedure is suggested. The solution is considered as surfaces that bound different layered bodies with a chosen physical property. Random surfaces are created by adding random numbers to a starting surface and it is calculated the field by the use of an algorithm that works in a Fourier domain. According to the Markov chain simulation, the solution obtained in a certain time of the iterative process depends only to the strictly previous one and not to the entire process. At each iteration of the Markov chain, several surfaces are created and the one used as starting model in the next iteration is selected according to the value of a goodness coefficient. The algorithm used to calculate the field is fast and guarantees the production of large number of surface maintaining a low computational cost. The final model obtained shows a possible shape of different bodies with homogeneous distribution able to produce a field that adequately match an observed anomaly field. By this strategy, it is possible to smooth cell-shaped surface obtained from other geophysical interpretations.

As regard the mapping method, a new strategy (Simple Guided Clustering, SGC) is proposed. A cluster based strategy is used for the first time in potential field map transformation. It is shown how terraced maps can be obtained using a simple reclassification of the field values. The selection of the centers can be conducted in an automatic way, exploring the kernel smoothing function of the distribution of the data itself. The terraced map obtained by this method is a representation of distinct units and the passage between terraces occurs in an abrupt way. The terraced maps obtained using the presented strategy shows units in which the field value is

constant, and do not present “blocky” features as produced by other algorithms. With these features the terraced map it is a good qualitative interpretation tool. Starting from a terraced map, it is produced an apparent physical property horizontal distribution map and its abrupt variations resembling what is generally displayed in a geological map.

Concerning the inversion of potential field data, a strategy (Sharp-Edged Inversion) is presented. The strategy allows to obtain volume where the retrieved physical property do not vary smoothly. Following the strategy, the model is classified in units and the distribution at depth could be related in an unambiguous way to geological features (i.e. faults or stratified layers). The number of the units presented in the model it has to be fixed a priori, and the associated physical values for the units can be choice according to external knowledges or selected automatically through the minimization of the fuzzy c-means function. Three steps compose the strategy; the first and the last are inversion with different constraints and associated weights, the second one is conducted by clustering the output of the first smooth inversion. The final output integrate the geological interpretation step into a model able to produce a field that match the observed geophysical anomaly. The final obtained model is a valid model in both geophysical and geological point of view.

Future research can be focused in the last strategy presented in this thesis. The forward problem solver can be extended to solve more than one unknown surfaces or to correlate some criteria in the choosing of the random numbers range. The range can be correlated to some characteristic of the observed input data or on some feature of the goodness coefficient. Another possible upgrade of this strategy is to adopt a way to constrain the surfaces or portions according to available information.

The inversion strategy can be applied in a different “environment”. The one proposed in this work need a volume discretized in a set of prismatic cells that gives to the final solution an aspect too strictly connected to the cell shape. A possible perspective is to use a different volume discretization, such as a model built up with tetrahedrons with different dimension according to different resolution in the three dimensions.

## **ACKNOWLEDGEMENTS**

I would like to express my sincere gratitude to my advisor Prof. Giovanni Florio for the continuous support of my Ph.D study and related research, for his patience and motivation.

I thank the Geoscience Data Repository of Natural Resources, Canada, for making available the gravity data relative to the Noranda-Val-d'Or (<http://gdr.agg.nrcan.gc.ca/gdrdap/dap/search-eng.php>).

I want also sincerely thank the reviewers, Dr. Roman Pasteka and Prof. Agata Siniscalchi, for the insightful comments and criticism that helped me to improve the final versions of this work.

