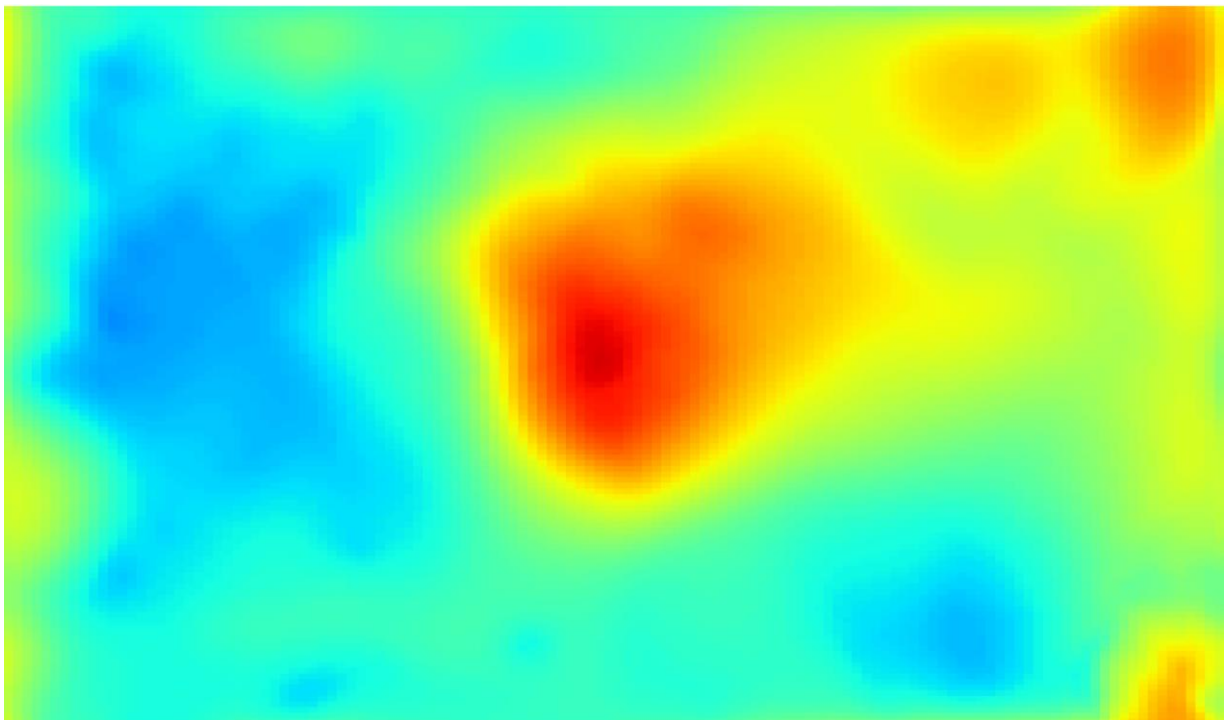


Forward and Inverse modelling of gravity anomaly for geothermal applications:

Case study of Los Humeros



Master Thesis

Katarzyna M. Sejan

Utrecht University



Universiteit Utrecht

TNO innovation
for life



Universiteit Utrecht

TNO innovation
for life

Forward and Inverse modelling of gravity anomaly for geothermal applications:

Case study of Los Humeros

Master Thesis

Final Version

Date of Submission:	13 th August 2018
Author:	Katarzyna Mira Sejan
Student number:	5791332
Credits:	37.5 ECTS
Master Programme:	Earth Structure and Dynamics
Track:	Physics of the Deep Earth and Planets
Supervisors:	Mark Vrijlandt (TNO) Dr. Fred Beekman (UU)
Host Organisation:	Applied Geosciences Group Netherlands Organisation for Applied Scientific Research (TNO)
Institute:	Department of Earth Sciences Faculty of Geosciences Utrecht University

ABSTRACT

Forward modelling of gravity anomaly may be used to validate structural models of the subsurface, particularly when the deeper subsurface is poorly constrained. When the modelled gravity anomaly correlates well to the measured anomalies, the structural model is validated. Otherwise, an inverse modelling with data assimilation will allow to reduce the uncertainties of the density anomaly present in the structural model.

Basin3D software developed by TNO, is a gravity anomaly modelling software for both forward and inverse modelling. The methodology of the software is based on previous work in the field by Caratori Tontini et al. (2009) and Cooley and Tukey (1965), and is implemented in Java. This study presents the results of benchmark tests performed on the forward modelling module of Basin3D, versus Geomodeler software (Geophysics, 2015), which uses highly comparable methodology. Identical structural model of Los Humeros geothermal complex (Evanno, 2017), and equivalent modelling parameters, have been used in both the forward modelling module of Basin3D and forward modelling module of Geomodeler. Results presented in this study have been assessed using Mean Absolute Error (MAE) for error calculation, and Root Mean Square Error (RMSE) to MAE ratio for evaluation of the error distribution, also Pearson's Correlation Coefficient (PCC) has been used for equivalence assessment. We found that the gravity anomaly modelled in Basin3D correlates very well with the gravity anomaly modelled in Geomodeler, reaching 0.999977 of correlation. Errors are also minimal, lower than 0.1 mGal, and the error distribution is satisfactory at a ratio equivalent of 1.252. Thus, we find that Basin3D performs well in the benchmark test, and it is a reliable tool for forward modelling of gravity anomaly.

Next, we use the same structural model of Los Humeros geothermal area and the results of the forward gravity modelling, together with the digitised map of gravity anomaly for the area, to evaluate the validity of the structural model by Evanno (2017). We find that the correlation between the modelled and observed gravity is very low, at only 0.24370, and the error is high at around 10 mGal. Thus, the modelled distribution of the anomalies and their amplitudes do not match the observed gravity anomalies. Therefore, to achieve a better match, we inversely model the gravity anomaly and corresponding densities with use of data assimilation of the measured values (digitised observed gravity map). The results presented in this study allow for identification of three main areas of density anomaly within the Los Humeros area, with the most significant low density anomaly located just below the Los Humeros caldera. Low density anomaly below the caldera may indicate

the presence of hotter material, possibly a magma chamber. There is no current consensus on the location, structure and extend of the magma chamber. Therefore, any constrains on its depth and location below the Los Humeros caldera are of great importance for the geothermal prospecting of the Los Humeros geothermal complex. Thus, based on the location of the density anomalies and the direction of change, we suggest two new structural models. First, we suggest a model with limestone layer of irregular thickness, thicker in the south-west of the model and thinner in the north-east, to compensate for the low and high gravity anomalies respectively. The second model is proposed with an additional low density layer, possibly hot material, reaching the shallowest depths just below the Los Humeros caldera. We suggest forward modelling of gravity anomaly will be performed in the future studies for the two proposed models, and then compared to the measured values of gravity anomaly. Thus, either validating or rejecting the hypothesis that the observed low gravity anomaly below the caldera is caused by a less dense material, possibly a magma chamber. Finally, we conclude that, both the forward and inverse modelling of the gravity anomaly for the Los Humeros case study produced results which will aid future research in the field.

ACKNOWLEDGEMENTS

I would first like to thank my thesis supervisor Mark Vrijlandt of the Applied Geosciences at TNO for the patient guidance, encouragement and advice during my research project. He has allowed this paper to be my own work, but steered me in the right direction whenever I needed it. His availability to answer my questions at any time of the day has cost him greatly, leaving many of his sandwiches half eaten. Thank you.

I would also like to thank Dr. Fred Beekman in the Tectonics Group of the Department of Earth Sciences at Utrecht University, for supervision of my thesis and advice on structuring my project. His apparent confidence in my abilities as an aspiring scientist, gave me the courage to be bold in the conclusions of this thesis. I am grateful to Jan Diederik van Wees of the Applied Geosciences at TNO for granting me this thesis project and acquainting me with the fascinating case study of Los Humeros geothermal complex.

Furthermore, I would like to thank Dr. Jon Limberger in the Tectonics Group of the Department of Earth Sciences at Utrecht University, for offering his time in the final stages of his PhD project, to help me get started with my master thesis. To Eszter Békési, in the Tectonics Group of the Department of Earth Sciences at Utrecht University, thank you for introducing me to the Applied Geosciences group at TNO and suggesting me as an intern student.

A very special mention goes to my partner, Ezequiel Julian Bianco Martinez, who has endlessly endured our dinnertime discussions on error distribution.

To the greatest, most determined and selfless person I know, my beloved grandmother Władysława Dyląg, thank you for teaching me that learning vowels is easier than I thought. You have taught me that to reach the peak you must climb the mountain. You have taught me how to fight for my dreams. I would also like to thank my mother, without whom I would not widen my horizons.

Finally, I would also like to thank my internship buddies, thank you for helping me sustain my caffeine addiction and for a cheerful banter on a daily basis.

TABLE OF FIGURES

FIGURE 1. PADDING OF THE STRUCTURAL MODEL BEFORE TRANSFORMING INTO FOURIER DOMAIN.....	4
FIGURE 2. PADDING OF THE STRUCTURAL MODEL IN X AND Y DIRECTIONS IN GEOMODELLER.	7
FIGURE 3. EXTEND OF THE TRANS MEXICAN VOLCANIC BELT (TMVB).	9
FIGURE 4. DIGITAL ELEVATION MODEL OF LOS HUMEROS AREA.....	11
FIGURE 5. GEOLOGICAL MAP OF LOS HUMEROS AREA BY CARRASCO-NÚÑEZ ET AL., 2017.....	11
FIGURE 6. CROSS-SECTION THROUGH LOS HUMEROS.	13
FIGURE 7. STRUCTURAL MODEL OF THE LOS HUMEROS AREA BY VERMA ET AL., 2011.	14
FIGURE 8. CUBE TEST MODEL.	3
FIGURE 9. EXTENDED CUBE TEST MODEL.	6
FIGURE 10. CROSS-SECTION OF THE LOS HUMEROS CALDERA FROM WEST TO EAST.....	1
FIGURE 11. 3D STRUCTURAL MODEL BY EVANNO, 2017.	2
FIGURE 12. CROSS-SECTION OF THE STRUCTURAL MODEL BY EVANNO, 2017.	2
FIGURE 13. STRUCTURAL MODEL OF LOS HUMEROS CREATED IN BASIN3D PREPROCESS MODULE..	4
FIGURE 14. SOUTH TO NORTH CROSS-SECTION OF THE LOS HUMEROS STRUCTURAL MODEL..	4
FIGURE 15. WEST TO EAST CROSS-SECTION OF THE STRUCTURAL MODEL OF LOS HUMEROS.	4
FIGURE 16. TOPOGRAPHY GRID OF THE LOS HUMEROS STUDY AREA.	6
FIGURE 17. FORWARD MODELLED GRAVITY ANOMALY – RESULT FROM BASIN3D.	7
FIGURE 18. FORWARD MODELLED GRAVITY ANOMALY – RESULT FROM GEOMODELLER.	7
FIGURE 19. COMPARISON OF FORWARD MODELLING IN BASIN3D VS. GEOMODELLER.....	9
FIGURE 20. VISUALISED DEGREE OF EDGE CUT-OFF IN THE ANALYSIS OF THE EDGE ERROR CONTRIBUTION.	10
FIGURE 21. ANALYSIS VERSUS EDGE CUT-OFF EQUIVALENT TO MODEL DEPTH, EXPRESSED IN THE PERCENT OF THE MODEL DEPTH.....	12
FIGURE 22. DIFFERENCE PLOT FOR 15 CELL CUT-OFF.	13
FIGURE 23. MEASURED GRAVITY MAP BY GARCÍA-ESTRADA (1992).....	16
FIGURE 24. DIGITISED MEASURED GRAVITY MAP WITH ORIGINAL UNITS OF 0.1 MGAL BY GARCÍA-ESTRADA (1992).	18
FIGURE 25. TOPOGRAPHY GRID OF THE LOS HUMEROS STUDY AREA VISUALISED IN THE QGIS SOFTWARE..	18
FIGURE 26. DIGITISED MEASURED GRAVITY MAP WITH ORIGINAL UNITS OF 0.1 MGAL BY GARCÍA-ESTRADA (1992), OVERLAID OF THE TERRAIN VIEW OF THE LOS HUMEROS STUDY AREA (FROM GOOGLE MAPS, 2018).	19
FIGURE 27. GRAVITY ANOMALY MAP (BASED ON GRAVITY MAP FROM GARCÍA-ESTRADA, 1992) WITH UNITS IN MGAL AND DISCRETISED TO 100 CELLS	19
FIGURE 28. FORWARD MODELLED GRAVITY ANOMALY. MODELLING BASED ON THE STRUCTURAL MODEL BY (EVANNO, 2017).....	21
FIGURE 29. GRAVITY ANOMALY MAP (BASED ON GRAVITY MAP FROM GARCÍA-ESTRADA, 1992) WITH UNITS IN MGAL AND DISCRETISED TO MATCH THE FORWARD MODEL DISCRETISATION.	22
FIGURE 30. DIFFERENCE BETWEEN THE FORWARD MODELLED GRAVITY ANOMALY AND THE OBSERVED DATA.....	22
FIGURE 31. INVERSE MODELLED GRAVITY ANOMALY AT DIFFERENT NUMBER OF ITERATIONS AND ENSEMBLES.....	27
FIGURE 32. INVERSE MODELLED GRAVITY ANOMALY USING DIFFERENT NUMBER OF OBSERVED DATA POINTS DURING DATA ASSIMILATION.	28
FIGURE 33. INVERSE MODELLED GRAVITY ANOMALY AT DIFFERENT VALUE OF THE OBSERVED DATA ERROR.	31
FIGURE 34. INVERSE MODELLED GRAVITY ANOMALY WITH A DENSITY SHIFT OF 500 KG/M ³ IN DIFFERENT LAYERS.	35
FIGURE 35. ALTERNATIVE STRUCTURAL MODEL OF LOS HUMEROS CREATED IN BASIN3D PREPROCESS MODULE.....	37
FIGURE 36. SOUTH TO NORTH CROSS-SECTION OF THE ALTERNATIVE LOS HUMEROS STRUCTURAL MODEL.....	37
FIGURE 37. WEST TO EAST CROSS-SECTION OF THE ALTERNATIVE LOS HUMEROS STRUCTURAL MODEL.....	37
FIGURE 38. FORWARD MODELLED GRAVITY ANOMALY. MODELLING BASED ON THE ALTERNATIVE STRUCTURAL MODEL.	39
FIGURE 39. FORWARD MODELLED GRAVITY ANOMALY. MODELLING BASED ON THE STRUCTURAL MODEL BY (EVANNO, 2017).....	39
FIGURE 40. GRAVITY ANOMALY MAP (BASED ON GRAVITY MAP FROM GARCÍA-ESTRADA, 1992).	40
FIGURE 41. DIFFERENCE BETWEEN THE FORWARD MODELLED GRAVITY ANOMALY (ALTERNATIVE STRUCTURAL MODEL) AND THE OBSERVED DATA.	40
FIGURE 42. INVERSE MODELLED GRAVITY ANOMALY WITH A DENSITY SHIFT OF 500 KG/M ³ IN BOTH LIMESTONE AND ANDESITE LAYERS (ALTERNATIVE STRUCTURAL MODEL)	42
FIGURE 43. NEW SUGGESTED STRUCTURAL MODELS FOR LOS HUMEROS STUDY AREA.	45
FIGURE 44. MEASURED GRAVITY ANOMALY FROM ARZATE ET AL. (2018)..	46

TABLE OF TABLES

TABLE 1. TABLE OF BOUNDARY AND GEOLOGICAL CONDITIONS USED IN STRUCTURAL MODEL BY VERMA ET AL., 2011.	13
TABLE 2. FIVE HYPOTHETICAL SETS (CASES) OF 4 ERRORS, AND THEIR CORRESPONDING MAE AND RMSE. FROM: WILLMOTT & MATSUURA, 2005.	2
TABLE 3. ABSOLUTE VALUE COMPARISON BETWEEN BASIN3D MODEL AND GEOMODELLER MODEL.....	5
TABLE 4. DIFFERENT EXPANSIONS OF GEOMODELLER MODEL INTO FOURIER DOMAIN VERSUS AN EXPANSION AT 100% AND EQUIVALENT OF BASIN3D SETTING.....	14
TABLE 5. INVERSE MODELLED GRAVITY ANOMALLY AT DIFFERENT NUMBER OF ITERATIONS AND ENSEMBLES COMPARED TO OBSERVED DATA – TOP SECTION.....	25
TABLE 6. INVERSE MODELLED GRAVITY ANOMALY AT DIFFERENT VALUE OF THE OBSERVED DATA ERROR COMPARED TO OBSERVED DATA..	30
TABLE 7. INVERSE MODELLED GRAVITY ANOMALY WITH A DENSITY SHIFT OF 500 KG/M ³ IN DIFFERENT LAYERS COMPARED TO OBSERVED DATA..	34

TABLE OF CONTENTS

ABSTRACT	I
ACKNOWLEDGEMENTS	III
TABLE OF FIGURES	IV
TABLE OF TABLES	V
INTRODUCTION	1
METHODOLOGY	3
<i>Basin3D Forward modelling</i>	3
<i>Basin3D Inverse modelling</i>	4
<i>Geomodeller Forward Modelling</i>	6
<i>Geomodeller Inverse Modelling</i>	7
GEOLOGY OF LOS HUMEROS	8
BENCHMARKING	15
<i>Structural Models</i>	15
<i>Methodology for visualising and comparing the results</i>	1
<i>Cube Test – spatial model</i>	2
<i>Cube Test – forward modelling parameters</i>	3
<i>Cube Test – results</i>	4
<i>Extended Cube Test – spatial model</i>	5
<i>Extended Cube Test – forward modelling parameters</i>	6
<i>Extended Cube Test – results</i>	7
<i>Los Humeros Model – spatial stratigraphic model</i>	1
<i>Los Humeros Model – forward modelling parameters</i>	5
<i>Los Humeros Model – results</i>	5
<i>Results</i>	5
<i>Edge thickness check</i>	10
<i>FFT check</i>	13
<i>Benchmarking Discussion</i>	14
LOS HUMEROS - INVERSE MODELLING	16
<i>Observed Gravity Data</i>	16
<i>Comparison to Forward model</i>	20
<i>Inverse Modelling Set-up</i>	22
<i>Iterations and Ensembles</i>	23
<i>Number of Observed Data Points</i>	25
<i>Observed Error</i>	29

<i>First Inverse Model of Los Humeros</i>	32
<i>Alternative Model of Los Humeros</i>	36
<i>Structural Model</i>	36
<i>Forward Model</i>	38
<i>Inverse Model</i>	40
<i>Inverse Modelling Discussion</i>	43
CONCLUSIONS	47
REFERENCES	48
APPENDIX I: ADDITIONAL BENCHMARKING FIGURES AND TABLES	51
APPENDIX II: ADDITIONAL MODELLING FIGURES AND TABLES	66

INTRODUCTION

The structure of the shallow and deep subsurface is rarely well described, due to the high cost of direct exploration (drilling deep wells) and limited downward extrapolation from field geology to only a few km. Thus, knowledge of the topmost 10 km of the subsurface is mostly obtained by employing geophysical methods, such as reflection seismics, potential field methods (gravity and magnetics), and electrical and electromagnetic methods (Burger et al., 2006). In this project we investigate the use of the gravity method to determine the structure of the subsurface.

If initially horizontal sedimentary layers are deformed by folding or offset by faults, or the subsurface structure is inhomogeneous due to other factors, the resulting lateral and vertical changes in density generate a change ('anomaly') in the gravitational acceleration at the Earth's surface. Conversely, when gravity measurements at the Earth's surface show the presence of gravity anomalies, these indicate the occurrence of density variations in the subsurface. The challenging aspect of the gravity method is the extraction of the information about the structure of the subsurface, which is unknown, based on the gravity anomalies measured at the surface. This is not a straightforward problem as the gravity anomalies depend on both, the magnitude of and the distance to, the density contrast between the layers. Moreover, the measured gravity anomaly at each surface location is not caused by a single density contrast, but by multiple density contrasts that might be present in the subsurface.

To tackle these issues, both forward and inverse modelling methodologies have been developed, and implemented in various software. The comparison between forward modelled and observed gravity anomalies at the Earth's surface, allows the validation of the structural models of the subsurface. If the agreement between the forward modelled gravity anomaly and the observed gravity anomaly is poor, the inverse modelling may aid in characterisation of the problem areas. The forward operator in gravity modelling is non-linear, meaning that in inverse modelling an infinite number of parameter (density) configurations could satisfy the observations, or that the multiple solutions fall within the noise level, or that the solution may not exist at all (Glegola, 2013). Thus, there must be regularisation constraining the inverse model, for example, by using data assimilation of the observed measurements and a limited variation of density that is bound to the physical range (Emerick & Reynolds, 2013 [a]). Forward modelling of the gravity anomaly, is an important preliminary step before the data assimilation and inverse modelling can be performed. The accuracy of the gravity anomaly computed based on the structural model must be high, as these values determine the

amount of shift (within the allowed range), to the density parameter when data assimilation is performed.

Basin3D software developed by TNO and implemented in Java (van Wees, 2016; Struijk, 2016; den Hollander, 2016), it is a forward and inverse gravity modelling software based on the methodology of Caratori Tontini et al. (2009) and Cooley-Tukey algorithm (Cooley & Tukey, 1965). The first aim of this study is to benchmark the forward modelling module of the Basin3D software against comparable methodology. For that purpose, the forward modelling module of the Geomodeller software is used (Geophysics, I., 2015). Both Basin3D and Geomodeller, are based on work of Plouff (1976), Okabe (1979) and Holstein (2003), in the field of gravity anomaly calculation, and work of Bhattacharyya (1967), Parker (1972) and Cooley & Tukey (1965) on equations for calculations in the Fourier Domain. Thus, the methodology of Basin3D is considerably comparable to the methodology of Geomodeller. We use the same structural model, with identical discretisation and densities, in both Basin3D and Geomodeller. Next, we investigate the performance of the Basin3D forward module and determine the level of confidence in the quality of the results.