## REFERENCES

- Ascione A., Cinque A., Improta L., Villani F., 2003. Late Quaternary faulting within the Southern Apennines seismic belt: new data from Mt. Marzano area (Southern Italy). *Quaternary International* 101–102, 2003, 27–41.
- Barreca G., Maesano F.E., 2012. Restraining stepover deformation superimposed on a previous fold-and-thrust-belt: a case study from the Mt. Kumeta–Rocca Busambra ridges (western Sicily, Italy). *J. Geod.* 55, 1–17.
- Barreca, G., Bruno, V., Cocorullo, C., Cultrera, F., Ferranti, L., Guglielmino, F., Guzzetta, L., Mattia, M., Monaco, C., Pepe, F., 2014. Geodetic and geological evidence of active tectonics in southwestern Sicily (Italy). *J. Geodyn.* 82, 138–149.
- Bhattacharyya, B.K. 1978. Computer modeling in gravity and magnetic interpretation. *Geophysics*, 43, 912–929.
- Berrino G. and Camacho A., 2008, 3D Gravity Inversion by Growing Bodies and Shaping Layers at Mt. Vesuvius (Southern Italy). *Pure appl. geophys.* 165, 1095–1115
- Bezdek J. C., 1981, *Pattern recognition with fuzzy objective function algorithms*: Plenum Press.
- Blakely R.J. 1996. *Potential Theory in Gravity & Magnetic Applications*. Cambridge University Press.
- Bowman, A. W., and A. Azzalini, 1997, *Applied smoothing techniques for data analysis*: Oxford Science Publications.
- Brancaccio L. 1968. Genesi e caratteri delle forme costiere nella Penisola Sorrentina. *Bollettino della Società dei Naturalisti in Napoli* 77, 247–274.
- Camacho A, Montesinos F. G. and Vieira R., 2000. Gravity inversion by means of growing bodies. *Geophysics*, 65 (1) 95–101.
- Carrozzo M.T, Luzio D., Margiotta C., Quarta T., 1991. *Carta Gravimetrica d'Italia, scala 1:500.000* C.N.R. - P.F.G. 1991.
- Catalano R., Franchino A., Merlini S., Sulli A., 2000. Central western Sicily structural setting interpreted from seismic reflection profiles. *Mem. Soc. Geol.It.* 55, pp. 5–16.
- Catalano, R., Merlini, S., Sulli, A., 2002. The structure of western Sicily, central Mediterranean. *Pet. Geosci.* 8, 7–18.
- Catalano, R., Valenti, V., Albanese, C., Accaino, F., Sulli, A., Tinivella, U., Gasparo Morticelli, M., Zanolta, C., Giustiniani, M., 2013. Sicily's fold–thrust belt and slab rollback: the SI.RI.PRO. Seismic crustal transect. *J. Geol. Soc.* 170, 451–464.
- Cella, F., and Fedi, M., 2012, Inversion of potential field data using the structural index as weighting function rate decay: *Geophysical Prospecting*, 60, 313–336.

- Cinque A., Lambiase S., Sgrosso I., 1981. Su due faglie dell'alta valle del Sele legate al terremoto del 23 novembre 1980. *Rendiconti della Società Geologica Italiana* 4, 127–129.
- Cooper G. R. J. and D. R. Cowan, 2009. Terracing potential field data: *Geophysical Prospecting*, 57, 1067–1071, doi: 0.1111/j.1365-2478.2009.00791.x.
- Cordell L. and A. E. McCafferty, 1989. A terracing operator for physical property mapping with potential field data: *Geophysics*, 54, 621–634, doi:10.1190/1.1442689.
- D'Addezio G., Pantosti D., Valensise G., 1991. Paleoearthquakes along the Irpinia fault at pantano di S. Gregorio Magno, Southern Italy. *Il Quaternario* 4 (1a), 121–136.
- Davis J.C., 2002. *Statistics and Data Analysis in Geology* (3rd ed.). 6: 487-492.
- Dimri, V., 1992, *Deconvolution and Inverse Theory- Application to geophysical problems*: Elsevier.
- Dimri, V.P. 1998. Fractal behavior and detectability limits of geophysical surveys. *Geophysics*, 63, 1943–1946.
- DISS Working Group 2010. Database of Individual Seismogenic Sources (DISS), Version 3.1.1: A compilation of potential sources for earthquakes larger than M 5.5 in Italy and surrounding areas. <http://diss.rm.ingv.it/diss/>, © INGV 2010 - Istituto Nazionale di Geofisica e Vulcanologia.
- Duda, R., and Hart, P., 1973, *Pattern classification and scene analysis*: Wiley-Interscience, New York, 482 p.
- Farquharson, C. G., M. R. Ash, and H. G. Miller, 2008, Geologically constrained gravity inversion for the Voisey's Bay ovoid deposit: *The Leading Edge*, 27,64–69, doi: 10.1190/1.2831681.
- Fedi M., Florio G., 2006. SCALFUN: 3D analysis of potential field scaling function to determine independently or simultaneously Structural Index and depth to source. 76th SEG Annual Meeting, New Orleans, 963-967.
- Florio G., Fedi M., Rapolla A., 2009. Interpretation of regional aeromagnetic data by the scaling function method: the case of Southern Apennines (Italy). *Geophysical Prospecting* 57, 479–489.
- Galli P.A.C., Giocoli A., Peronace E., Piscitelli S., Quadrio B., Bellanova J., 2014. Integrated near surface geophysics across the active Mount Marzano Fault System (southern Italy): seismogenic hints.
- Gardner, G. H. F., Gardnerl. W., and Gregorya. R., 1974. Formation velocity and density - the diagnostic basics for stratigraphic traps. *Geophysics* 39, 770–780.
- Ghisetti C., Gorman A.R., Grasso M., Vezzani L., 2009. Imprint of foreland structure on the deformation of a thrust sheet: the Plio-Pleistocene Gela Nappe (southern Sicily, Italy). *Tectonics*, 28, TC4015.
- Gibbons, J.D., 1985. *Nonparametric Statistical Inference*, 2nd ed., M. Dekker.
- Gordon, A., 1999. *Classification* (2nd ed.), London: Chapman and Hall/CRC Press.
- Hartigan, J., 1975, *Clustering algorithms*: John Wiley and Sons, New York, 351 p.