A case study based on the structural model of Los Humeros geothermal field in Mexico is used primarily to benchmark the Basin3D against Geomodeller. Then, we compare the forward modelled gravity anomaly of Los Humeros to the observed values, and further investigate the density distribution in the subsurface by using inverse modelling and data assimilation. Los Humeros geothermal field is of a great interest, due to its operational power plant, which is the 3rd largest geothermal power plant in Mexico with total installed capacity of 94 MW, 58 drilled wells and 27 wells in production (Carrasco-Núñez et al., 2017; Peiffer et al., 2018). At present, the depth and extend of the magma chamber is poorly constrained, and most of the geological studies in the area focus on the shallower subsurface (Carrasco-Núñez et al., 2017). Thus, a second aim of this study is to find an indication of density anomaly below Los Humeros geothermal complex.

METHODOLOGY

Basin3D Forward modelling

The gravity field, or vertical gravity anomaly Δg_z at the Earth's surface ($z=0$), can be calculated for the density distribution $\rho(x)$ using the following equation:

$$\Delta g_z(x_0) = G \int_{\mathbb{R}^3} \frac{\rho(x)(z - z_0)}{|x - x_0|^3} dx, \quad Eq. 1$$

where, G ($N * m^2/kg$) is the gravitational constant, $\rho(x)$ (kg/m^3) is the density and $xx = (x_0; y_0; z_0)$ is the position of the measuring point. A numerical calculation of the gravity field in the spatial domain is time consuming and requires substantial computing memory. Alternative much faster method of the calculation was proposed by Caratori Tontini et al. (2009), in which the integration in the spatial domain is expressed as a convolution in the Fourier domain. This method is based on using the Fourier Transform to first express the gravity equation (Equation 1) in the Fourier domain, and then using the convolution theorem along with Fourier differentiation theorem, to simplify it to a multiplication:

$$\mathcal{F}_{3D}[\Delta g_z(x_0)] = i4\pi G \frac{k_z}{|k|^2} \mathcal{F}_{3D}[\rho]. \quad Eq. 2$$

where, \mathcal{F}_{3D} is the 3D Fourier operator, $x = (x; y; z)$ is the spatial vector, $k = (k_x; k_y; k_z)$ is the 3D wave vector and R_3 is the domain where $f(x)$ is defined.

Implementation of this methodology means that the gravity field for the 3D volume can be obtained by first applying the Fourier Transform, then the gravity equation (Equation 2) in the Fourier domain, and finally transforming back to the spatial domain by applying the Inverse Fourier Transform (den Hollander, 2016). This transforming of the calculation is faster and more memory efficient than a numerical computation.

Basin3D software uses the Caratori Tontini et al. (2009) methodology in Java based on the Cooley-Tukey algorithm (Cooley & Tukey, 1965), this means the input must be a 3D density array with dimensions 2^n by 2^n by 2^n . The density array in the spatial domain is described by prismatic finite size grid with dx , dy and dz , and the points containing the density are located within a regular grid, and stored in the whole volume of the cell. Before the transformation into the Fourier domain, the density array is padded with zeros by exactly the same number of cells as the model size in the z-plus

direction (below zero). Then, the densities are extrapolated outwards (Figure 1). Once this is done, the spatial domain is transformed to the Fourier domain, then gravity computation is performed and finally the inverse transform is applied. Once transformed back to the spatial domain the results of the gravity anomaly can be read for a given depth.

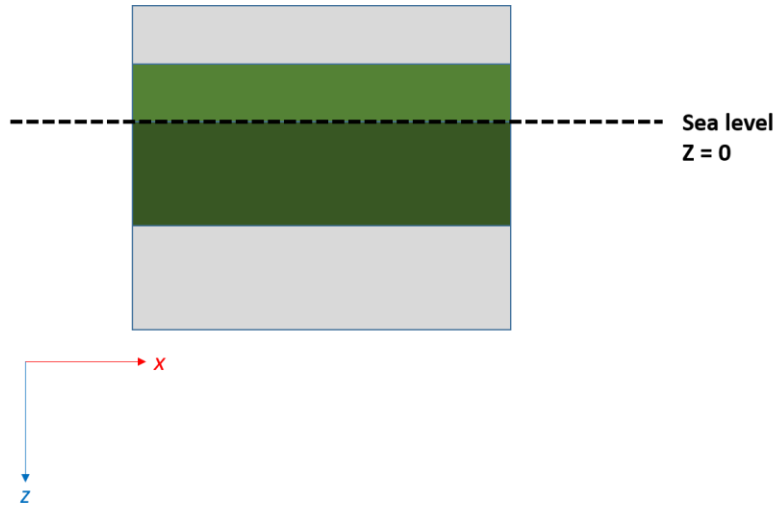


Figure 1. Padding of the structural model before transforming into Fourier Domain. Model space – green, padding – grey.

Basin3D Inverse modelling

The subsurface is not directly observed, but rather indirectly measured using methodologies that carry assumptions and errors. Thus, we can pose an inverse problem that aims to compute a more accurate representation of the subsurface.

$$y = F(x) + e \quad \text{Eq. 3}$$

where, y is the observation vector, $F(x)$ is the gravity equation and e is the error. In case of gravity, the forward operator is a non-linear function, thus the inverse problem is ill posed (Tarantola, 2005). This means that an infinite number of parameter (density) configurations could satisfy the observations, or that the multiple solutions fall within the noise level, or that the solution may not exist at all (Glegola, 2013). To overcome this problem, a regularisation must be introduced, and here it is in the form of data assimilation of the observed measurements (Emerick & Reynolds, 2013 [a]).

Additionally, the variation to the density parameter is bounded to occur only in the specified layers and only within the physical range.

Data assimilation of the observed measurements into the numerical model allows to forecast the parameter (density) in the model to fit the observed data. During the data assimilation, the observed values are compared to the modelled ones (Emerick & Reynolds, 2013 [a]). If they do not match, and after taking to account the uncertainties in both, a new forecast value for the parameter is computed that better matches the observed values and still falls within the bounds of the model (Glegola, 2013). Basin3D uses the ensemble smoother with multiple data assimilation (which is an extension of the Kalman filter) as the data assimilation method (Evensen, 1994). This implies that the data assimilation is in Bayesian framework and uses a global update on all available data for very non-linear systems, and with some additional Monte-Carlo methods (Ensemble Kalmar Filter) to predict the inflated covariance matrix of the measurement errors (Oliver et al., 2008; Emerick & Reynolds , 2013 [b]; Glegola, 2013).

This methodology is implemented in Java (van Wees, 2016; Struijk, 2016; den Hollander, 2016), and the main input are the observed and modelled values both shifted to zero. The layer to be varied during the data assimilation must be selected, and the parameters specifying the variation must be set: firstly the value range for the density variation, and secondly a method for adjusting the density values, either scaling or shifting. Furthermore, range of cells that determines the smoothness of the density change during the data assimilation, also must be chosen. Finally, the type of the value distribution during the density variation must be specified, either a triangular or uniform distribution. Once the settings for the model parameters are set, the file with observed data is uploaded. This file must include xyz coordinates, measured parameter values, chosen measurement error and depth of the measurement. Lastly, the number of the iterations can be chosen. One iteration has been found to be sufficient for a linear problem, however for non-linear problems four iterations are necessary (Emerick & Reynolds, 2013 [b]). Number of ensembles must also be specified, with larger numbers allowing more accurate results. The resulting forecast density and the corresponding gravity anomaly can be read at chosen depth.

Geomodeller Forward Modelling

Geomodeller software has two options for forward gravity modelling, the spatial convolution option and the Fourier Transform option, both calculating the tensor grid of the potential fields. The Fourier Transform methodology of Geomodeller is based on the same theory as the Fourier transform methods of the Basin3D. Geomodeller also incorporates multi-threaded option in their software.

Geomodeller methodology uses closed-form equations for gravity anomaly calculations following the work of Plouff (1976), Okabe (1979) and Holstein (2003), and also the equations for calculations in the Fourier Domain following the papers of Bhattacharyya (1967), Parker (1972) and Cooley & Tukey (1965). Detailed descriptions of Geomodeller methodology can be found in papers by authors associated with Geomodeller software (Calcagno et al., 2008; Chiles et al., 2004; Guillen et al., 2008; Guillen et al., 2004). Basin3D methods are based on the work of Caratori Tontini et al. (2009), which was developed separately but is a continuation of earlier work in the field. Thus, methodologies of Basin3D and the Geomodeller are significantly comparable.

Before the transformation of the model from spatial to Fourier domain, it must be padded. In Geomodeller, padding is done by repeating the mirrored image of the model on each of the model sides (Figure 2) in x- y- directions. In the z-direction the vertical expansion of the model space is described in the % size of the existing model (called FFT %), and is recommended by the authors of the software to be set at a standard starting value of 400%.

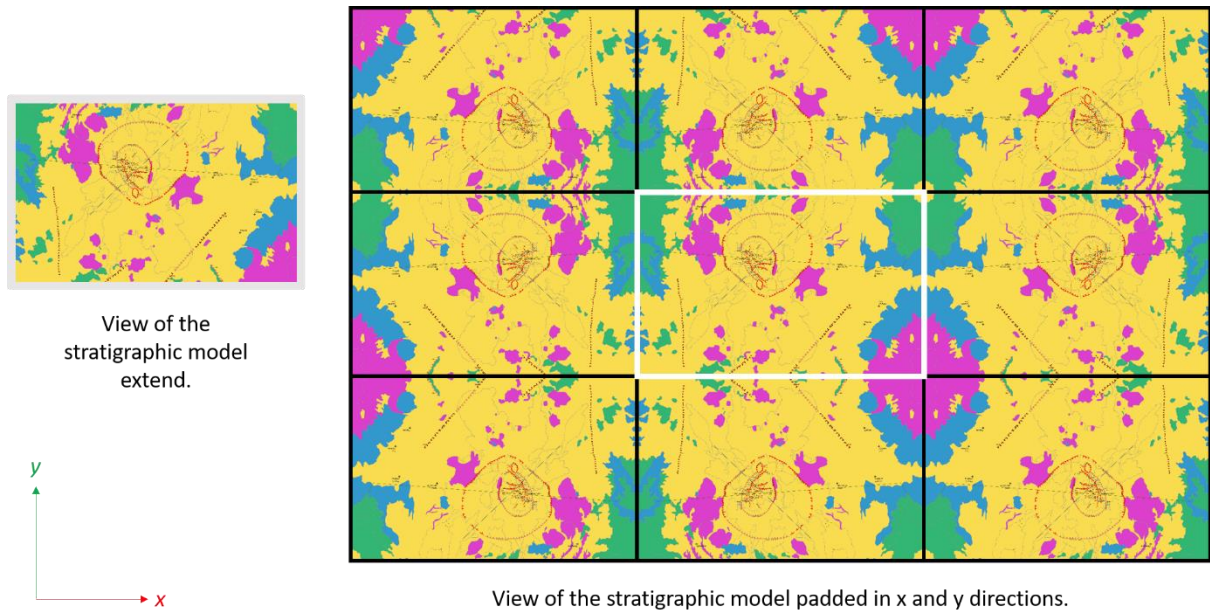


Figure 2. Padding of the structural model in x and y directions in Geomodeller, before transformation to Fourier Domain.

Geomodeller Inverse Modelling

Inverse modelling in the Geomodeller, similarly to Basin3D, uses Monte Carlo type inversion constrained by the observed data, structural model and the *a priori* probability density function (pdf). The *a priori* probability density function is a Gaussian function based on the density values incorporated in the structural model. The inverse modelled density is allowed to vary within the pdf, and must result in a fit with the gravity anomaly within the misfit error (Husson et al., 2018). The main difference of the approach to the inverse modelling between Basin3D and Geomodeller, is that in Geomodeller not only the modelled parameter can be modified, but also the geometry of the model by assigning the cell in question to a neighbouring layer (Gullien et al., 2008). Furthermore, likelihood functions are used in this methodology, which determine the acceptance or rejection of the current model into the ensemble of models for final model computation (Gullien et al., 2008). The inverse modelling module of Geomodeller will not be used in this study.

GEOLOGY OF LOS HUMEROS

The area of Los Humeros is a volcanic complex, located in the eastern part of Trans-Mexican Volcanic Belt (TMVB), 180km east of Mexico City (Figure 3). At around 97°28'0''W and 19°45'0''N, there is 25-30 km wide, basalt-andesite-rhyolite caldera of Los Humeros, and within it a younger 5-8 km oval caldera of Los Potreros (Carrasco-Núñez et al., 2017).

Los Humeros Pleistocene caldera is the northernmost volcano of the Serdán Oriental Basin (SOB), characterised by cinder cones, large rhyolitic domes, and basaltic and rhyolitic maars. The SOB is a limited basin, with the Cofre de Perote-Citlaltépetl volcanic chain to the east (large andesitic stratovolcanoes and dome complexes, Figure 3), and Tlaxco-Cerro Grande range to the west (Miocene andesitic volcanic rocks) (Carrasco-Núñez et al., 2017). Tlaxco-Cerro Grande range overlies a ≤ 3000 -m-thick (Carrasco-Núñez et al., 2017), highly deformed Mesozoic limestone, which sometimes crops out within the SOB (Figure 3). To the north of the SOB, where the Mesozoic sedimentary basement is found at shallower depths, it is deposited on top of the Teziutlán Massif, a Paleozoic crystalline complex made of metamorphic and intrusive rocks, including green schists, granodiorites and granites (Quezadas-Flores, 1961; Viniegra, 1965). Just below the Los Humeros cladera, overlaying the limestone basement is a 1.5 km thick (López-Hernández, 1995) andesitic formation, Teziutlán Andesite, associated with the Plio-Pleistocene volcanism.

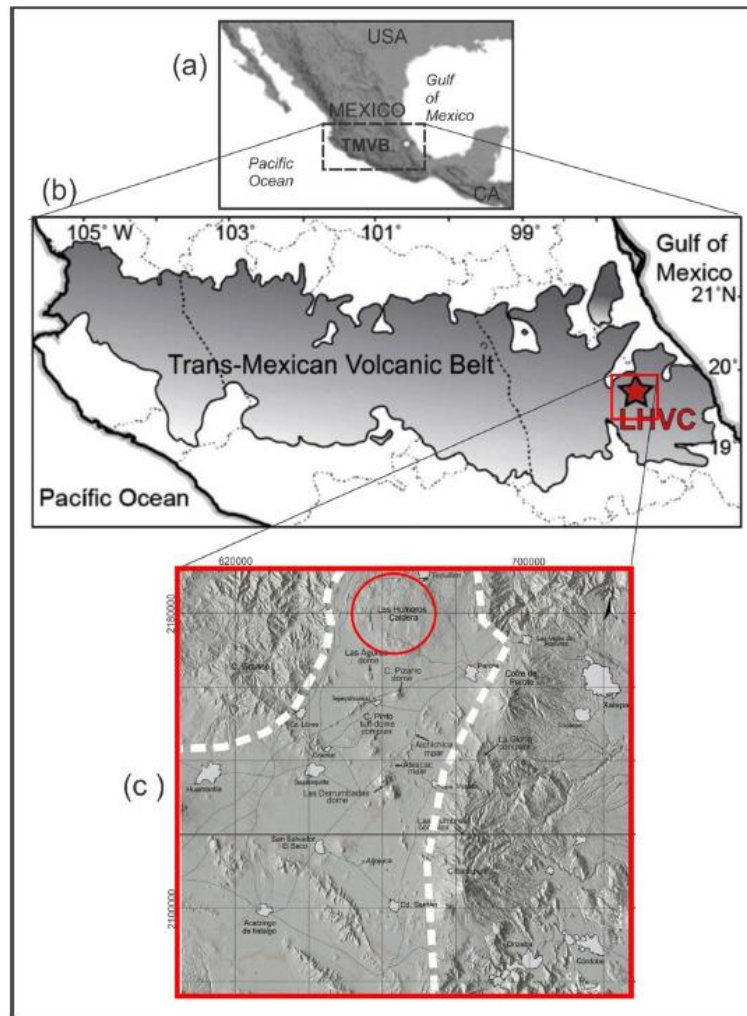


Figure 3. Extend of the Trans Mexican Volcanic Belt (TMVB) (a) and the location of Los Humeros caldera (LHVC) (b). Serdán Oriental Basin (dashed line) and the Los Humeros caldera (red circle) (c). From: Arzate et al., 2018.

The Pleistocene caldera of Los Humeros has formed after at least two major caldera-forming eruptions, several plinian eruption, alternating dacitic and rhyodacitic dome-forming eruptions, and more recently by basaltic and basaltic andesite lava flows and cinder cones (Carrasco-Núñez et al., 2017). The widely accepted development of the caldera has been described in four stages (Ferriz and Mahood, 1984). Firstly, the pre-caldera rhyolite lavas were disrupted by large caldera-forming eruption at 460 – 450 ky, which deposited large volumes of high silica rhyolitic Xáltipan ignimbrite, forming the large irregular Los Humeros caldera of 15-21 km (Figure 4). During the second stage extending from 360 ky to 140 ky, high-silica rhyolite domes were emplaced within the existing caldera, followed by multiple voluminous plinian and sub-plinian eruptions depositing rhyodacitic-andesitic pumice fallout layers, called ‘Faby Tuff’ (Carrasco-Núñez et al., 2017). Third stage started

from 140 ky, which amounts to a large caldera-forming eruption, that formed the Los Potreros caldera of 9 km (Figure 4), and deposited compositionally zoned Zaragoza ignimbrite (from rhyolitic to andesitic to rhyolitic pumice) (Carrasco-Núñez et al., 2017). Fourth stage of the geological development of the Los Humeros area, comprised of eruption of rhyodacitic and andesitic lavas (60ky), deposition of pumice known as Xoxocitic Tuff (~50ky), and breccia and pyroclastic flows known as Llano Tuff (~28ky). More recent volcanism includes andesitic and basaltic lava flows, mainly the eruption from Los Humeros caldera, which produced basaltic andesite and basalt deposits in the southern part of the caldera rim (Figure 5) (Carrasco-Núñez et al., 2017).

The rim of the Los Humeros caldera still remains partly unknown. However, Norini et al. (2015) have performed a more detailed investigation of the structural characteristics of the inner caldera, Los Potreros. The study resulted in mapping of several faults and detailed outline of the rim of the Los Potreros caldera. Another study, based on integrated well-log and seismic data (Rodríguez, 2000), proposed a cross-section of the area with a limestone basement starting from around 2000 meters below the surface (around 1000 meters above sea level). The cross-section also included an intermediate layer of Mio –Pliocene volcanic deposits, mainly andesitic lavas between the Los Humeros forming deposits and the limestone basement. The exact thickness of the limestone basement is unknown, and the presence and extend of any underlying crystalline rock is not discussed. Most of the studies on the Los Humeros area are limited to the uppermost 4km. Well data gathered from 45 wells within the Los Potreros caldera, give information about the horizons of the Pliocene volcanics and the underlying limestone (Giordano et al., 2013). The Pliocene volcanics are found as deep as at depth of 1650 meters above the sea level, suggesting a maximum thickness of the Los Humeros volcanics at 1200 meters. Top of limestone was reported as deep as 250 meters above sea level, and as shallow as 1340 meters above sea level. There is no well data available for calculating or estimating the thickness of the limestone layer.

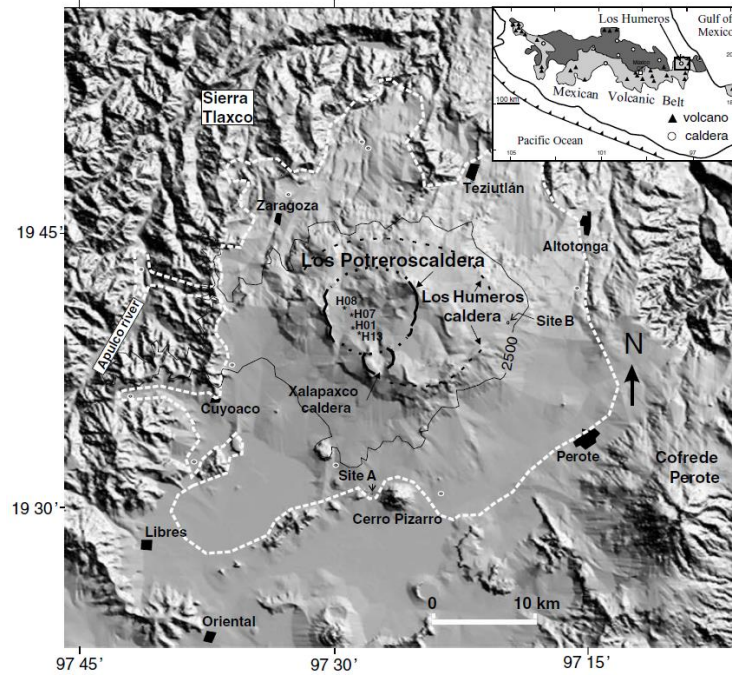


Figure 4. Digital elevation model of Los Humeros area. The Los Humeros caldera and Los Potreros caldera are located in the centre of the image. Extend of the Zaragoza ignimbrites – white dashed line. Inset shows the location of the digital elevation model within the Mexican Volcanic Belt. From: Carrasco-Núñez et al., 2017.

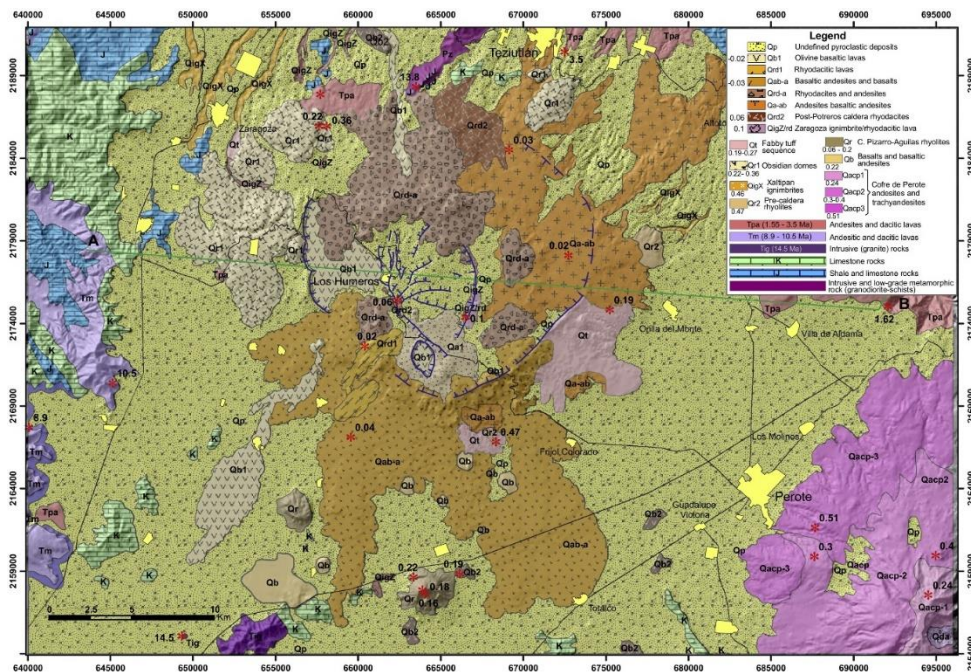


Figure 5. Geological Map of Los Humeros area by Carrasco-Núñez et al., 2017.

Compositional characteristics of the Los Humeros volcanic system shows, as measured by x-ray fluorescence data, continuous trend from mafic to high-silica rhyolites products (Verma, 2000). Carrasco-Núñez et al. (2017) studied in depth the Zaragoza ignimbrites, and in combination with the field studies of the outcrop and the well data from Cedillo (1999), identified the main compositional groups of the area to be at the uppermost the *Late Pleistocene andesites, basalts and pyroclastics*, followed by the *Xáltipan Ignimbrite*, then the *Plio-Quaternary andesites*, and the *Tertiary Tuffs, andesites and basalts*, and the *basement of Mesozoic limestone* (Figure 6). Carrasco-Núñez et al. (2017) study does not inform us about the densities of the mentioned sections. The most widely cited study of the geochemistry of rock formation for the Los Humeros volcanic systems is this of Verma (1982), however, this study does not include a density analysis. The densities of the rocks widely used in the literature, including modelling experiments by Verma himself, were measured in another study performed by Contreras et al. (1990). This study however, is not available for review, and only the density values cited by Verma in his modelling studies are available (Verma et al., 1990; Verma, 2011; Verma, 2013). Table 1 shows the parameters of a model by Verma et al. (2011; 2013), initially build with generalised densities (Verma, 1985), and then updated with densities from Contreras et al. (1990) (Verma, 1990). Verma et al. (2011) consider Los Humeros volcanic system to have four main sections, the uppermost section 4 consists of 1km thick deposit of acid volcanic rocks with density of 2360 kg/m³ (Figure 7). Section 3 is a group of intermediate volcanic rocks, represented by the density of 2394 kg/m³, and thickness of 1km. In his models, Verma assumes the thickness of Limestone to be 1km, and the density equals to 2180 kg/m³. Additionally, there is a basement of intrusive volcanic rock, which for the purpose of the models have 17 km thickness, and density of 2680 kg/m³. Verma's models have been set up for purpose of the heat flow calculation and emplacement of the magma chamber, thus additionally, the models also include a magma chamber at the depth of 5km (Figure 7). These studies give an alternative insight to the deeper structure of the Los Humeros system, in contrast to the previously mentioned geological studies, which only focus on the uppermost 4km of the strata.

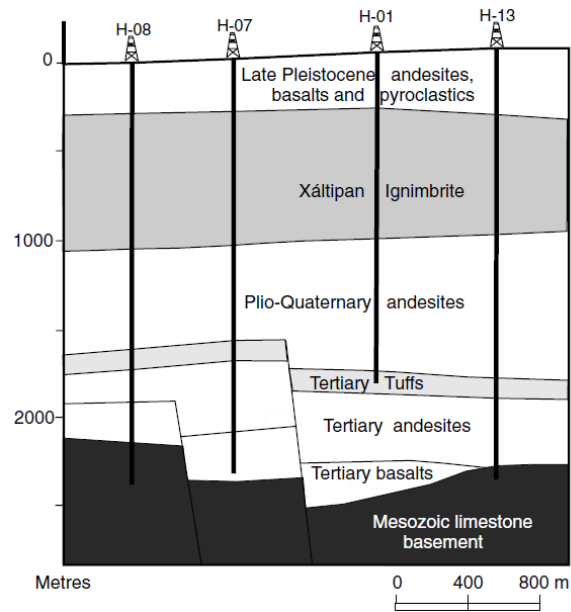


Figure 6. Cross-Section through Los Humeros. Example of the topography and depth of Mesozoic Limestone basement. From: Cedillo, 1999.

Physical property (units)	Emplacement of magma chamber
<i>Mesh construction information</i>	
Total modeled length in the x direction (km)	30
Number of control volumes in the x direction	120
Total modeled length in the y direction (km)	30
Number of control volumes in the y direction	120
Total modeled length in the z direction (km)	20
Number of control volumes in the z direction	80
Mesh size $\delta x, \delta y, \delta z$ (km)	0.25, 0.25, 0.25
<i>Emplacement conditions</i>	
Chamber volume (V_{cham}) (km^3)	1400
Chamber radius (r_{cham}) (km)	8.5
Depth of the top of the chamber (d_{cham}) (km)	5
Depth of the chamber centroid (d_{cham} ; km)	8.125
Depth of the bottom of the chamber (km)	11.125
Magma emplacement temperature (T_{cham}) ($^{\circ}\text{C}$)	1350
<i>Characteristics of the four main geological sections</i>	
Thickness of section 1 – intrusive rocks (km)	17.0
Thermal conductivity (W/mK)	2.843
Specific heat (J/kg K)	914
Density (kg/m^3)	2680
Thickness of section 2 – limestone (km)	1.0
Thermal conductivity (W/mK)	2.705
Specific heat (J/kg K)	854
Density (kg/m^3)	2180
Thickness of section 3 – intermediate volcanic rocks (km)	1.0
Thermal conductivity (W/mK)	1.673
Specific heat (J/kg K)	1009
Density (kg/m^3)	2394
Thickness of section 4 – acid volcanic rocks (km)	1.0
Thermal conductivity (W/mK)	1.738
Specific heat (J/kg K)	885
Density (kg/m^3)	2360
<i>Boundary conditions</i>	
Surface temperature (T_s) ($^{\circ}\text{C}$)	25
Temperature gradient (ΔT_g) ($^{\circ}\text{C}/\text{km}$)	30
<i>Simulation time</i>	
Discretization time (δt) (year)	10
Total simulation time (t) (year)	500,000

Table 1. Table of boundary and geological conditions used in structural model by Verma et al., 2011.

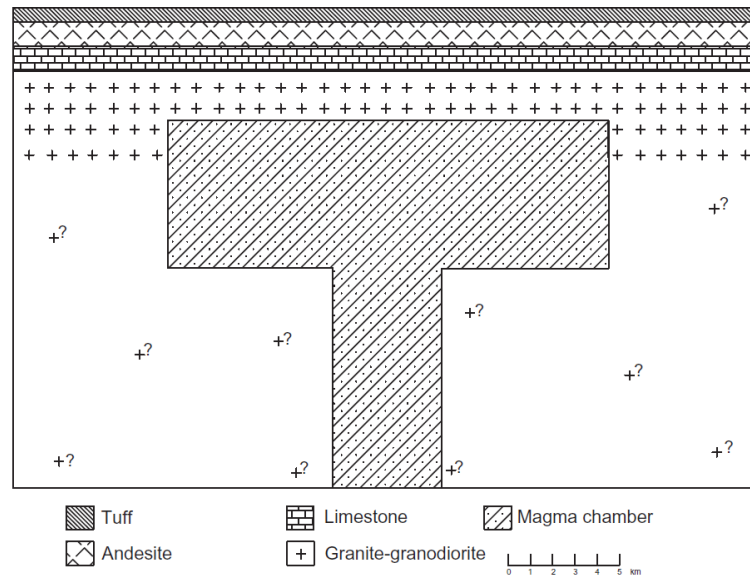


Figure 7. Structural model of the Los Humeros area by Verma et al., 2011. Shallow magma chamber surrounded by higher density crustal layer.

The study area of Los Humeros is an active geothermal system, with a currently operational geothermal power plant, which is the 3rd largest geothermal power plant in Mexico with 94 MW total installed capacity, 58 drilled wells and 27 wells in production (Carrasco-Núñez et al., 2017; Peiffer et al., 2018). As previously mentioned, the exact location, depth and extend of the magma chamber below the Los Humeros caldera is not known. It is however, an important information when estimating the geothermal potential of the field, especially when considering the duration of the existing extraction projects and their possible expansion. Thus, in this study we will use gravity forward and inverse modelling to identify the possible density anomalies within a structural model of the Los Humeros system.

BENCHMARKING

Structural Models

All the spatial models have been built in the Basin3D Preprocess module, which allows to build 3D stratigraphic models and save them in voxel file format. The Preprocess module allows for input of spatial model extend in xyz directions. Both min and max for each direction are specified separately, and cell discretization in xyz dimensions, can also be specified separately. Additionally, the discretisation in the z direction can be adjusted for a specific depth to be more or less coarse than the rest of the model. This feature will not be used in this study, to allow the resulting voxets to be used with Geomodeller, and to minimize the spatial domain roundup errors when using the same voxel in both softwares.

The input data necessary for the model building are the spatial extends of the model and the cell discretisation as mentioned above. Additionally, the spatial information about the stratigraphic layers must be specified, this includes the depth (or elevation) of the top surface of the layer and its thickness. Also, the geological ID (name) of the formation must be specified for each layer, this must correspond to the formation ID in the rock properties file (editable), which was created based on Hantschel and Kauerauf (2009). We focus on the density of the layers, thus other properties like porosity, remain constant.

Basin3D Preprocess module will output the master file (spatial file) of the calculated voxel as a .vo file, additionally all other parameters will be stored in binary files and referenced in master file. These parameters, among others, include layer ID, depth (Z) and density (RHO). Additionally, a .csv file is output containing all the spatial coordinates of the voxel and corresponding parameters, for each cell of the model.

Methodology for visualising and comparing the results

Results of the Basin3D have been automatically saved as voxel file (.vo) and also as a comma separated file (.csv). Thus, there is no need for additional processing of the data before plotting and analysing. However, the results of Geomodeller are output in a voxel file (.vo) and in a raster file (.ers). To make the data more accessible we import the raster file into QGIS software and then save it as a geometry file (.xyz). The data in a geometry file format can then be read in Excel, where appropriate headers can be added to match the formatting of the .csv files output by Basin3D.

We use Python programming language to plot and analyse the results of Basin3D and Geomodeller. Firstly, we compute the mean value of the data and remove it from the data set, doing this for both Basin3D and Geomodeller data sets. This allows us to use the computed gravity anomaly, and not the absolute values, which will make it easier for comparison between the data sets. After gridding the data with respect to the x and y cell coordinates, we plot each data set individually as an anomaly map. We also plot a histogram for each of the data sets, to check the distribution of the values. As a next step in the analysis of the results we compare the two data sets to each other, to do this we subtract the data from each other and plot them as a map and histogram of the differences.

Additionally, we calculate several different measures of error for the comparison between the results of Basin3D and Geomodeller. We use Pearson's Correlation Coefficient (PCC) to measure the similarity between the modelled gravity anomaly in Basin3D and the modelled gravity anomaly in Geomodeller. PCC is used due to its sensitivity to a linear relationship between the two data sets (Buda & Jarynowski, 2010). In case of benchmarking the modelled gravity anomaly, this measure allows us to identify if the locations of the gravity anomalies match well between the two models. To measure the difference between the two techniques, we use Mean Absolute Error (MAE), which is the mean of the absolute errors, i.e. it is the average absolute difference between the same element in the two data sets (between y_i and x_i). MAE is a dimensioned measure of model performance as it has the same scale as the data (Willmott & Matsuura, 2005). We also use Root Mean Square Error (RMSE), mainly for checking the variability within the distribution of error magnitudes. RMSE is a square root of the average of squared errors, also dimensioned measure, thus it is very sensitive to outliers (Willmott & Matsuura, 2005). When frequency distribution of error magnitudes increase, the MAE remains the same but the RMSE increases, this is well illustrated in the Table 2 (from Willmott & Matsuura, 2005). Furthermore, RMSE is always higher or equal to MAE, since it is based on the

squares of errors and thus it has high sensitivity to outliers. We can see in Table 2 that closer the RMSE is to MAE, better the distribution of error magnitudes or average error magnitude. We will use this quality to assess the error distribution when comparing our two data sets, and as we will discuss below, this will help to pinpoint the possible cause of the differences. Due to the prevalent misuse of RSME in the earth science literature, it is important to stress that RSME is used in this study not as a measure of ‘average error’ or to compare the two models, but as means to evaluate the changing distribution of error magnitudes when investigating specific scenarios.

Variable	Case 1	Case 2	Case 3	Case 4	Case 5
e_1	2	1	1	0	0
e_2	2	1	1	0	0
e_3	2	3	1	1	0
e_4	2	3	5	7	8
$\sum e_i $	8	8	8	8	8
MAE	2	2	2	2	2
$\sum e_i ^2$	16	20	28	50	64
RMSE	2.0	2.2	2.6	3.5	4.0

Table 2. Five hypothetical sets (cases) of 4 errors, and their corresponding MAE and RMSE. From: Willmott & Matsuura, 2005.

Cube Test – spatial model

A cube test was designed to check the spatial orientation of the model in Geomodeller and the input and output parameters of the software. The file formats of the Geomodeller proved to be a challenge, the xyz orientation of the model and the corresponding results had to be checked with a simple cube model. To simplify the model, each cell is a cube as well as the overall shape of the model (Figure 8). The dimensions of the cell in xyz directions are 100m by 100m by 100m, and the dimensions of the whole model in x- and y- direction are 6.4km by 6.4km. In the z direction the model is 3.2km above zero (sea level) and 3.2km below zero. There are three layers with different densities, including layer one (blue in Figure 8) has the density of 2360 kg/m³ corresponding to light pyroclastics like tuff, then layer two (green) has a much higher density of 2680 kg/m³ corresponding to granite, and the last layer (red) has a density of 2650 kg/m³ corresponding to andesite. These densities have been chosen to match some of the layers of the case study of Los Humeros, and are evenly distributed within the assigned layers, there is no averaging or interpolation of the values. The top tuff layer has a thickness

of 800 meters, the granite layer has a thickness of 1200 meters and the third bottom layer (andesite) extends from 1000 meters above sea level, all the way to the bottom of the model.

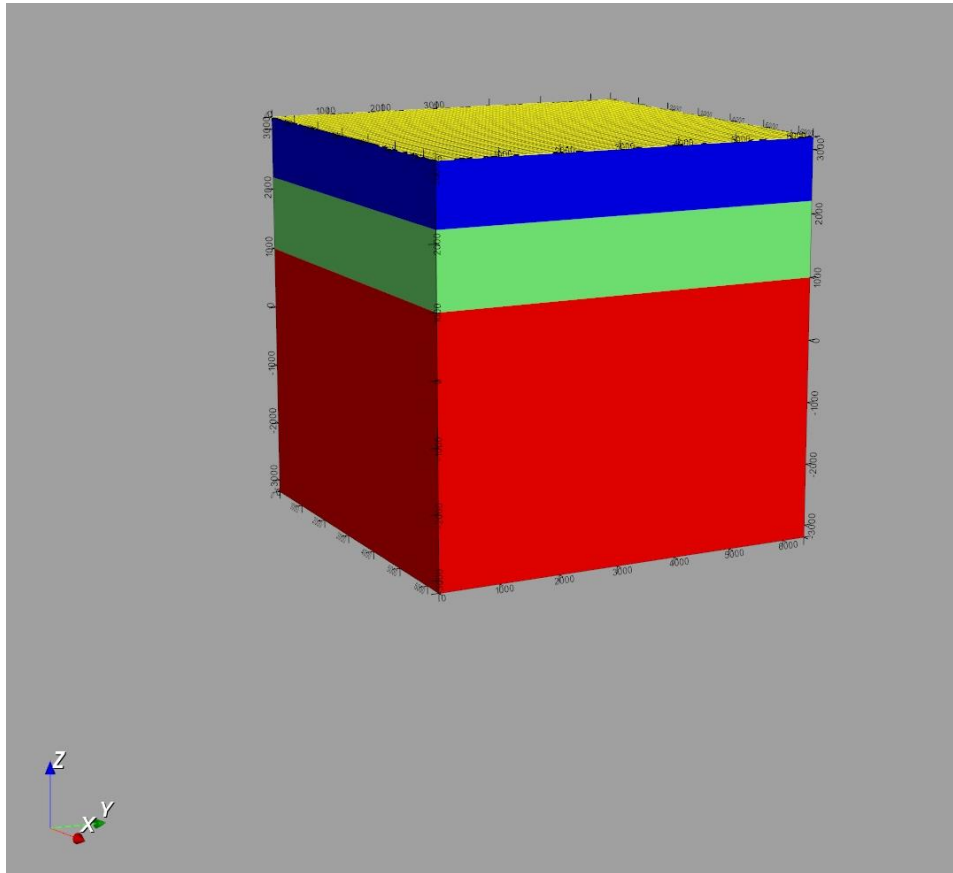


Figure 8. Cube Test Model. First benchmark model for testing the model orientation in Geomodeller software. Cell discretisation is 100m in every direction, also the model is cubic at 6.4 km in every direction. Top (blue) layer is set to density of pyroclastics at 2360 kg/m³, middle (green) layer represents granite at 2680 kg/m³ and at the bottom (red) is the andesite layer, at 2650 kg/m³.

Cube Test – forward modelling parameters

A master file of the voxel, which holds the spatial information of the model, along with the Z parameter stored in the separate binary file, is used to set up a 3D model in Geomodeller. The layer ID parameter, also stored in a binary file, is used to assign lithology, in this case for purely visual purpose. The capabilities of the Geomodeller that allow to input and edit lithology information is not used in these forward modelling exercises. The only necessary lithology information to successfully model the forward gravity response in Geomodeller is the density distribution within the model. We

provide this information by using density voxel, which has been created, together with all the other parameters of the model, in Basin3D and stored in separate binary file.

Other forward modelling settings of the Geomodeller have been set including extend of the padding when converting to the Fourier Domain, and the border effect. In Geomodeller, for each case of forward modelling the padding, simply called 'FFT', must be set to 100% (and not as it is set by default to 400%), to match the Basin3D settings.

The cube test model was run at -2550 meters (middle of the cell) in Basin3D, which is a depth within the top tuff layer at 2550 meters above sea level. Due to the different spatial orientation of the voxets used by Geomodeller, the voxel master file had to be edited so that it could be read by Geomodeller. This Cube Test has been designed to test the spatial orientation of the model in the Geomodeller's forward modelling module. The forward modelling has been run on two opposite depths, to compare the result to the Basin3D results for top tuff layer, at -2550 meters, and for bottom andesite layer at 2550 meters.

Cube Test – results

As mentioned above, forward modelling has been run on two opposite depths, to compare the result to the Basin3D results for top tuff layer, at -2550 meters, and for andesite layer at 2550 meters. Due to lack of topography in any of the layers, it was expected that the output model will have uniform absolute value for gravity throughout the surface (x- y- directions) of the measurement. The results for the measurement at 2550 m.a.s.l. in Basin3D (-z) was 285.69 mGal, and in Geomodeller (+z) was 287.84 mGal (Table 3). To make sure the orientation of the model is correct in Geomodeller, we have also run the measurement at 2550 m.b.s.l. in Basin3D (+z) and in Geomodeller (-z), with results of -279.50 mGal and -281.54 mGal respectively (Table 3). The absolute value of the gravity measured at 2550 m.a.s.l. differs between Basin3D and Geomodeller by ~2.15 mGal, and similarly at 2550 m.b.s.l. the difference is ~2.04 mGal (Table 3). As a result of the symmetry within the model in z-direction, the absolute values of the two measurements are similar, although as expected based on the layer densities, the measurement taken above sea level is slightly higher. Thus, we perform another test with structural model asymmetric in z-direction to further check the orientation of the model in Geomodeller, results are described below.

Absolute value comparison at different depths – Basin3D Model vs. Geomodeller Model								
Model	Depth [m]	FFT [%]	Gravity [mGal]		PCC	MAE [mGal]	RMSE [mGal]	RMSE to MAE Ratio
			Basin3D	Geomodeller				
Cube Test	2550	100	285.69	287.84	-0.662194	2.14811	2.14811	1
	-2550	100	-279.50	-281.54	0.001989	2.04169	2.04169	1
Long Cube Test	2550	100	578.86	465.71	-0.000362	113.15184	113.15184	1
	-2550	100	-125.28	-100.85	-0.639359	24.42823	24.42823	1

Table 3. Absolute value comparison between Basin3D Model and Geomodeller Model. FFT stands for Fast Fourier Transform expansion and is expressed in the percent of the model size below $z=0$. PCC is the Parson's Correlation Coefficient, MAE is the Mean Absolute Error [mGal] and RMSE stands for Root Mean Square Error [mGal]. RMSE/MAE ratio is a measure of error distribution, value of 1 (minima) implies a uniform error distribution.