- Hathaway, R. J., and J. C. Bezdek, 2001, Fuzzy c-means clustering of incomplete data: IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics, 31, 735–744.
- Hinze, W. J., R. R. B. von Frese, and A. H. Saad, 2013, Gravity and magnetic exploration: Cambridge University Press.
- Ialongo, S. 2013. Depth resolution in potential field inversion: theory and applications. Ph.D. thesis, University of Naples 'Federico II'.
- Improta L., Zollo A., Bruno P.P., Herrero A., Villani F., 2003. High-resolution seismic tomography across the 1980 (Ms 6.9) Southern Italy earthquake fault scarp. Geophysical Research Letters 30 (10), 1494.
- Kearey P., Brooks M. and Hill I., 2002, An Introduction to Geophysical Exploration. Blackwell Publishing, Oxford.
- Keating, P., 1992, Density mapping from gravity data using the Walsh transform: GEOPHYSICS, 57, 637–642.
- Klemela J., 2009. Smoothing of Multivariate Data. Density Estimation and Visualization. Wiley and Sons, New Jersey. 8: 195-211.
- Kyprianidis, J. E., H. Kang, and J. Döllner, 2009, Image and video abstraction by anisotropic Kuwahara filtering: Computer Graphics Forum, 28, 1955–1963.
- Lane, R., D. FitzGerald, A. Guillen, R. Seikel, and P. McInerey, 2007, Lithologically constrained inversion of magnetic and gravity data sets: Preview, 129,11–17
- Lelièvre, P. G., 2009, Integrating geologic and geophysical data through advanced constrained inversions: Ph.D. thesis, University of British Columbia.
- Li, X., 2016, Terracing gravity and magnetic data using edge-preserving smoothing filters: Geophysics, 81, no. 2, G41–G47.
- Li, Y., and D. W. Oldenburg, 1996, 3-D inversion of magnetic data: Geophysics, 61, 394–408.
- Li Y., Oldenburg D.W., 1998. 3 D inversion of gravity data. Geophysics 63, 1, 109-119.
- Li Y., and Oldenburg D., 2003, Fast inversion of large-scale magnetic data using wavelet transform and logarithmic barrier method: Geophysical Journal International, 152, 251-265.
- Lo Re D., Florio G., Ferranti L., Ialongo S., Castiello G., 2016. Self-constrained inversion of microgravity data along a segment of the Irpinia fault. Journal of Applied Geophysics, 124, 148–154.
- MacQueen, J. B., 1967, Some Methods for classification and Analysis of Multivariate Observations. Proceedings of 5-th Berkeley Symposium on Mathematical Statistics and Probability, Berkeley, University of California Press, 1, 281-297.
- Meccariello, M., Ferranti, L., Barreca, G., Palano M., 2017. New insights on the tectonics of the Lampedusa Plateau from the integration of offshore, on-land and space geodetic data. Ital. J. Geosci., Vol. 136, No. 2, pp. 206-219.