Extended Cube Test – spatial model

The initial Cube Test model has been extended in z dimension, this allows more control over the model orientation in Geomodeller, and will further allow to compare the results between the Basin3D and Geomodeller. The dimensions of the new extended model in x- and y- direction are the same, both for the cells and for the overall model. In the z dimension the cell size has not changed, but the model has been elongated by doubling the number of cells below sea level. As in Cube Test the model has 3 layers with varying densities, these has not changed and the same densities are still corresponding to tuff, granite and andesite. The only difference in this new model is the extend of the final layer, andesite, starting as previously from 1000 meters above sea level, but this time extending to 6.4 km below sea level. Thus, the thickness of the layer 3 is now 7.4 km, and the overall thickness of the model is 9.6km (Figure 9).

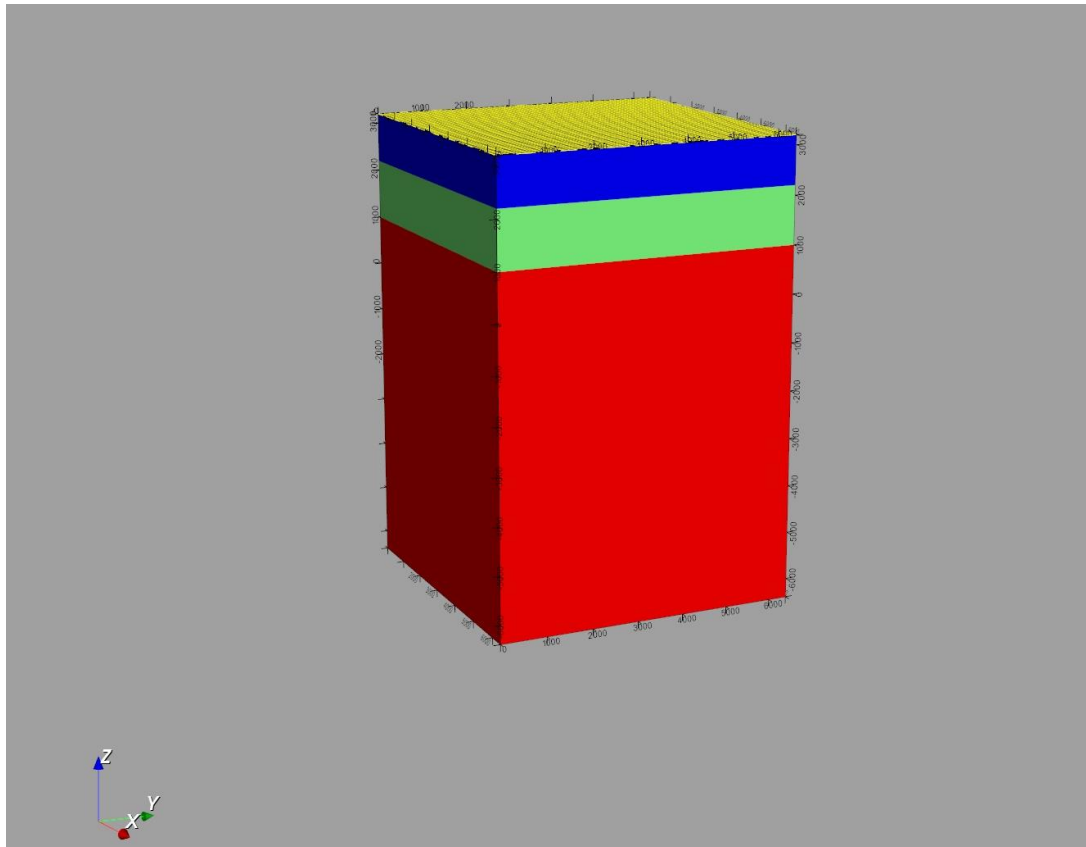


Figure 9. *Extended Cube Test Model. Second benchmark model for testing the model orientation in Geomodeller software. Cell discretisation is 100m in every direction, model is 6.4 km by 6.4 km in x- y-directions, and 9.6 km in depth (z-direction). Top (blue) layer is set to density of pyroclastics at 2360 kg/m³, middle (green) layer represents granite at 2680 kg/m³ and at the bottom (red) is the andesite layer, at 2650 kg/m³.*

Extended Cube Test – forward modelling parameters

As in the case of the Cube Test model mentioned above, the master file of the voxet, which holds the spatial information of the model, along with the Z parameter, are used to set up a 3D model in Geomodeller. The layer ID parameter is used to assign lithology for purely visual purposes, and the density distribution within the model is given by the density parameter. The density parameter is stored as a voxet, which has been created, together with all the other parameters of the model, in Basin3D and stored in separate binary file.

Forward modelling settings for the Extended Cube Test are, as in the Cube Test, set to match the Basin3D settings. The extend of the padding in z-direction when converting to the Fourier Domain is set to 100% (and not as it is set by default to 400% - as explained in *Geomodeller Forward Modelling*

section). The Extended Cube Test has been run at the same depths as Cube Test, at -2550 meters in Basin3D and at both -2550 meters and 2550 meters in Geomodeller.

Extended Cube Test – results

We extended the initial test model in the z-direction below sea level, to produce asymmetry within the model. The tops of the layers remained the same, with only the thickness of the bottom andesite layer increasing. The results for the measurement at 2550 m.a.s.l. in Basin3D (-z) was 578.86 mGal, and in Geomodeller (+z) was 465.71 mGal (Table 3). Furthermore, the measurement at 2550 m.b.s.l. in Basin3D (+z) and in Geomodeller (-z), yielded results of -125.28 mGal and -100.85 mGal respectively (Table 3). The absolute value of the gravity measured at 2550 m.a.s.l. differs between Basin3D and Geomodeller by ~113.15 mGal, but at 2550 m.b.s.l. the difference is only ~24.43 mGal (Table 3). The large difference in absolute values at 2550 m.a.s.l. is unexpected, as none of the modelling parameters were changed, only the extent of the model in z-direction below sea level has increased. Possible causes of this discrepancy will be discussed further below.

Los Humeros Model – spatial stratigraphic model

Los Humeros caldera is located within the Trans-Mexican Volcanic Belt (TMVB), and more specifically in the Serdán-Oriental Basin (SOB). The extent of the area of interest is mapped in Figure 5 (from Carrasco-Núñez et al., 2017), which has been used as an input for the stratigraphic model of Los Humeros built by Evanno (2017). Other inputs in the model include cross-sections from previously mentioned studies of Norini et al. (2015), and Carrasco-Núñez et al. (2017; Figure 10), as well as data from 16 wells supplied by the Comision Federal de Electricidad (CFE), and Digital Terrain Model (DTM). The structural model of Los Humeros has been created in Geomodeller (Geophysics Intrepid, 2015), using the potential-field methodology for interpolation from points to surfaces (Lajaunie et al., 1997). The output model extends from easting 640000 (19.47 latitude) to 696000 (19.8 latitude), and northing 2154000 (-97.133 longitude) to 2190000 (-97.65 longitude), exactly as the input map of Carrasco-Núñez et al. (2017; Figure 11). The overall size of the model is 56 km in x-direction (easting) by 36 km in y-direction (northing), and extends 5km above the sea level (elevation) and 7km below the sea level (depth). The output mesh grid resolution is 250m in every direction for 3D display and for the Tsurf files of the interfaces (Evanno, 2017). There are four formations in the model (Figure 11; cross-section Figure 12) that correspond to the cross section from Carrasco-Núñez et al. (2017; Figure 11), these are named Post Caldera Volcanics G1 (light yellow in Figure 10 ; yellow in Figure 11), Caldera Volcanics G2 (orange in Figure 10 ; magenta in Figure 11), Post Caldera Volcanics G3 (red in Figure 10 ; blue in Figure 11), and lastly Basement G4 (green in both figures).

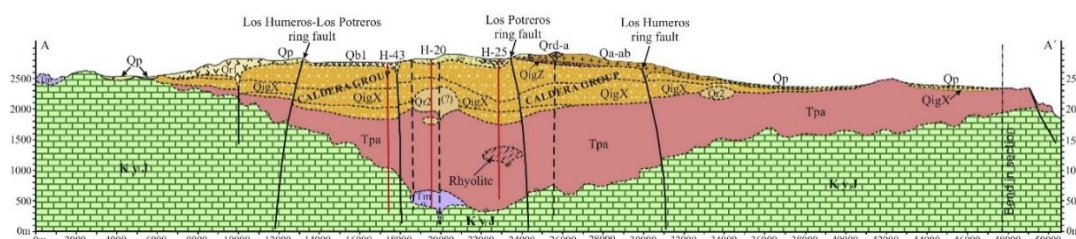


Figure 10. Cross-section of the Los Humeros caldera from West to East, also marked in Figure 11. Top (yellow) layer is pyroclastics, second (orange) layer is ignimbrites, the third (red) layer is andesite, and the basement is limestone (green). From: Carrasco-Núñez et al., 2017.

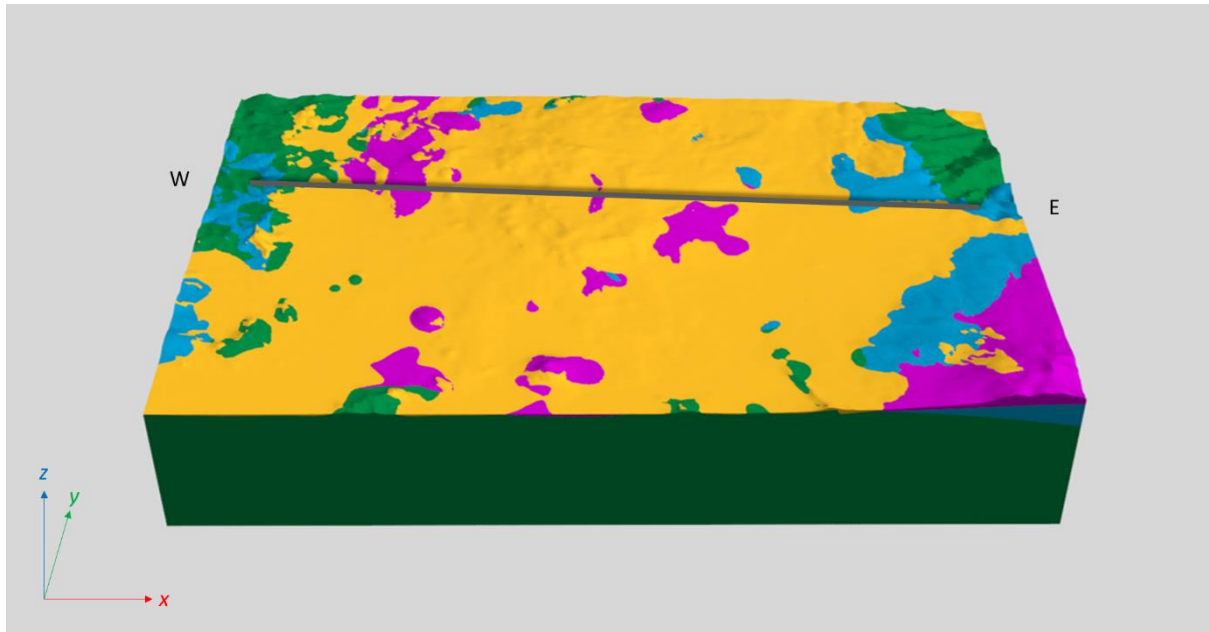


Figure 11. 3D structural model by Evanno, 2017. Grey line marks the cross-section from West to East illustrated in Figure 10 and Figure 12. Top (yellow) layer corresponds to pyroclastics, second (magenta) layer is an ignimbrite deposit, under lies the andesite layer (blue) and at the bottom is the limestone layer (green).

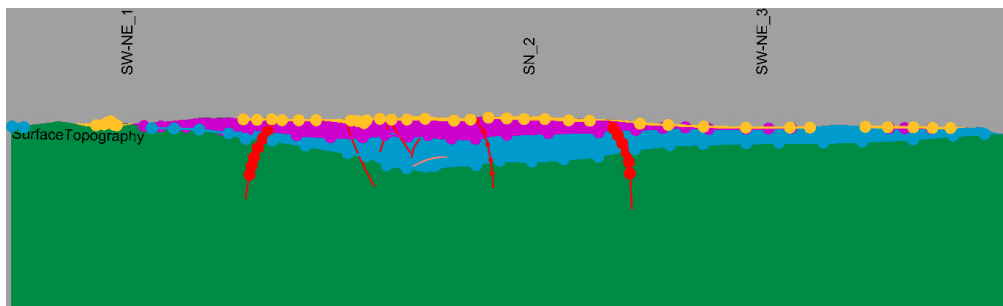


Figure 12. Cross-section of the structural model by Evanno, 2017, corresponding to the cross-section by Carrasco-Núñez et al. (2017) (Figure 10). Top (yellow) layer corresponds to pyroclastics, second (magenta) layer is an ignimbrite deposit, under lies the andesite layer (blue) and at the bottom is the limestone layer (green).

The model by Evanno (2017) has been used as a basis for the structural model of Los Humeros used in this study, the Tsurf files of the interfaces have been exported from Geomodeller, and imported into Basin3D. The extend of the Evanno model (2017) has also been preserved, but the discretisation in x-, y- and z-direction has been changed to suit the methodology for forward and inverse modelling in both Basin3D and Geomodeller. Number of cells in both x- and y-directions is set to 128, this results in the cells size of 0.4375 km in x-direction (West-East) and cell size of 0.28125 km in y-direction (South-North). Discretisation of the model in the z direction is the most important for the conversion

to Fourier Domain, thus we have set the cell size to 0.12 km, so that the number of cells below the $z=0$ is 64 cells and there are 40 cells above $z=0$.

Our Los Humeros model has 4 stratigraphic layers followed from the model of Evanno (2017), with an additional 'air' layer above the topography set to the Bouguer correction density of 2670 kg/m³ and a bottom crustal layer set to 2680 kg/m³ following Verma et al. 2013, who specified it as intrusive rock (granite) section 1 (Table 1). The Post Caldera Volcanics G1 layer (top tuffs) has been set to density of 2360 kg/m³ as Verma's acid volcanic rocks section 4 (Table 1), and Caldera Volcanics G2 (ignimbrites) have been assigned the density of Verma's intermediate volcanic rocks section 3 (Table 1) at 2394 kg/m³. The Post Caldera Volcanics G3 (andesite) are set to the average andesite value of 2650 kg/m³ from Hantschel and Kauerauf (2009), and lastly Basement G4 is a limestone basement with density of 2180 kg/m³ Verma's limestone section 2 (Table 1). We illustrate the completed structural model in Figure 13 as a top view above the caldera at 3312.5 m.a.s.l. with marked caldera in dashed line, and two cross-sections. The layer values from one through to six, correspond to layers starting from the top layer of 'air' (1), then top tuffs (2), ignimbrites (3), andesite (4), and down to limestone (5) and the bottom layer of crustal rock (6). South to North cross-section of the model (Figure 14) shows the extend of each layer below the caldera, accurately presenting the limestone layer (5), which at times reaches over 9km in thickness. Furthermore, the andesite layer (4) is visibly mostly located directly under the caldera, only with small extension towards the East visible in the West-East cross-section (Figure 15).

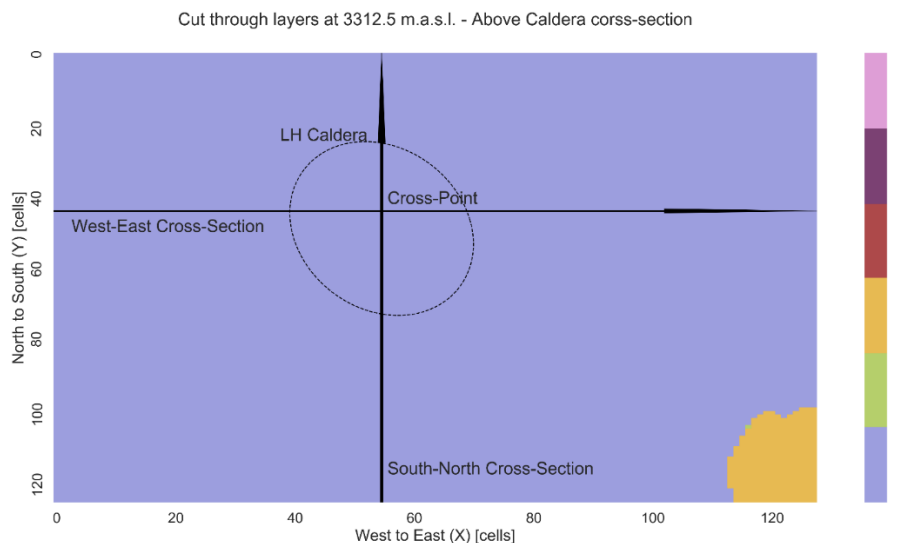


Figure 13. Structural model of Los Humeros created in Basin3D Preprocess module. Layer 1 (violet) represents air layer above topography, Layer 2 (green) is the top tuffs also mentioned as pyroclastics, Layer 3 (yellow) is the ignimbrites, Layer 4 (dark red) is the Andesite, Layer 5 (purple) is the limestone basement and Layer 6 (pink) is the crustal layer.

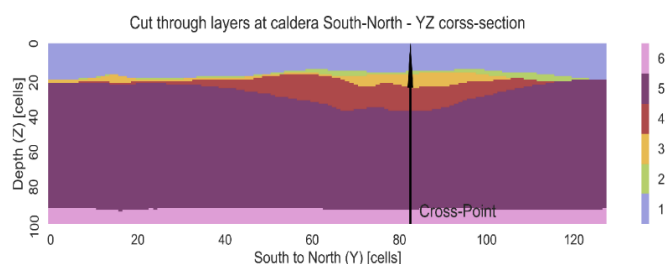


Figure 14. South to North cross-section of the Los Humeros structural model. Layer 1 (violet) represents air layer above topography, Layer 2 (green) is the top tuffs also mentioned as pyroclastics, Layer 3 (yellow) is the ignimbrites, Layer 4 (dark red) is the Andesite, Layer 5 (purple) is the limestone basement and Layer 6 (pink) is the crustal layer.

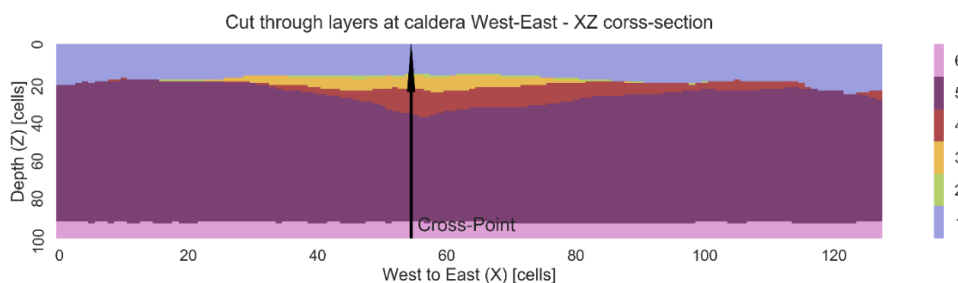


Figure 15. West to East cross-section of the structural model of Los Humeros. Layer 1 (violet) represents air layer above topography, Layer 2 (green) is the top tuffs also mentioned as pyroclastics, Layer 3 (yellow) is the ignimbrites, Layer 4 (dark red) is the Andesite, Layer 5 (purple) is the limestone basement and Layer 6 (pink) is the crustal layer.

Los Humeros Model – forward modelling parameters

Los Humeros model has been run at -3312.5 meters in Basin3D, and at corresponding 3312.5 meters in Geomodeller.

The master file of the voxel, which holds the spatial information of the model, along with the Z parameter stored in the separate binary file, are used to set up a 3D model in Geomodeller. The layer ID parameter is used to assign lithology for purely visual purposes, and the density distribution within the model is given by the density parameter. The density parameter is stored as a voxel, which has been created, together with all the other parameters of the model, in Basin3D and stored in separate binary file.

Other forward modelling settings of the Geomodeller have been set including extend of the padding when converting to the Fourier Domain, and the border effect. In Geomodeller, for each case of forward modelling the padding, simply called 'FFT', must be set to 100% (and not as it is set by default to 400%), to match the Basin3D settings.

Los Humeros Model – results

Results

Structural model of Los Humeros has been used for forward modelling of gravity anomaly in Basin3D and Geomodeller. The topography of the area has been mapped in Figure 16, where a rough outline of the Los Humeros caldera is marked in a dashed line. The highest peak of the caldera is at 3130.3 meters above sea level, thus we compute all of the forward and inverse models in the cell just above the caldera, at elevation of 3312.5 meters. The results for the measurement are presented in Figure 17 and Figure 18, where the same dashed line is marked out representing the Los Humeros caldera. The average value of the modelled absolute gravity has been subtracted from the respective models, thus data has been shifted to around zero, and the two maps represent normalised gravity anomaly in unit of mGal. Results of the Basin3D forward model range between 510.02 mGal and 536.28 mGal in absolute values, with average of 519.40 mGal. After the shift the anomaly ranges between -9.38 and 16.88 mGal. The Geomodeller forward model has lower absolute values between 430.53 and 458.017 mGal, and average of 440.30 mGal. However, after the subtraction of the mean, the gravity anomaly values are in the range of -9.78 and 17.71 mGal. Thus, the forward modelled gravity anomaly

has a comparable range in both Basin3D and Geomodeller. Some of the plots and data used in this analysis, not included here, can be found in *Appendix I*.

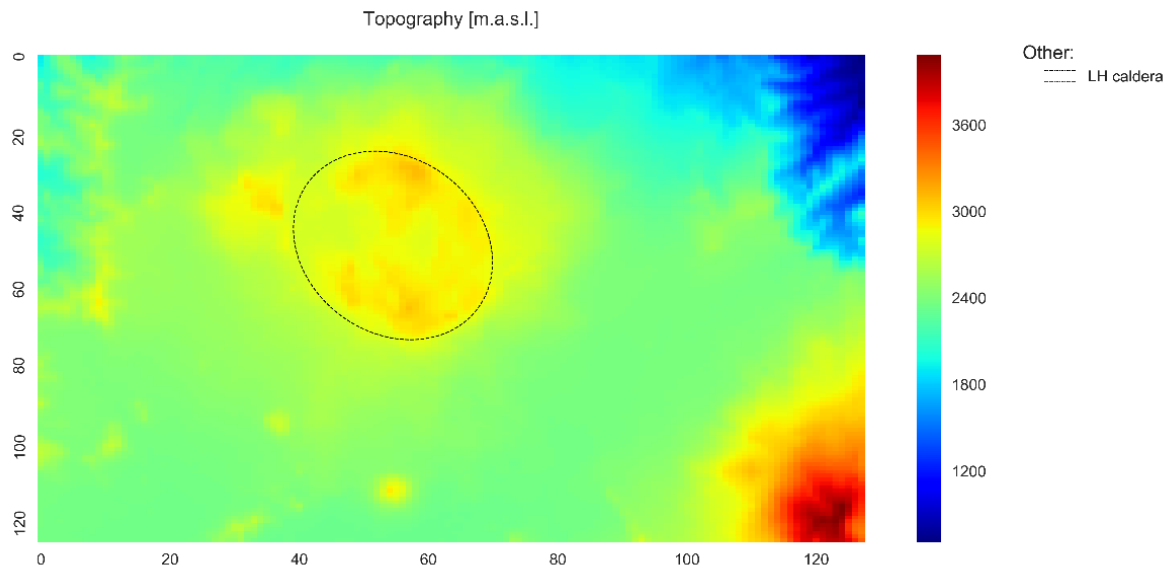


Figure 16. Topography grid of the Los Humeros study area. Extracted from the structural model of Evanno (2017). Discretisation is 0.5 km by 0.5 km (x- y- directions). Dashed line marks the maximum extend of the Los Humeros caldera, which reaches the elevation of 3130.3 m.a.s.l.

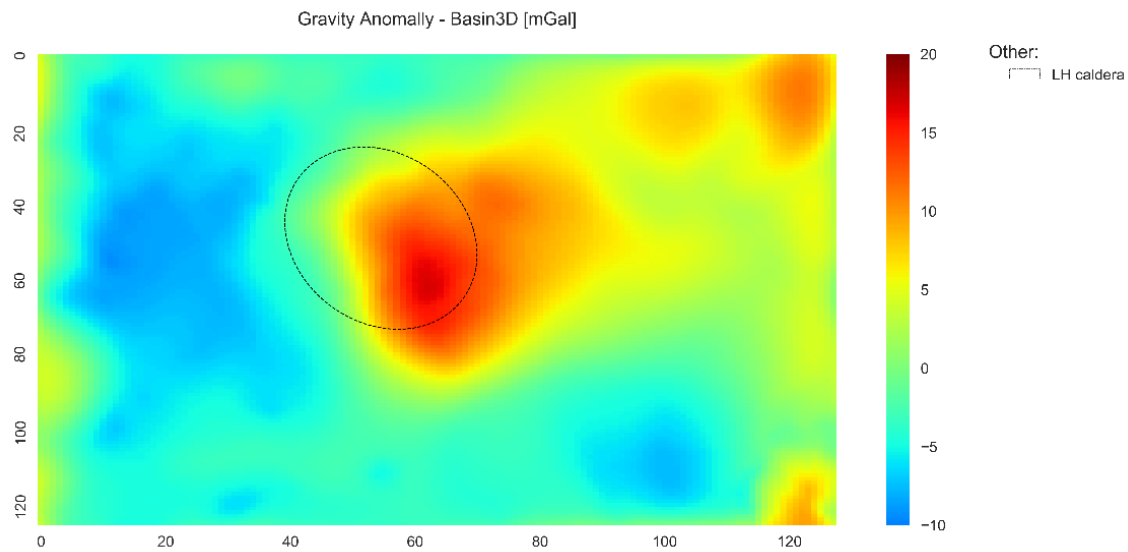


Figure 17. Forward modelled gravity anomaly – result from Basin3D. Modelling based on the structural model by (Evanno, 2017). Gravity values range from -9.38 mGal to 16.88 mGal.

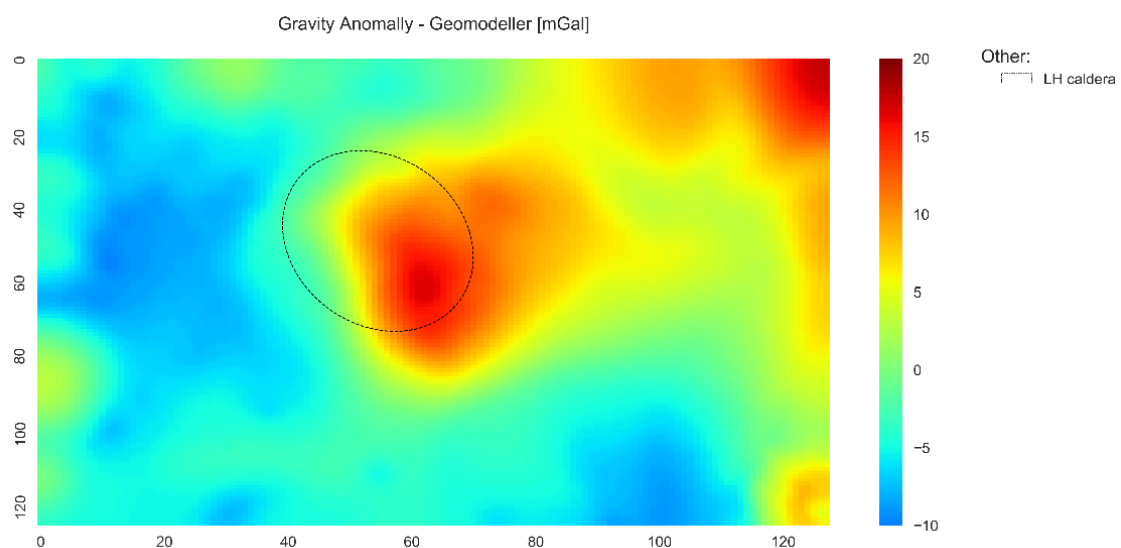


Figure 18. Forward modelled gravity anomaly – result from Geomodeller. Modelling based on the structural model by (Evanno, 2017). Gravity values range from -9.78 mGal to 17.71 mGal.

Distribution of the high and low gravity anomalies for both models is visualised in Figure 17 and Figure 18, illustrating a great agreement between the models. Basin3D model and Geomodeller model, both show a high gravity anomaly in the south-east part of the Los Humeros caldera. Based on the image alone, this high gravity anomaly looks virtually the same in both models. Furthermore, both models display an extensive low gravity anomaly in the western part, and a smaller one in the south-east of the model. Although, there are visible differences between the two models, particularly the low

gravity in the Geomodeller extends further towards the edges of the model and has a slightly higher magnitude (Figure 18). Moreover, the high gravity anomaly in the north-east section of the models is much larger and higher in magnitude in the Geomodeller model, than in the Basin3D model. In Figure 19, the histograms in the second row (Basin3D – left, Geomodeller – right), show the distribution of gravity values in the models. Both models have the same bimodal distribution, with one higher peak at around -4 mGal and a smaller one at 5 mGal. The increase number of values below -5 mGal and above 15 mGal, in comparison to the Basin3D model, is also visible in the histogram of Geomodeller model. The visualised difference between the Basin3D forward gravity model and the Geomodeller forward gravity model, highlights the edges of the models as the main areas of dissimilarity (third row, left column, Figure 19). Over half of the cells (over 8000 cells) in the models are the same with a 0 mGal error in the difference histogram (third row, right column, Figure 19), and most of the modelled cells stay within an error of 2 mGal, with only few cells reaching a difference as high as 6 mGal. This distribution of the error can be quantified by computation of Mean Absolute Error (MAE), Root Mean Square Error (RMSE) and the ratio between RMSE and MAE. We use MAE to describe the difference (error) between the two models, and the ratio of RMSE to MAE describe the error distribution. In this case MAE is very low at 0.741 mGal, confirming the visibly low error in the difference histogram. Moreover, the RMSE/MAE ratio is very large at 2.072, indicating the error between the two data sets is badly distributed, suggesting there are few values that are much higher or much lower than the rest. These higher errors are visible as outliers in the scatterplot of the Basin3D result versus Geomodeller result (bottom row, Figure 19). From the difference map we have identified, that the cells contributing to these errors (outliers) are at the edges of the model. Thus, we identify them as edge errors and not errors in the canter of the models. Regardless of this significant difference at the edges, the Pearson's Correlation Coefficient (PCC) is relatively high at 0.97227, again confirming that the rest of the models' cells match well. Below we investigate the effect of the edges on the correlation between Basin3D model and the Geomodeller model. Some of the plots and data used in this analysis, not included here, can be found in *Appendix I*.

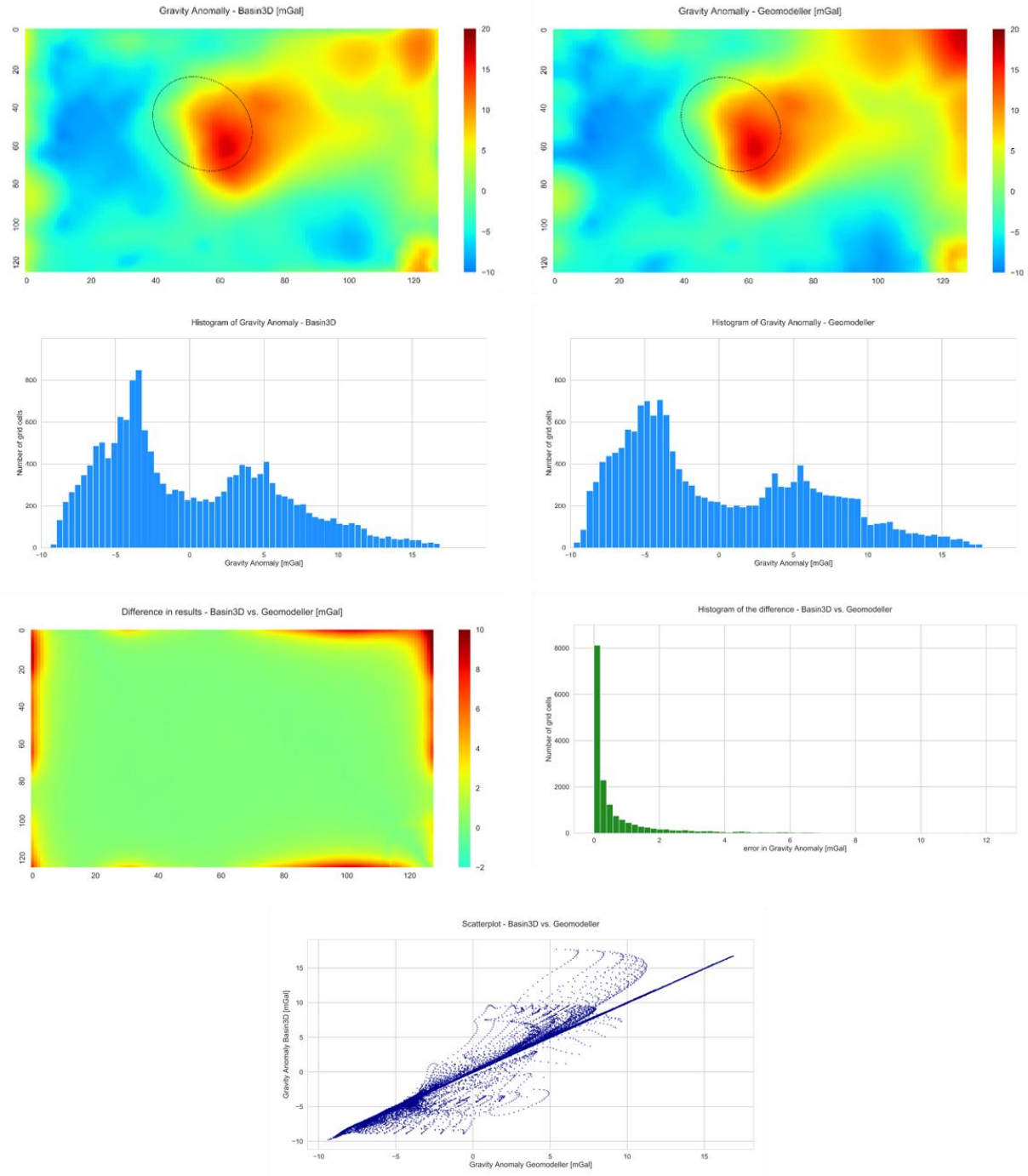


Figure 19. Comparison of forward modelling in Basin3D vs. Geomodeler. Top row: gravity anomaly modelled in Basin3D (left) and gravity anomaly modelled in Geomodeler (right). Second row: histogram of the gravity anomaly values modelled in Basin3D (left) and Geomodeler (right) – bimodal with peaks at -4 mGal and 5 mGal. Third row: difference between Basin3D result and Geomodeler result (left) and the histogram of the difference (absolute) values (right). Bottom row: gravity anomaly values of Basin3D vs. Geomodeler – scatterplot.

Edge thickness check

The benchmarking of Basin3D forward model versus Geomodeler forward model yielded good preliminary results. However, in the previous section we explained that all of the highest errors are located on the edges of the models. As a result of these errors, both imaging and numerical comparisons, prevent us from examining the differences between the models at their centre, where the caldera of Los Humeros is located. Thus, we removed the edges of the forward models, both Basin3D and Geomodeler, and compared the results. We started with 5 cells and then increase up to 40 cells with increments of five. It is important to mention that the discretisation in x-direction is different to the discretisation in y- direction, thus each removed cell is a distance of 0.4375 km in the x-direction and 0.28125 km in the y-direction. We visualised the degree of the cell cut-off in the Figure 20, which illustrates the eight increasingly larger edges removed from the models and a Los Humeros caldera marked with a dashed line. The depth of the models below the measurement is 11.3125 km. To accurately calculate the gravity anomaly caused by a buried object at a certain depth, there must be an edge buffer around that object, equivalent in distance to the burial depth. Thus, regardless of reaching the caldera outline at 25 cells cut-off, we went further to achieve the edge distance equivalent to the depth below measurement in the y-direction. We reached the depth equivalent edge thickness in the x-direction at 26 cells, but in y-direction it is only reached at 40 cells. Some of the plots and data used in this analysis, not included here, can be found in *Appendix I*.

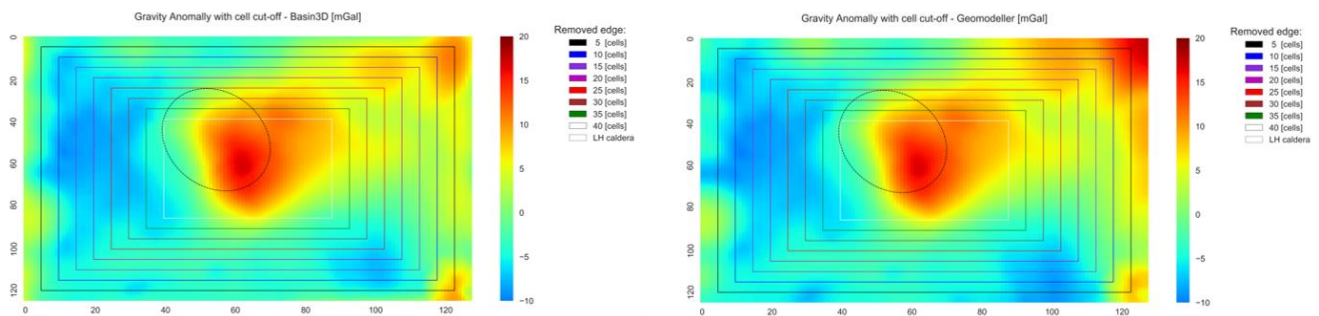


Figure 20. Visualised degree of edge cut-off in the analysis of the edge error contribution. Black – 5 cells, Blue – 10 cells, Purple – 15 cells, Magenta – 20 cells, Red – 25 cells, Burgundy – 30 cells, Green – 35 cells, White – 40 cells. Dashed line marks out the outline of Los Humeros caldera.

In Figure 21 we plot the changes to the correlation, MAE and error distribution based on the depth equivalent edge cut-off, for x-direction in left column and y-direction in the right column. The top row (Figure 21) shows a decrease of MAE with respect to increasing edge cut-off. The error falls below 0.1 mGal at 70% and 45% of the depth equivalent distance, for the x-direction and y-direction respectively. This represents a seven-fold decrease in the mean absolute error, indicating that the further away from the edges, the better the match between the Basin3D model and Geomodeller model. Similarly, correlation (PCC) of the two models increases substantially for the first 15 cells cut-off (second row, Figure 21), but starts to level off at 20 cells, equivalent to 60% in x-direction and 40% in y-direction, reaching 0.99959. The error distribution also improves, although not as drastically as the other parameters (bottom row, Figure 21), reaching only 1.3 after 40 cell cut-off. The majority of the edge error effects remains within 15 cells from the edge, equivalent to 58% (x-direction) and 37% (y-direction) of the model depth, as seen by a drop in the maximum difference between the models to 1mGal (Figure 22). In this test we have shown that there is a strong edge effect when comparing the Basin3D forward model to the Geomodeller forward model, we will discuss the possible source of this difference in the *Benchmarking Discussion* section. Some of the plots and data used in this analysis, not included here, can be found in *Appendix I*.

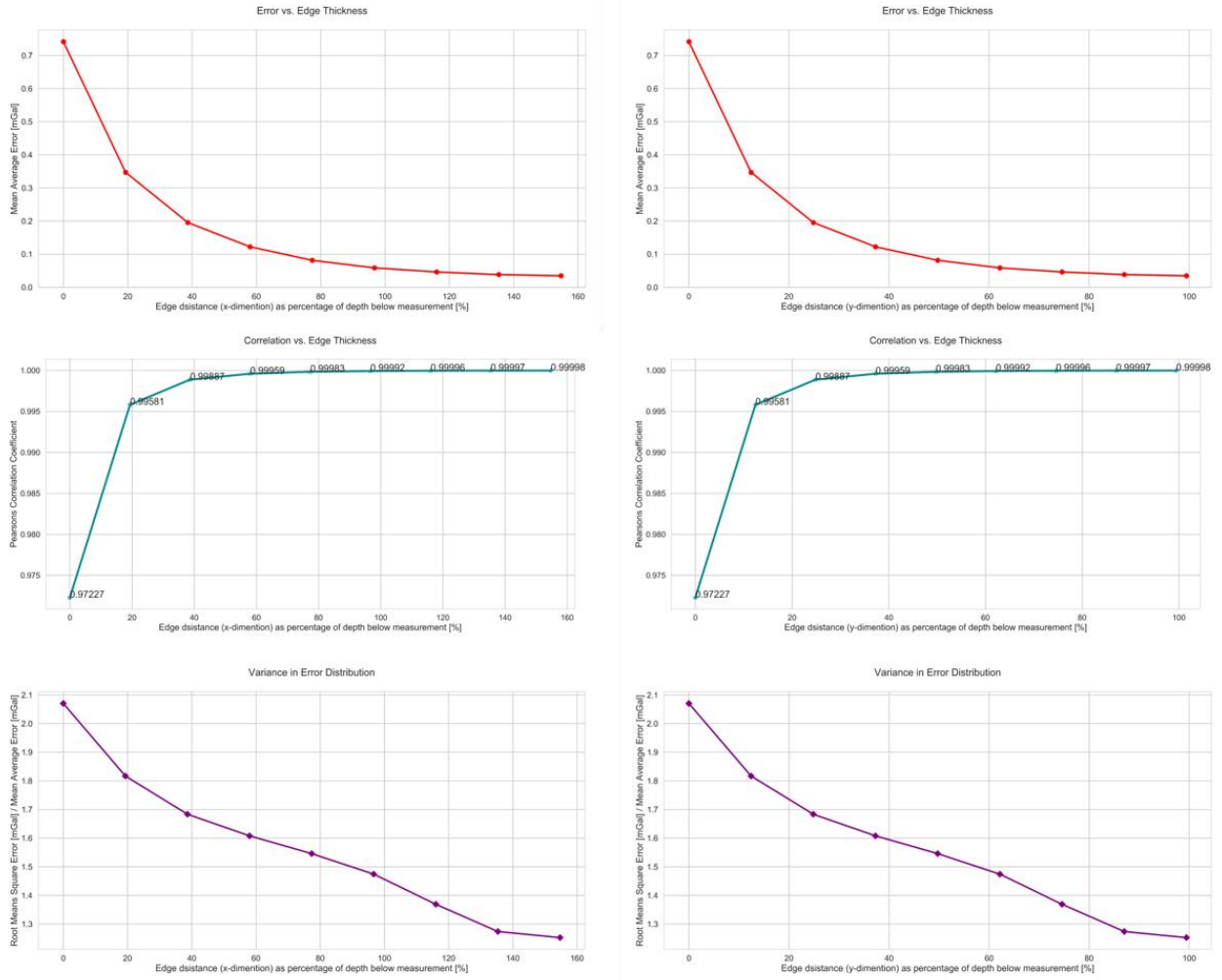


Figure 21. Analysis versus edge cut-off equivalent to model depth, expressed in the percent of the model depth. Top row: Mean Absolute Error (MAE) vs. depth equivalent edge cut-off for x-dimension (left) and y-dimension (right). Second row: Pearson's Correlation Coefficient (PCC) vs. depth equivalent edge cut-off for x-dimension (left) and y-dimension (right). Bottom row: Variance in error distribution vs. depth equivalent edge cut-off for x-dimension (left) and y-dimension (right).