- Meccariello, M., 2017. Tectonic evolution and current deformation of the NW Sicily Channel and the Lampedusa Plateau based on multi-resolution seismic profiles analysis. PhD Thesis, University of Naples Federico II.
- Menke, W., 1989. Geophysical Data Analysis: Discrete Inverse Theory. 1st Edition, Academic Press, San Diego, 289.
- Monaco C., S. Mazzoli and L. Tortorici 1996 Active thrust tectonics in western Sicily (southern Italy): the 1968 Belice earthquake sequence. *Terra Nova*, 8, 372-381.
- Montanari, D., Minissale, A., Doveri, M., Gola, G., Trumpy, E., Santilano, A., Manzella, A., 2017. Geothermal resources within carbonate reservoirs in western Sicily (Italy): A review. *Earth-Science Reviews* 169 (2017) 180–201.
- Oldenburg, D. W., and Y. Li, 1994, Subspace linear inverse method, *Inverse Problems*, 10, 915 - 935.
- Oldenburg, D. W., and Y. Li, 2005, Inversion for applied geophysics: A tutorial, in D. K. Butler, ed., *Near-surface geophysics: SEG*, 89-150.
- Pantosti D., Valensise G., 1990. Faulting Mechanism and Complexity of the November 23, 1980, Campania-Lucania Earthquake, Inferred From Surface Observations. *Journal of Geophysical Research* 95, 15319-15341. doi: 10.1029/JB095iB10p15319
- Pantosti D., Schwartz D.P., Valensise G., 1993. Paleoseismology along the 1980 surface rupture of the Irpinia fault: implications for earthquake recurrence in the Southern Apennines, Italy. *Journal of Geophysical Research* 98 (B-4), 6561–6577.
- Parker, R. L., 1972, The rapid calculation of potential anomalies: *Geophysical Journal of the Royal astronomical Society*, 31, 447–455.
- Pham, D. L., 2001, Spatial models for fuzzy clustering: *Computer Vision and Image Understanding*, 84, 285–297.
- Phillips, J. D., and Simpson, R. W., 2015, Sharpening the boundaries - 3D terracing applied to physical property inversions: 85th Annual International Meeting, SEG, Expanded Abstracts, 1536–1540.
- Portniaguine and Zhdanov, 1999. Focusing geophysical inversion images. *Geophysics*, 64, 874-887.
- Reynolds J. M., 1997, *An introduction to applied and environmental geophysics*, John Wiley & Sons Ltd, Chichester, West Sussex.
- Sen, M.K., Stoffa, P.L., 2013. *Global Optimization Methods in Geophysical Inversion*. Cambridge University Press.
- Simpson, R. W., R. C. Jachens, V. E. Langenheim, and T. G. Hildenbrand, 2008, Drainage-divide approach to finding boundaries of geologic bodies using gradients of 2D potential field anomalies and 3D tomographic velocity anomalies: Fall Meeting, American Geophysical Union, Extended Abstracts, GP43B-0800

- Sheath, P., and Sokal, R., 1973, Numerical taxonomy: Freeman, San Francisco, 573 p.
- Sun, J., and Y. Li, 2010, Inversion of surface and borehole gravity with thresholding and density constraints: 80th Annual International Meeting, SEG, Expanded Abstracts, 1798–1803.
- Sun J., Li Y., 2015. Multidomain petrophysically constrained inversion and geology differentiation using guided fuzzy c-means clustering. *Geophysics*, 80 (4), 1-18.
- Talwani, M., and Ewing, M. 1960. Rapid computation of gravitational attraction of three-dimensional bodies of arbitrary shape. *Geophysics*, 25, 203–225.
- Talwani, M., J. L. Worzel, and M. Landisman, 1959, Rapid gravity computations for two-dimensional bodies with application to the Mendocino submarine fracture zone: *Journal of Geophysical Research*, 64(1), 49–59, doi: 10.1029/JZ064i001p00049.
- Tikhonov A. V., Arsenin V. Y., 1977. *Solution of ill-posed problems*: John Wiley & Sons.
- Tou, J., and Gonzalez, R., 1974, *Pattern recognition principles*: Addison-Wesley, Reading, Mass., 377 p.
- Tryon, R.C., 1939. *Cluster Analysis: Correlation Profile and Orthometric (Factor) Analysis for the Isolation of Unities in Mind and Personality*. Edwards Brothers, Ann Arbor.
- Urquhart, W.E.S. and D.W. Strangway, 1985, Interpretation of part of an aeromagnetic survey in the Matagami area of Quebec: in W.J. Hinze, ed., *The Utility of Regional Gravity and Magnetic Anomaly Maps*: SEG, 426–438.
- Ward, W. O. C., P. B. Wilkinson, J. E. Chambers, L. S. Oxby, and L. Bai, 2014, Distribution-based fuzzy clustering of electrical resistivity tomography images for interface detection: *Geophysical Journal International*, 197, 310–321.
- Williams, S. E., J. D. Fairhead, and G. Flanagan, 2005, Comparison of grid Euler deconvolution with and without 2D constraints using a realistic 3D magnetic basement model: *Geophysics*, 70, no. 3, L13–L21.

## WEBSITE CITATIONS

VIDEPI PROJECT, <http://unmig.sviluppoeconomico.gov.it/videpi>