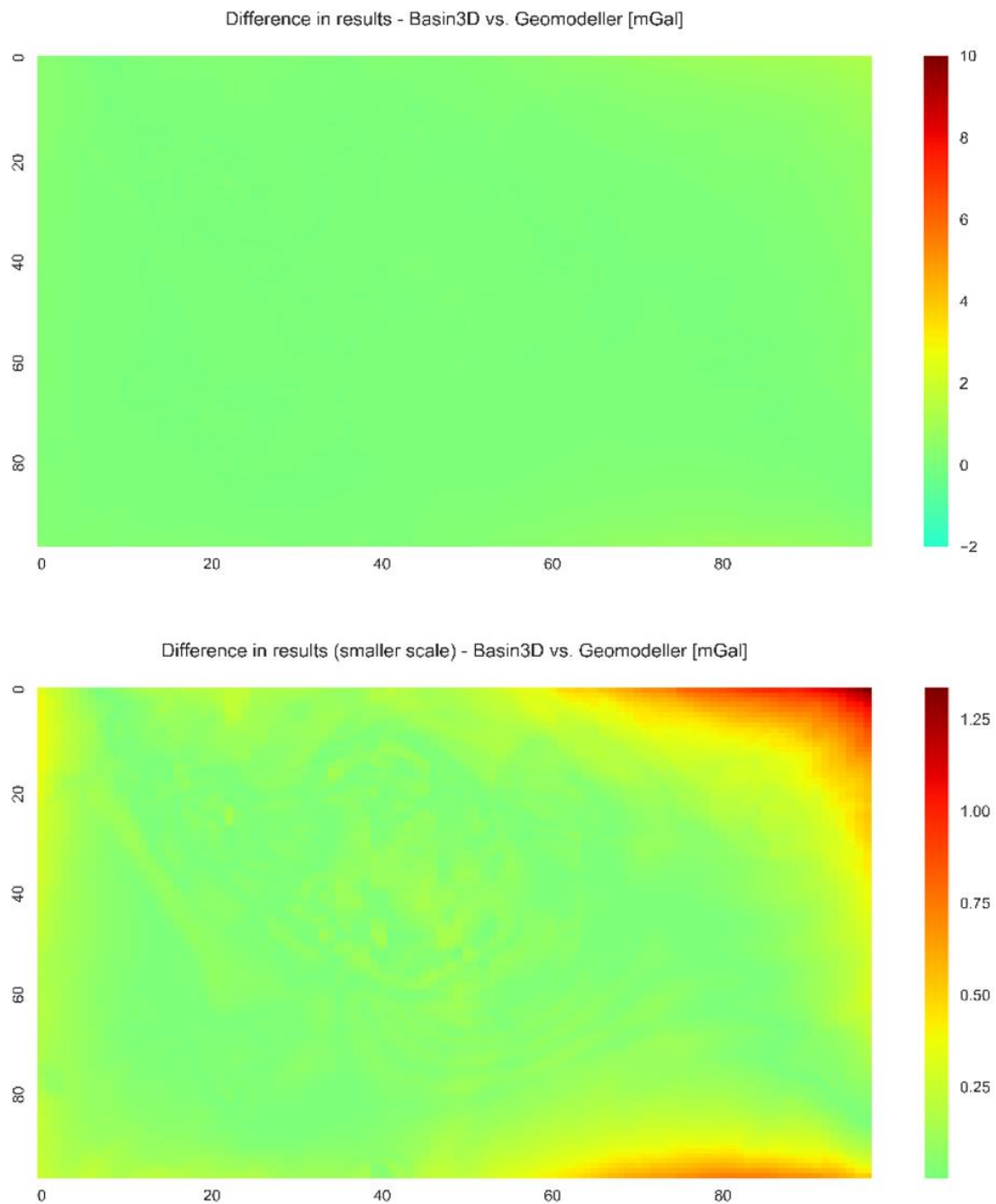


Figure 22. Difference plot for 15 cell cut-off: equivalent to 58% (x-direction) and 37% (y-direction) of the model depth. Error between the two models drops to around 1mGal.

FFT check

Absolute gravity values modelled in Basin3D differ from those modelled in Geomodeller, despite the same Fast Fourier Transform (FFT) setting. We check the functionality of the Geomodeller FFT setting by comparing different expansions to expansion at 100% of the model. In Table 4 the PCC virtually

does not change with each next expansion, as the distribution of the high and low gravity anomalies remains the same, despite the changing absolute gravity values. However, the amplitude of the gravity anomalies changes slightly as revealed by the increasing MAE and worsening error distribution. Thus, although the absolute values of the gravity anomaly change when a different FFT expansion is applied, the distribution of the gravity anomalies and their amplitudes do not change considerably. Some of the plots and data used in this analysis, not included here, can be found in *Appendix I*.

Comparison of Geomodeller Models at different FFT expansions vs. at FFT 100%						
Model	Depth [m]	FFT [%]	PCC	MAE [mGal]	RMSE [mGal]	RMSE to MAE Ratio
LH-G-fft125	3312.5	125	0.999991	0.04016	0.04664	1.16120
LH-G-fft150	3312.5	150	0.999977	0.06718	0.07824	1.16465
LH-G-fft200	3312.5	200	0.999950	0.09540	0.11167	1.17057
LH-G-fft400	3312.5	400	0.999893	0.13417	0.15874	1.18312
LH-G-fft600	3312.5	600	0.999881	0.14040	0.16647	1.18572
LH-G-fft1000	3312.5	1000	0.999877	0.14203	0.16861	1.18718

Table 4. Different expansions of Geomodeller model into Fourier domain versus an expansion at 100% and equivalent of Basin3D setting. FFT stands for Fast Fourier Transform expansion and is expressed in the percent of the model size below $z=0$. PCC is the Parson's Correlation Coefficient, MAE is the Mean Absolute Error [mGal] and RMSE stands for Root Mean Square Error [mGal]. RMSE/MAE ratio is a measure of error distribution, value of 1 (minima) implies a uniform error distribution.

Benchmarking Discussion

Basin3D software performs very well in the benchmarking test. The correlation with the commercially available Geomodeller software is very high, reaching 0.99959, when the edge effects are removed. Furthermore, the error between the two software is minimal at 0.1 mGal with removed edge effects, confirming that not only the gravity anomaly distribution is the same, but also their amplitude. This is a great result, verifying that Basin3D is a reliable software for forward modelling of the gravity anomaly. The small difference on the edges of the models can be explained by the fact that the Basin3D model is not padded. The option to pad the model in Basin3D is available, but uses nearest neighbour extrapolation to pad the model in x- y- space. If this padding method was used in the benchmarking test, it would not allow a direct comparison with Geomodeller, thus it is not used in this study. Geomodeller automatically pads the model by mirroring it on the each side, as explained

in the *Geomodeller Forward Modelling* section. This padded versus unpadded forward modelling is easier to analyse, than the results of two different padding methodologies. In this case the unpadded versus padded modelling causes the discrepancies between the Geomodeller and Basin3D models, lowering the correlation to 0.97227 and increasing the error to 0.741 mGal. The padding methodology used in Geomodeller is an interesting solution to the lack of density information on the boundaries of the model. Unfortunately, it only works with symmetrical models (preferably rectangular in x- y- directions) and could not be applied in Basin3D, which allows an input of irregularly shaped model. The structural models in Basin3D are allowed to be irregularly shaped by using a surface file that specifies extend of the model (x- y- directions) in terms of individual cells within the model space. Thus, in this case nearest neighbour extrapolation is a better methodology.

Inconsistent absolute gravity value due to the changes of the expansion before transformation into the Fourier domain (Geomodeller), does not affect the final results, as the modelled gravity is shifted to around zero. Therefore, the comparison between the gravity anomalies modelled in Basin3D to that modelled in Geomodeller is not affected.

The structural model used for Basin3D forward modelling and Geomodeller forward modelling was identical, including the discretisation, layer topography and densities. Furthermore, the computation of both software's is based on the same methodology following the algorithm of Cooley and Tukey (1965), and all of the calculation parameters have been set to the same values. Thus, we are confident that the benchmarking of Basin3D versus Geomodeller was performed accurately. Hence, we consider the small number of case studies, as the only limitation of the benchmarking process.

Forward modelling of gravity anomaly in Basin3D has been benchmarked versus forward modelling module of Geomodeller, with great results and Basin3D is proved to be a reliable tool.

LOS HUMEROS - INVERSE MODELLING

Observed Gravity Data

At the moment of this study there are no data available on the observed gravity anomaly in Los Humeros area. We have however access to a publication from 1992 by García-Estrada, which includes a Bouguer anomaly map available as an image within the article (PDF file) (Figure 23). The source of this map and the measurements that were its basis, are not described by the author. Furthermore, there is no mention of any other metadata, including the coordinate system, units, interpolation methods or measurement errors. The units are only described as gravimetric units, thus we assume, it refers to non SI unit equal to 0.1 mGal, which seems to correspond well with the anomaly values within the map.

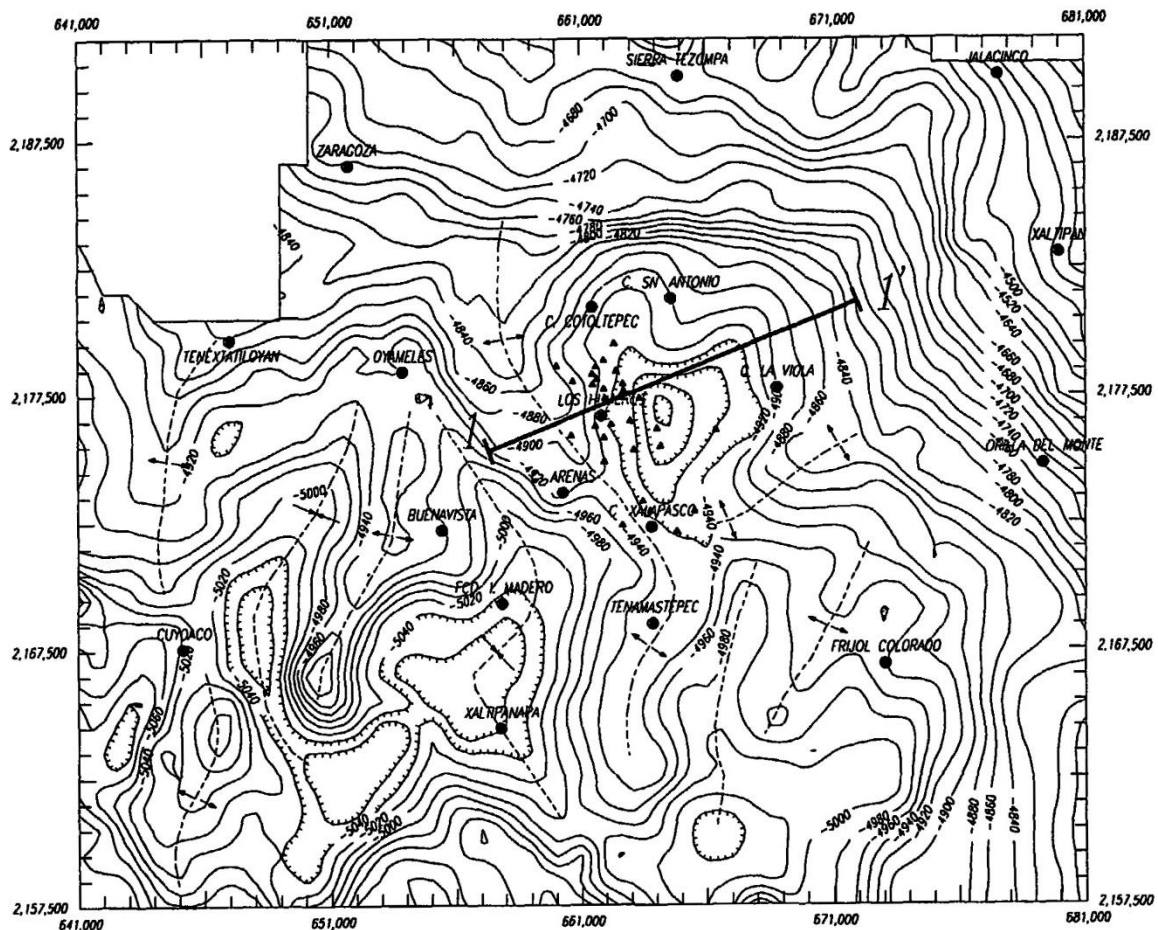


Figure 23. Measured gravity map by García-Estrada (1992), units are standard gravimetric units i.e. 0.1 mGal. The projection of the map is assumed to WGS84/UTM14N.

Although there are many uncertainties in accuracy of this map, it is the only available gravity anomaly map for the study area. Thus, we use it for our data assimilation in the inverse modelling of the gravity anomaly of Los Humeros. First, we digitise the map using open source QGIS software, converting the map into polylines, setting the coordinate reference system to WGS84/UTM14N and then interpolating by triangular interpolation (Triangulated Irregular Networks (TIN)) to a raster. Resulting map (Figure 24) has a very fine grid of cells, 300 in x direction and 250 in y directions, and cell size of 130 m by 130 m. We use a topography shape file from the Los Humeros structural model to compare the extend of the digitised gravity map to the topography (Figure 25). Los Humeros caldera is marked with a dashed line in both maps, as well as the outline of the digitised gravity anomaly map (Figure 24 and Figure 25, and Figure 26). The available gravity anomaly data extends only over a part of the structural model of the Los Humeros area, thus, limiting the data assimilation effect to this area only. Furthermore, the topography of the area within the black line is varied and covered in thick vegetation. Considering all these aspects of the area and its size at 39.08 km by 32.57 km, we decide that the 97 data points for observed gravity anomaly (approximately 3 km by 3km cell size) is a reasonable number. The newly rasterised map is shown in Figure 27, with the Los Humeros caldera marked in a dashed line.

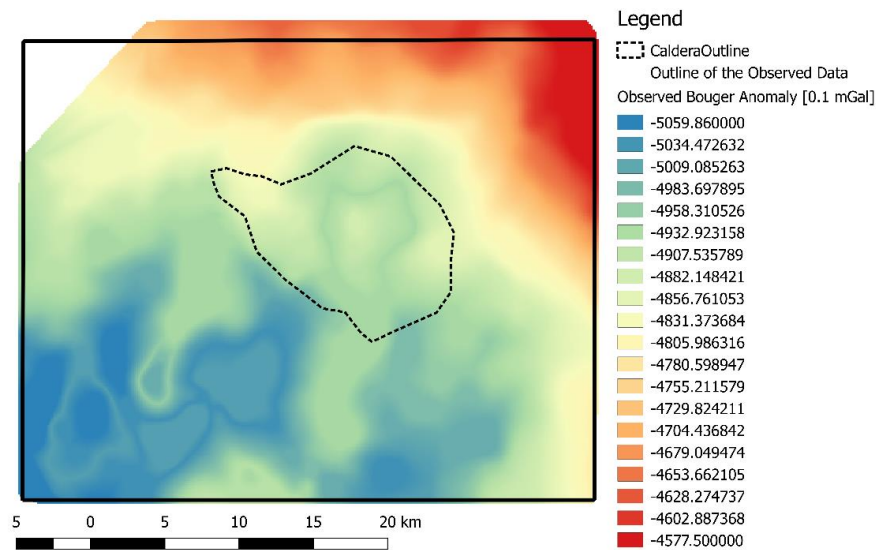


Figure 24. Digitised measured gravity map with original units of 0.1 mGal by García-Estrada (1992). Dashed line marks the outline of the Los Humeros caldera. Black solid line marks the outline of the map that fits into the model space of the structural model by Evanno (2017).

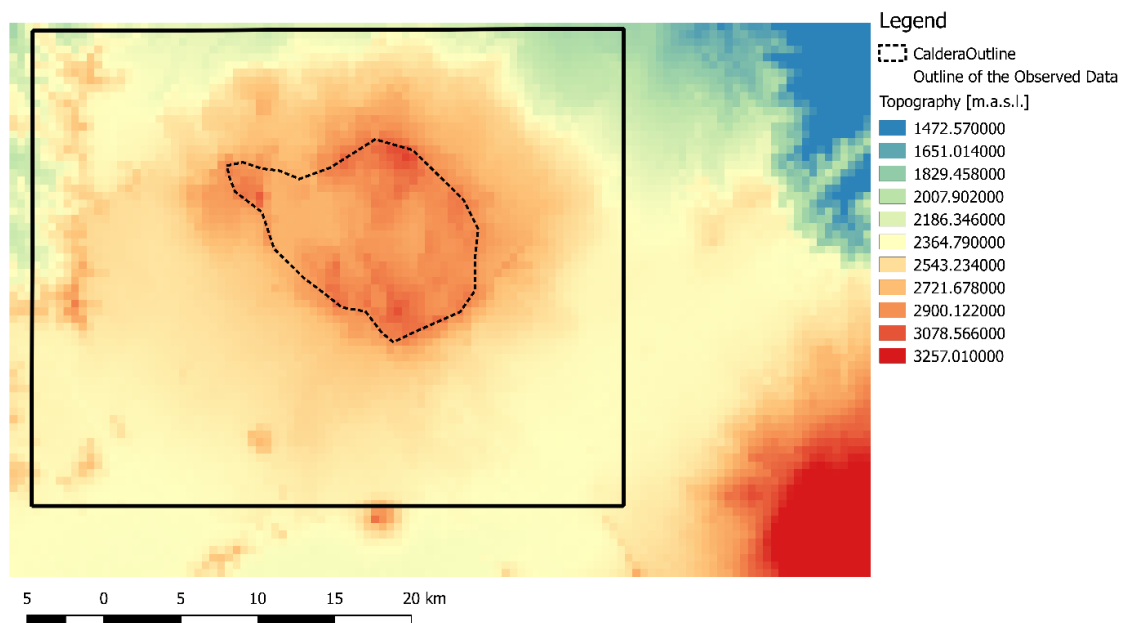


Figure 25. Topography grid of the Los Humeros study area visualised in the QGIS software. Extracted from the structural model of Evanno (2017). Discretisation is 0.5 km by 0.5 km (x- y- directions). Dashed line marks the maximum extend of the Los Humeros caldera, which reaches the elevation of 3130.3 m.a.s.l. Black solid line marks the outline of the observed gravity map that fits into the model space of the structural model by Evanno (2017).

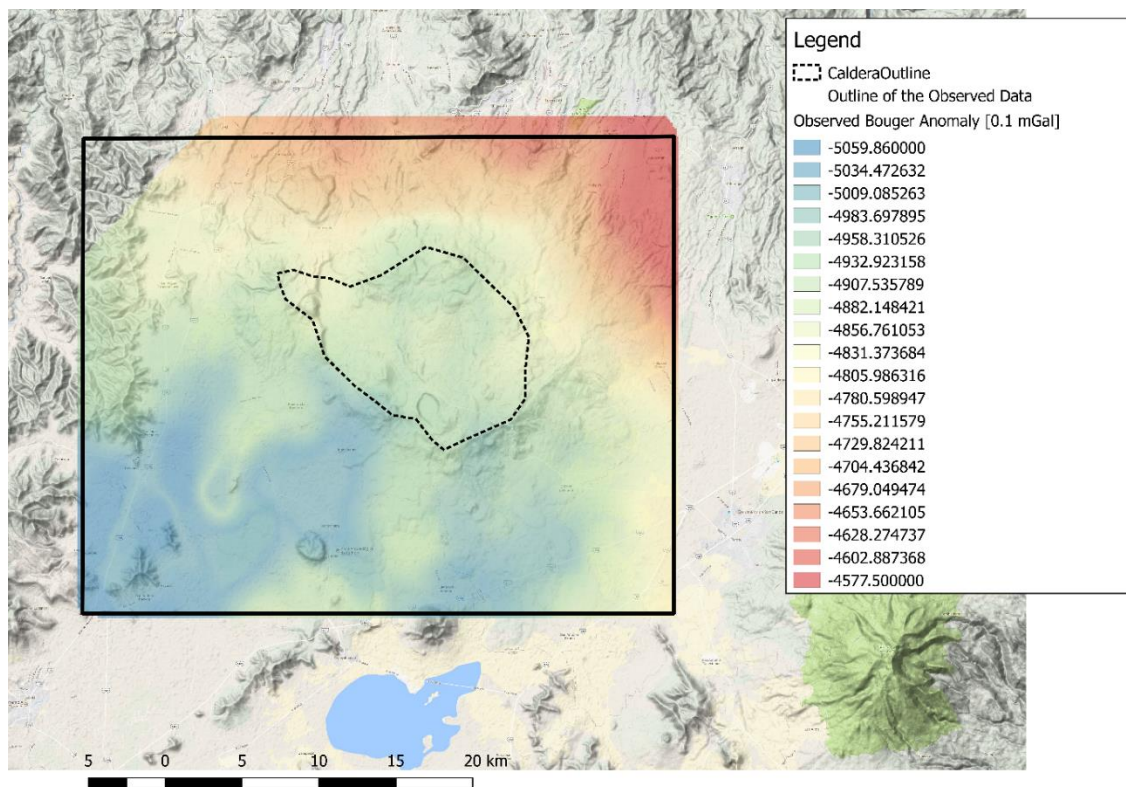


Figure 26. Digitised measured gravity map with original units of 0.1 mGal by García-Estrada (1992), overlaid of the terrain view of the Los Humeros study area (from Google Maps, 2018). Dashed line marks the outline of the Los Humeros caldera. Black solid line marks the outline of the map that fits into the model space of the structural model by Evanno (2017).

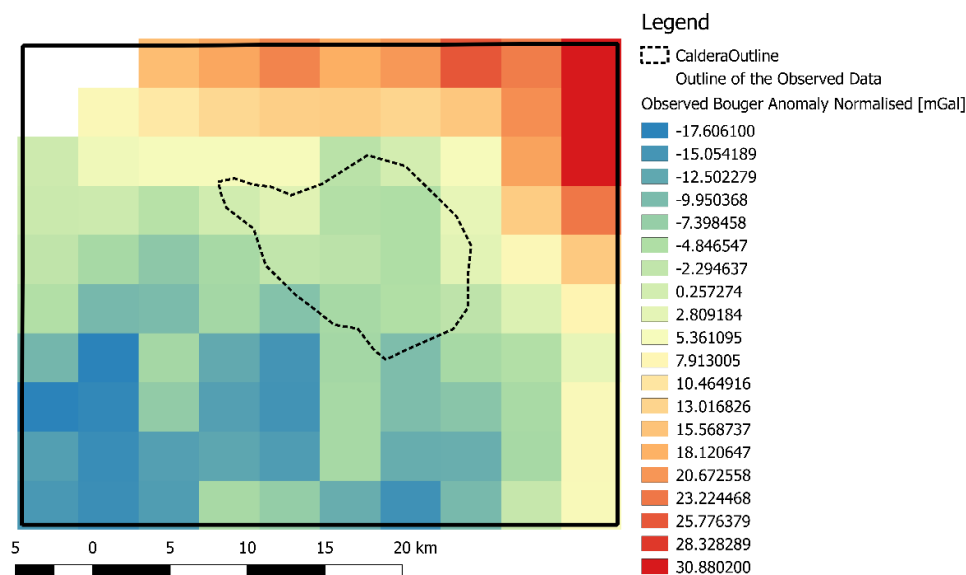


Figure 27. Gravity anomaly map (based on gravity map from García-Estrada, 1992) with units in mGal and discretised to 100 cells – allowing 97 data points. Dashed line marks the outline of the Los Humeros caldera. Black solid line marks the outline of the map that fits into the model space of the structural model by Evanno (2017).

Comparison to Forward model

The digitised gravity anomaly map (Figure 29), at first glance, seems quite different to the forward model of the Los Humeros area (Figure 28), described in the previous sections. The observed data range between -23 and +37 mGal, and the forward modelled gravity ranges between -9 and +17 mGal. An important feature is the high gravity anomaly south-east of the caldera, a result of the thicker bowl-shaped feature within andesite layer. This high anomaly does not appear in the observed data, where the area of caldera is within average values. The map of the observed measurements reveals high anomalies in the north-east and low anomalies in the south-west, which are somewhat present in the forward model, however with a much smaller amplitude and different pattern. These differences between the forward modelled gravity anomaly and measured gravity anomaly are illustrated by subtracting the measured values from modelled values, as in the Figure 30. Naturally, the plot confirms that the forward model is too high in the south-west and too low in the north-east. Furthermore, the high anomaly due to the thicker andesite, in the south-east section of the caldera, is clearly visible in the difference plot. Numerical measures of difference also show a very low linear correlation between the modelled and observed values. First, we compare the observed data points with corresponding points of the forward model output, utilising the automatic output of the Basin3D software (obs_mod.csv file; columns OBS and MOD_MEAN). The point by point comparison of the 97 data points, between observed values and forward modelled values, yields correlation (PCC) of 0.26128. In contrast the correlation performed on the model cells versus the observed map, both with the same discretisation and area, resulted in the PCC value of 0.24370. Thus, the correlation deteriorates between model and the observed data, due to the interpolation of both data sets to populate the model space and the map space. Moreover, the error between the two data sets is slightly lower in the point by point comparison, than in the model to map comparison. First yields the MAE of 10.301 mGal, whereas second MAE of 10.364 mGal. The RMSE is also higher in the model to map comparison at 13.491 mGal, but in the point by point comparison at 12.434 mGal. This result is purely due to the interpolation within the observed map; in the forward model Basin3D software does not change the result of its cell by cell computation and there is no data assimilation performed. The QGIS software used to produce the observed map, however, has been set to triangular interpolation. Thus, we conclude that the worsening correlation between the model and the map is due to the interpolation of the map. As expected, the errors distribution (RMSE to MAE ratio) is better in the point by point analysis at 1.207, whereas the model to map error distribution is

at 1.302. Regardless of the comparison, point by point or model to map, the correlation between the observed and modelled data is very low, almost completely uncorrelated. And due to the error distribution, we also know that it is nearly uniformly uncorrelated. In the next section we test the performance of the data assimilation and the influence of the number of observed data points on the results. Some of the plots and data used in this analysis, not included here, can be found in *Appendix II*.

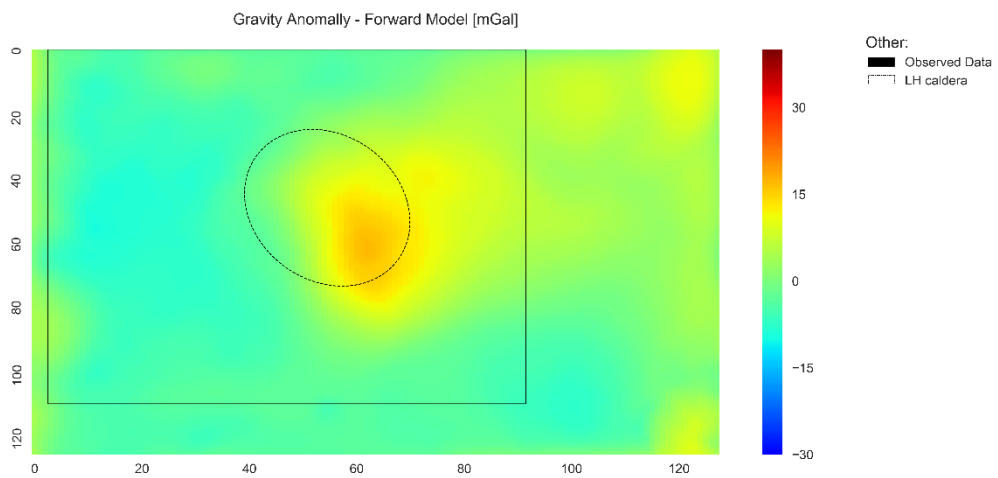


Figure 28. Forward modelled gravity anomaly. Modelling based on the structural model by (Evanno, 2017). Gravity values range from -9.38 mGal to 16.88 mGal. The data is identical to Figure 17, but the colour bar limits have been change to allow the comparison with observed gravity data in Figure 29. Dashed line marks the outline of the Los Humeros caldera. Black solid line marks the outline of the observed data map that fits into the model space of the structural model by Evanno (2017).

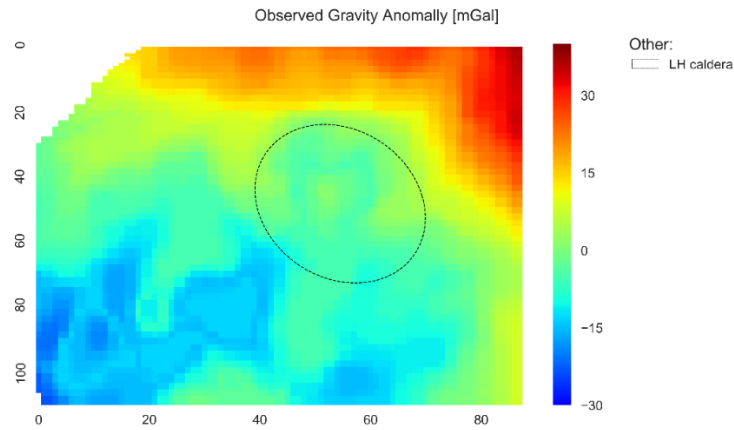


Figure 29. Gravity anomaly map (based on gravity map from García-Estrada, 1992) with units in mGal and discretised to match the forward model discretisation – allowing 9657 data points. Dashed line marks the outline of the Los Humeros caldera.

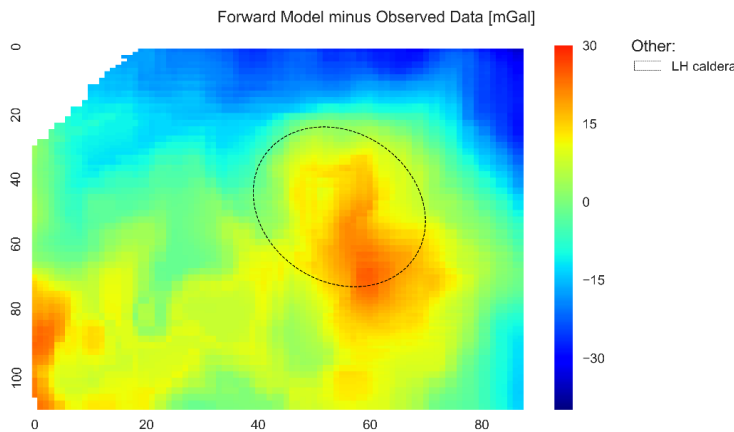


Figure 30. Difference between the forward modelled gravity anomaly and the observed data. High value areas are those in which the modelled anomaly is too high, and low value areas represent the sections of the model with too low gravity anomaly [mGal]. Dashed line marks the outline of the Los Humeros caldera.

Inverse Modelling Set-up

As described above in the Methodology Section, there are a number of parameters that must be set for inverse modelling. Firstly, the observed gravity anomaly data must be input as a file with specified measurement error. Secondly, the layer which will be calibrated must be specified, and other calibration parameters must be decided. Lastly, the number of ensembles and iterations must be set. We test each of these three setting groups by using our Los Humeros model, and comparing the results to the observed data map which has spatial discretisation corresponding to the model (9657 cells). We name the models as follows: first letter stands for 'K' thick or 'N' thin limestone layer in the structural model; 'o' stands for observations and the following letter indicates, if the number of

observations is large 'B' (9657 data points) or small 'S' (97 data points); 'e' stands for error (observation error) and the number following it indicates the error set in mGal; After the dash, the 'I' stands for iterations and the number indicates the iterations chosen for this model; The next letter 'e' stands for ensembles and the following number specifies the ensembles run in the model; The final section of the model name explains which layer has been chosen for density shift, 'L' for limestone and 'A' for Andesite, and the variation 'v' followed by the maximum value in kg/m³. As an example, in the first check we only change the number of iterations and ensembles, we set the first model to one iteration and 100 ensembles, thus the model name is KoSe20-I1e100Lv1000 (this translates to: structural model with thick limestone layer run with 97 observed data points and observation error at 20 mGal, the inverse settings are set to one iteration and 100 ensembles and the density variation is allowed in limestone layer to be varied by maximum of 1000 kg/m³). The second model has been set to one iteration and 1000 ensembles (model KoSe20-I1e1000Lv1000), third to four iterations and 100 ensembles (model KoSe20-I4e100Lv1000), and fourth to four iterations and 1000 ensembles (model KoSe20-I4e1000Lv1000). The layer chosen to be affected by the changes is the limestone layer, due to the conflicting information on its thickness (Arzate et al., 2018; Verma et al., 1990; Verma, 2011; Verma, 2013; Carrasco-Núñez et al., 2017; Evanno, 2017). We allow a maximum shift in density by 1000 kg/m³ from the current value of 2180 kg/m³, with triangular distribution and 10 cell range. There are 97 observed measurement data points with an error set to 20 mGal.

Iterations and Ensembles

The results of the first check are shown in Figure 31, images in the left column represent the gravity anomaly as a result of the inverse density modelling. The right column shows the resulting density variation in the limestone layer at depth of 687.5 meters, just below the andesite layer. Simply comparing the imaged density variation, we can notice the similarities between the models with 100 ensembles (first and third row), and the models with 1000 ensembles (second and fourth row). Models run with only 100 ensembles seem to show the density distribution less clearly and with more 'noise', when compared to the models with 1000 ensembles. Models with the higher number of ensembles show smoother progression of lower density to higher density anomalies. Also, the lower and higher density anomalies are located in the areas expected to be shifted, thus corresponding to the higher and lower gravity anomalies (Figure 31).

Furthermore, in Table 5 we compare the resulting gravity anomaly to the observed data, including as previously, point by point comparison and model to map comparison. Point by point comparison is the more reliable measure of difference for the assessment of data assimilation, due to the lack of the effect from interpolation. We do, however, include the model to map comparison, to highlight an unexpected differences in error distribution. Correlation in point by point comparison (PCC points), shows that both models with 100 ensembles, KoSe20-I1e100Lv1000 and KoSe20-I4e100Lv1000, are not as well correlated as the models with 1000 ensembles. Correlation coefficient for these two models is 0.91423 and 0.85767 respectively, further showing that the model with four iterations performs the worst when compared to the observations. MAE (MAE points) and RMSE (RMSE points) are also higher for the models with 100 ensembles than the models calculated with 1000 ensembles (Table 5). Yet, the error distribution is better for the lower ensembles models, due to the limited amount of samples used to calculate the mean. Evidently, a lower number of ensembles means that there is less chance to draw an outlier value in the values randomly selected in the given number of ensembles. Furthermore, this attribute of the RMSE/MAE ratio is due to the fact that it cannot be used to compare two series with different sample sizes, as described by Willmott & Matsuura (2005).

Additionally, the density range after the data assimilation and inverse modelling is much larger in the 100 ensemble models, taking almost the full allowed range set for this computation (Table 5). On the contrary, models run with 1000 ensembles, KoSe20-I1e1000Lv1000 and KoSe20-I4e1000Lv1000, correlate much better to the observed data and have lower errors, in both the point by point comparison and the model to map comparison. There is a very small difference between these two models, with correlation only improving for the one iterations model at third decimal place and the errors at a one decimal place (0.1 mGal). However, this very small improvement in correlation could be a result of the better values selected randomly in data assimilation, therefore we assume the correlation between the model with one iteration is equal to the model with four iterations. Thus, we confirm that the linear approach, i.e. one iteration, is correct for the inverse gravity modelling, using data assimilation and variation of the density values. Furthermore, we conclude that 100 ensembles for a model with 97 observed data points, are insufficient for an accurate computation of the inverse model of gravity anomaly. This is then compensated by large range in density variation and less constrained spatial distribution of the variation, resulting in 'noisy' data.

Therefore, based on this simple check, we conclude that the sufficient number of ensembles for accurate density variation, must be significantly larger than the number of observed data points. We have found that a number of ensembles that is equal to ten times the number of observed data points gives a reliable result. Consequently, further analysis of the Los Humeros models and any further check will be performed at one iteration and one thousand ensembles. Some of the plots and data used in this analysis, not been included here, can be found in *Appendix II*.

Comparison of iteration and ensemble set-up vs. Observed data														
Model	Iterations	Ensembles	Observed Data Points	Depth [m]	PCC points	MAE points [mGal]	RMSE points [mGal]	RMSE to MAE Ratio points	New Rho Range All Cells [kg/m3]	New Rho Range At Depth [kg/m3]	PCC voxel	MAE voxel [mGal]	RMSE voxel [mGal]	RMSE to MAE Ratio voxel
KoSe20-I1e100Lv1000	1	100	97	3312.5	0.91423	4.211	5.283	1.255	1712-3001	1703-2570	0.90057	6.726	8.153	1.212
KoSe20-I1e1000Lv1000	1	1000	97	3312.5	0.93524	3.666	4.903	1.337	1974-2872	1974-2443	0.92475	6.307	7.605	1.206
KoSe20-I4e100Lv1000	4	100	97	3312.5	0.85767	5.199	6.628	1.275	1729-3029	1729-2619	0.84346	7.364	9.114	1.238
KoSe20-I4e1000Lv1000	4	1000	97	3312.5	0.93332	3.677	4.790	1.303	1965-2855	1965-2409	0.91879	5.903	7.144	1.210
Comparison of Data points number in assimilation vs. Observed data														
KoSe20-I1e1000Lv1000	1	1000	97	3312.5	0.93524	3.666	4.903	1.337	1974-2872	1974-2443	0.92475	6.307	7.605	1.206
KoBe20-I1e1000Lv1000	1	1000	9657	3312.5	0.99481	1.034	1.318	1.275	1688-2966	1688-2663	0.99011	7.581	7.719	1.018

Table 5. Inverse modelled gravity anomaly at different number of iterations and ensembles compared to observed data – top section. Comparison of two models with different amount of observed data points during data assimilation versus the observed data points – bottom section. PCC is the Parson’s Correlation Coefficient, MAE is the Mean Absolute Error [mGal] and RMSE stands for Root Mean Square Error [mGal]. RMSE/MAE ratio is a measure of error distribution, value of 1 (minima) implies a uniform error distribution. PCC, MAE and RMSE/MAE ratio have been included for two types of comparison, ‘points’ comparison stands for point to point comparison in which values for a given observed point and values in the model space corresponding to the same point are compared. ‘Voxel’ comparison stands for comparison of the whole model to the whole observed data map, and includes interpolation errors. Additionally, New Rho Range columns, for both all of the cells within a layer (‘cells’) and only the layer cells at a given depth (‘depth’), give the new density range [kg/m3] as a result of data assimilation. Note: The depth specified in the table is the depth of the observation (measurement) and not the depth at which the density values have been extracted. New limestone density has been analysed at the depth of 687.5 meters.

Number of Observed Data Points

As a second check, we compare the KoSe20-I1e1000Lv1000 model mentioned above, to the KoBe20-I1e1000Lv1000 model with higher number of observed data points. For both models parameters are set at one iteration and 1000 ensembles, limestone layer shift with range of 1000 kg/m3, and observation data error at 20 mGal. The density variation is allowed to occur with triangular distribution within a range of 10 cells. KoSe20- model has 97 observed data points in data assimilation, whereas KoBe20- model has 9657 observed data points in the data assimilation. The high number of data points, at 9657 is equal to the number of cells in the model, meaning that each

cell of the model has an assigned observation value. This number of data points is however unrealistic for the area of 39.08 km by 32.57 km, with varied topography and thick vegetation cover, and has only been achieved by interpolation of the digitised anomaly map (*Observed Gravity Data* section). Therefore, we only use this high number of data points as a benchmark for the performance of data assimilation based on the number of observations. The amount of the observation values, their distribution throughout the model and their density, together with the allowed cell variation range for the density, will affect the performance of the data assimilation. This is due to the fact that not every model cell (on a 2D surface in x- y- directions) will hold an observation value, if observation values are scarce and only occur every few cells. Let's assume that the distance between observation points as expressed in number of model cells is 20. If we assign a range of variation for the density change to be at 5 cells, then during the data assimilation only that 5 cells will be affected by density change. Thus, the remaining cells between the observed data points will be unaffected. If however, the range for the density change will be set to 20, the effect of the neighbouring observed data points will overlap, causing the density to vary more smoothly. Due to this effect, the PCC points for the model with number of observations equal to the number of model cells is at 0.99481, but the model with the lower number of observations is at 0.93524. Also, the errors are much lower for the model with larger number of observations, only at around 1 mGal, and error distribution is also improved.

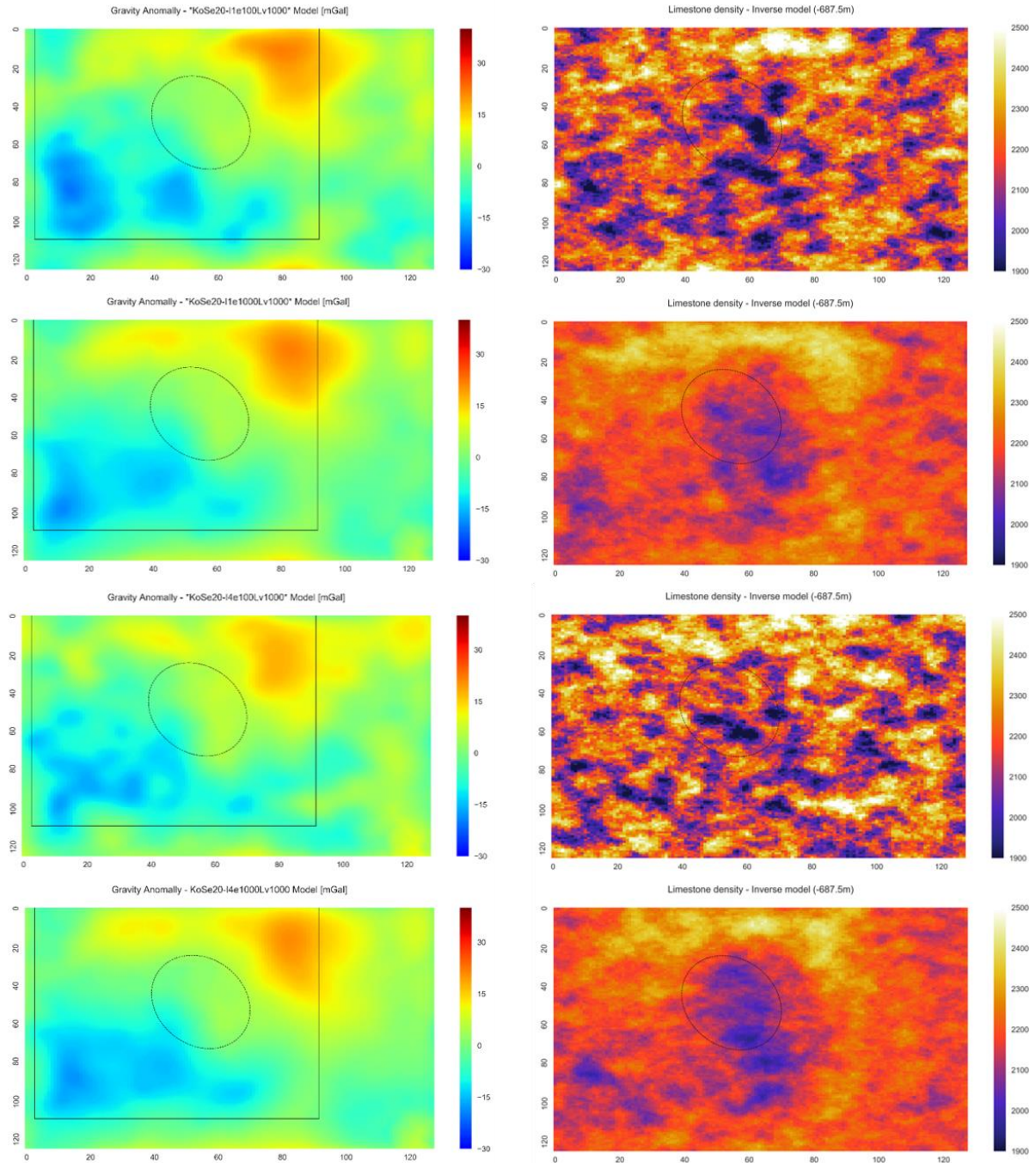


Figure 31. Inverse modelled gravity anomaly at different number of iterations and ensembles (left) [mGal], and visualised density shift for limestone layer at depth of 687.5 meters [kg/m³] (right). Top row: model run at one iteration and 100 ensembles. Second row: model run at one iteration and 1000 ensembles. Third row: model run at four iterations and 100 ensembles. Bottom row: model run at four iterations and 1000 ensembles. Dashed line marks the outline of the Los Humeros caldera. Black solid line marks the outline of the observed data map that fits into the model space of the structural model by Evanno (2017).

The images in the right column in Figure 32 show the newly computed density for the limestone layer for KoSe20- model (top) and KoBe20- model (bottom). The new density range is larger in the model with more observations used in the assimilation, and the density image seems more 'noisy' but does

still show well defined high and low anomalies. This ‘noisiness’ could possibly be explained by the insufficient number of ensembles, at 1000 it is ten times lower than the number of observations (9657 data points). Increasing the number of ensembles to match the number of data points, would allow for a smoother density variation. Except, the running time is expected (on a standard Quad-Core PC) to increase tenfold with each next 10^n number of ensembles, from 1 hour for 100 ensembles to 10 hours for 1000 ensembles. Therefore, increasing the number of ensembles is not possible in this study. And as we have previously mentioned, the 9657 observed data points are only possible when interpolating the data or a map, and in practice it is more reasonable to gather somewhere around 97 measured data points for the study area. This would correspond to a grid of approximately 3km by 3km. The KoSe20- model performs very well based on this lower number of data points, with PCC points correlation at 0.93524 and errors at acceptable level of 3.7 mGal. Thus, we expect that with adjustment to other parameters, for example lowering the observed data error and allowing variability in andesite layer, we can achieve even better results. Some of the plots and data used in this analysis, not been included here, can be found in *Appendix II*.

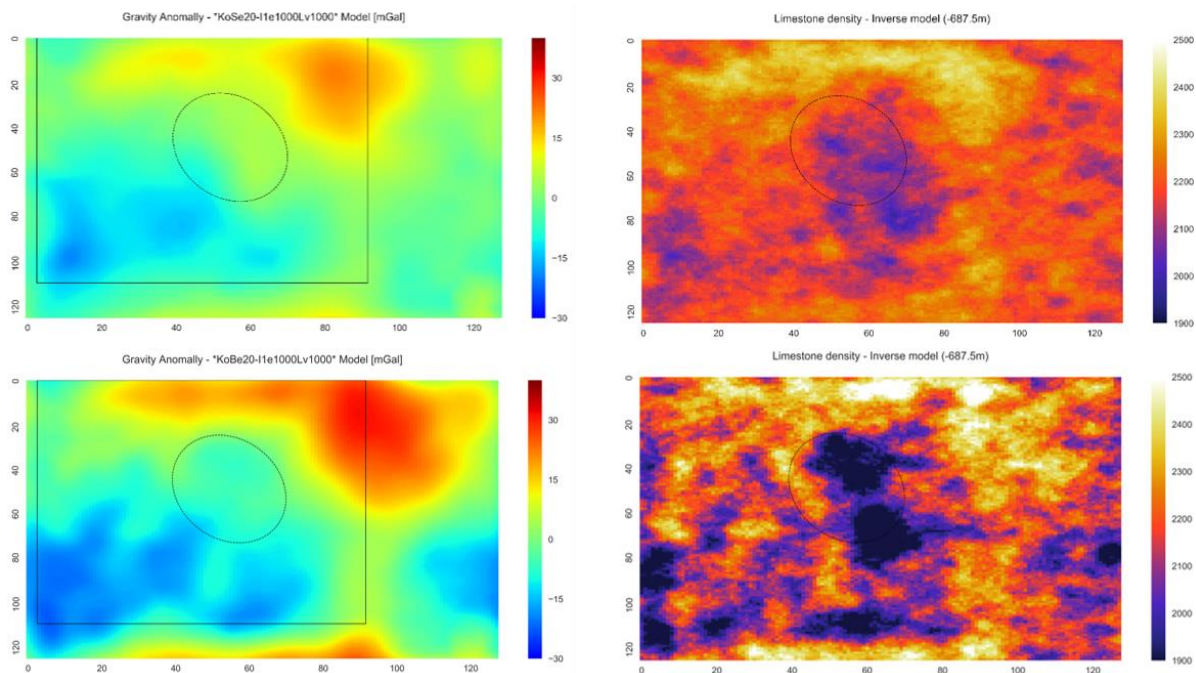


Figure 32. Inverse modelled gravity anomaly using different number of observed data points during data assimilation (left) [mGal], and visualised density shift for limestone layer at depth of 687.5 meters [kg/m3] (right). Top row: model run with data assimilation containing 97 data points. Bottom row: model run with data assimilation containing 9657. Dashed line marks the outline of the Los Humeros caldera. Black solid line marks the outline of the observed data map that fits into the model space of the structural model by Evanno (2017).

Observed Error

The observed data are in the range from -19 mGal to 33 mGal with a standard deviation (1σ) of 12.43 mGal, and the priori forward model to be used for data assimilation has a range of -9 mGal to 17 mGal, with a standard deviation (1σ) of 6.47 mGal. One standard deviation (1σ) of the observed data is much higher, almost double (2σ), than the standard deviation (1σ) of the data in the forward model. The forward model has $1\sigma = 6.47$, $2\sigma = 12.94$, $3\sigma = 19.41$, $4\sigma = 25.88$ and $5\sigma = 32.35$. And the highest absolute value of the observed data is at 33 mGal, meaning that potentially a value in the model would have to be shifted by five standard deviations or more (more for values below zero), if the error was set to be only 1 mGal. As an example, let's assume that the model and the observed data are highly correlated, so we know that the model value for a given point is higher or equal zero for a positive observed value at the same point, then:

$$(Observed\ Max\ Value - Min\ Model\ Value) - Error = Maximum\ Modelled\ Value\ Shift$$

$$(33\ mGal - 0\ mGal) - 1\ mGal = 32\ mGal$$

$$Maximum\ Modelled\ Value\ Shift = 32\ mGal$$

Thus, an error of 1mGal will give us a maximum shift of 32 mGal, corresponding to a shift just below five standard deviations of the model values. However, if we assume that the observations and the priori forward model are anti-correlated, we must consider scenario in which the lowest value of the model corresponds to a point with the highest value of the observations, then:

$$(Observed\ Max\ Value - Min\ Model\ Value) - Error = Maximum\ Modelled\ Value\ Shift$$

$$(33\ mGal - 1\ mGal) - (-9\ mGal) = 41\ mGal$$

$$Maximum\ Modelled\ Value\ Shift = 41\ mGal$$

In this case the maximum model value shift would be higher than five standard deviations for an observation error set to 1 mGal. The error should be set so that the maximum possible shift would not exceed five standard deviations. Due to the regularisation during the data assimilation, there is a very small probability that the modelled value will be shifted by more than four standard deviations, $4\sigma = 25.88$. Two standard deviations of observed data are almost equal to four standard deviations of our priori model. We check the maximum difference between the observed data points and the corresponding model data points, and we find that the maximum difference is 28.06 mGal. Thus, assuming a shift of four standard deviations that equals the shift of 25.88 mGal, and knowing that

the maximum difference between the values is 28.06 mGal, we calculate that the smallest error to satisfy the above restrictions is 2.18 mGal.

In Table 6 we present the correlation and MAE values for the models with different observed error. Starting with a large error of 20 mGal, we then decrease the value to 10 mGal, following by 5 mGal, 2 mGal and 1 mGal (Table 6). As we have explained above, both the observed error of 2 mGal and 1 mGal are too small. The error of 1 mGal, exceeds the five standard deviations shift, and the error of 2 mGal exceeds the four standard deviations shift, decreasing the likelihood of a high correlation. The correlation values in the point by point comparison (column PCC points) and the MAE (points) values are the best for the models with 5 mGal and 10 mGal error. A very high error of 20 mGal does not force the model values to shift enough towards the observations, resulting in a larger MAE error and worse correlation. Furthermore, the error distribution is not significantly affected by changing the observed error, as long as the error is within four standard deviations. Some of the plots and data used in this analysis, not been included here, can be found in *Appendix II*.

Comparison of observed data error in assimilation vs. Observed data												
Model	Observed Data Error [mGal]	Depth [m]	PCC points	MAE points [mGal]	RMSE points [mGal]	RMSE to MAE Ratio points	New Rho Range All Cells [kg/m3]	New Rho Range At Depth [kg/m3]	PCC voxel	MAE voxel [mGal]	RMSE voxel [mGal]	RMSE to MAE Ratio voxel
KoSe20-11e1000Lv1000	20	3312.5	0.93524	3.666	4.903	1.337	1974-2872	1974-2443	0.92475	6.307	7.605	1.206
KoSe10-11e1000Lv1000	10	3312.5	0.98144	1.880	2.546	1.354	1825-2933	1825-2542	0.96631	6.406	7.014	1.095
KoSe5-11e1000Lv1000	5	3312.5	0.99210	1.159	1.605	1.385	1755-3019	1755-2587	0.97251	7.472	7.874	1.054
KoSe2-11e1000Lv1000	2	3312.5	0.98305	2.080	2.632	1.265	1809-2982	1809-2556	0.95612	6.355	7.222	1.136
KoSe1-11e1000Lv1000	1	3312.5	0.99024	1.325	1.775	1.339	1704-2970	1704-2570	0.93341	6.573	7.470	1.136

Table 6. Inverse modelled gravity anomaly at different value of the observed data error compared to observed data. PCC is the Parson's Correlation Coefficient, MAE is the Mean Absolute Error [mGal] and RMSE stands for Root Mean Square Error [mGal]. RMSE/MAE ratio is a measure of error distribution, value of 1 (minima) implies a uniform error distribution. PCC, MAE and RMSE/MAE ratio have been included for two types of comparison, 'points' comparison stands for point to point comparison in which values for a given observed point and values in the model space corresponding to the same point are compared. 'Voxel' comparison stands for comparison of the whole model to the whole observed data map, and includes interpolation errors. Additionally, New Rho Range columns, for both all of the cells within a layer ('cells') and only the layer cells at a given depth ('depth'), give the new density range [kg/m3] as a result of data assimilation. Note: The depth specified in the table is the depth of the observation (measurement) and not the depth at which the density values have been extracted. New limestone density has been analysed at the depth of 687.5 meters.

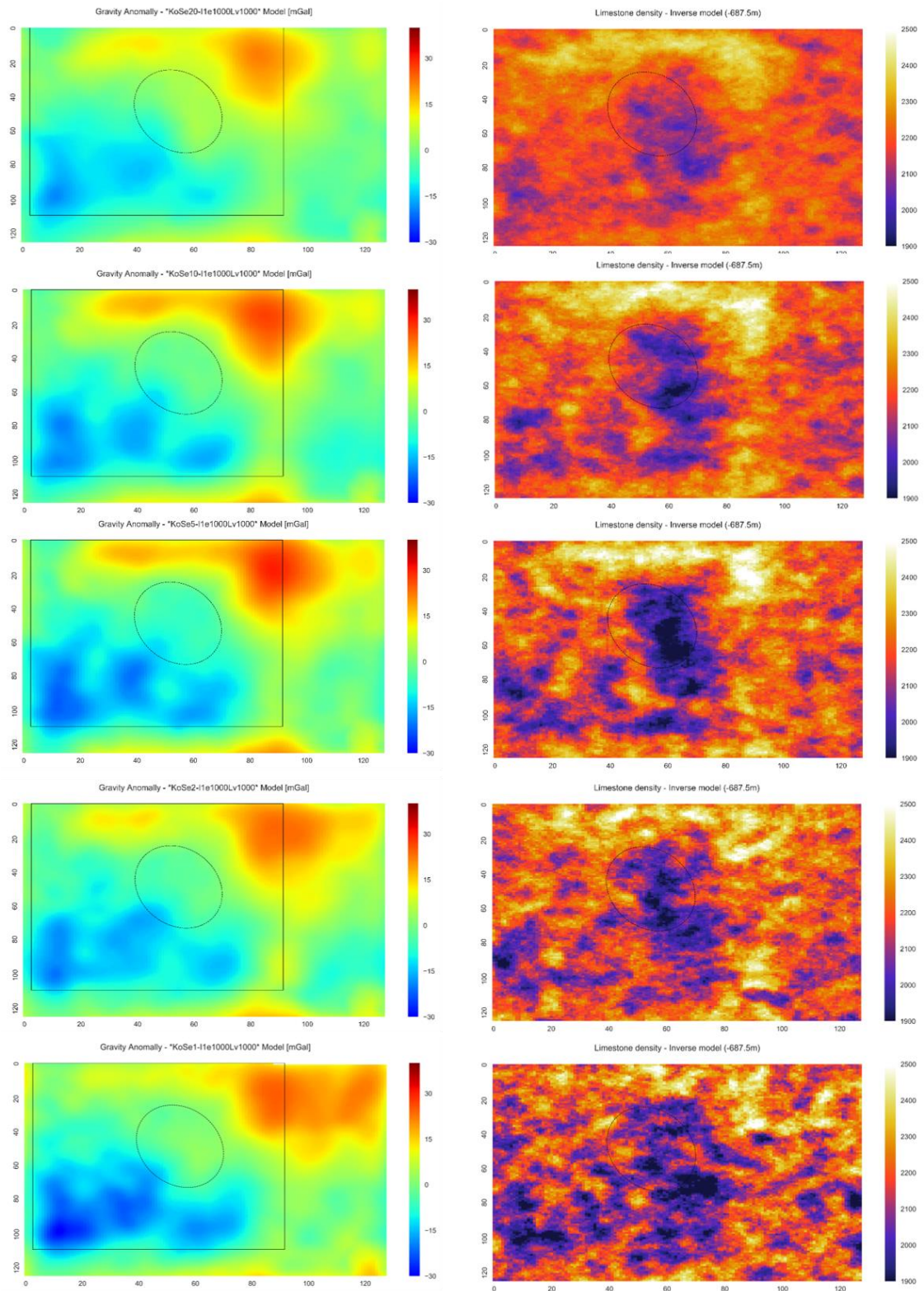


Figure 33. Inverse modelled gravity anomaly at different value of the observed data error (left) [mGal], and visualised density shift for limestone layer at depth of 687.5 meters [kg/m3] (right). Top row: model run with observed data error of 20 mGal. Second row: model run with observed data error of 10 mGal. Third row: model run with observed data error of 5 mGal. Fourth row: model run with observed data error of 2 mGal. Bottom row: model run with observed data error of 1 mGal. Dashed line marks the outline of the Los Humeros caldera. Black solid line marks the outline of the observed data map that fits into the model space of the structural model by Evanno (2017).

New density range (Table 6) becomes larger with a decreased observed error, this is expected as the larger shift of model values is reflected in the larger variation of the density. Density images (right column) for the limestone layer in Figure 33, show an increased divergence of the high and low anomalies from the mean. The density images for the observed errors of 20 mGal (top row), 10 mGal (second row) and 5 mGal (third row), retain their low and high anomaly pattern only changing the range of its values. However, the density variation of the lower observed errors became 'noisy', and the anomaly patterns undistinguishable. Next, we select the 5 mGal error as a suitable setting for the inverse modelling of the Los Humeros area, and limit the density variation to 500 kg/m³, so that the limestone layer values will fall within the physical values of limestone 1930-2900 kg/m³ (Heagy et al., 2016). Some of the plots and data used in this analysis, not been included here, can be found in *Appendix II*.

First Inverse Model of Los Humeros

Final inverse model of Los Humeros based on the existing structural model by Evanno (2017), was set to one iteration and one thousand ensembles, with 97 observed data points in the data assimilation and observed error of 5 mGal. The first layer chosen to be varied was the limestone layer, due to the conflicting information on its thickness (Arzate et al., 2018; Verma et al., 1990; Verma, 2011; Verma, 2013; Carrasco-Núñez et al., 2017; Evanno, 2017). The allowed maximum variation within the limestone layer was a shift of 500 kg/m³. The new density value after data assimilation is in range of 1850-2950 kg/m³ when taking to consideration all cells identified as limestone. Cells bordering on a surface with another layer may be assigned a density value different to that of the layer, during the building of structural model. This happens when a cell contains some part of the neighbouring layer, a weighted average is then taken of the two layers, and assigned as the density value of the cell. In the case of the limestone layer, which is located between the higher density andesite layer and a crustal layer, the densities on the layer surface would be averaged to be higher. Thus, the cell density range is higher than expected for limestone of density 2180 kg/m³ shifted by 500 kg/m³. An alternative approach is to look at the density at a given depth, if possible, where there are no intersections with another layer. At a depth of -687.5 meters, limestone layer is not affected by the averaging of the surface cells, and the density range is between 1850-2512 kg/m³ (Table 7). This range is consistent with the shift value of 500 kg/m³. Imaged density at the same depth in Figure 34 (top row, right column) shows less smoothness in density anomalies but clear high and low

anomalies. Point by point correlation of the modelled to observed values is very good at 0.98139, with MAE smaller than 2 mGal and the error distribution at 1.320. Comparison of the model to the map yields worse correlation and errors, as previously explained, due to interpolation of the data.

As a second scenario, we varied both the limestone layer and the andesite layer by 500 kg/m³. The new density range for andesite has lowered minima, due to the averaging of the cells on the border with limestone and ignimbrites. The density range of limestone has higher maxima due to the averaging on the surface with andesite. Cell range for limestone is 1891-2938 kg/m³, and for andesite is 2163-2812 kg/m³ (Table 7). The new range of the density at the depth of -687.5 meters for limestone layer is 1891-2534 kg/m³, and we can assume it to be accurate, because limestone does not intersect with another layer at this depth. However, the andesite layer has a very varied topography and does border with other layers at any given depth. A large area of andesite is located at a depth of 1812.5 meters above sea level, and situated just below the south-east part of the Los Humeros caldera. This andesite area (layer 4, third row, left column, Figure 34) is surrounded mostly by limestone (layer 5, third row, left column, Figure 34) and top tuff in the north-east (layer 2, third row, left column, Figure 34). In the centre of Los Humeros caldera, andesite layer is intersected by ignimbrites (layer 3, third row, left column, Figure 34). All those intersections of andesite with other layers make the minima and maxima calculation based on the layer ID less reliable. However, the calculated range still falls within the specified shift of 500 kg/m³, thus we assume it to be representative of the new density range after data assimilation. New layer densities have been imaged in Figure 34, where the image of limestone is in the second row (right column) and the image of andesite in third row (right column). In this scenario limestone layer has less density decrease in the low anomaly, than in the previous scenario, when only limestone was affected by the data assimilation. The low anomaly in the south-east part of the caldera, clearly visible in the first scenario, has now disappeared. This anomaly is now being compensated by the andesite layer, which decreased in density in the same south-east area of the caldera. And there is not much improvement on the correlation between the observed data points and the modelled values, with PPC at 0.98655, suggesting that the shift of 500 kg/m³ for both layers may be reduced, if a smaller density range is required. We will not perform further data assimilations to achieve this reduced range, as we believe this to be beyond the scope of this study.

As a third scenario, we set the density variation only to andesite layer with a shift of 500 kg/m³. Due to the limited extend of the andesite layer, the correlation of the modelled to the observed points

has dropped drastically to 0.53708. MAE in the point to point correlation has also increased up to 8.513 mGal. The increased range of the density shift was not sufficient to compensate for the limited extend of the andesite layer. Last row of the Figure 34 displays the imaged gravity anomaly (left column) and the newly modelled density (right column). It is instantly visible that the inverse model does not differ much from the forward model in Figure 28, with the exception of high anomaly shifted to north-east outside of the caldera. As the volume below the caldera was predominantly composed of the andesite layer, the high anomaly in the south-east of the caldera was allowed to be reduced to match the observed data. Similarly, the andesite layer extending to the east and north of the caldera, and facilitated the shift of the density in this area to match the high gravity anomaly in the observed data. These density variation is very well illustrated in the andesite density map (bottom row, right column, Figure 34).

Comparison of layer affected in assimilation vs. Observed data												
Model	Layer of density variation	Depth [m]	PCC points	MAE points [mGal]	RMSE points [mGal]	RMSE to MAE Ratio points	New Rho Range All Cells [kg/m3]	New Rho Range At Depth [kg/m3]	PCC voxel	MAE voxel [mGal]	RMSE voxel [mGal]	RMSE to MAE Ratio voxel
KoSe5-I1e1000Lv500	Limestone	3312.5	0.98139	1.930	2.548	1.320	1850-2950	1850-2512	0.96639	5.842	6.481	1.109
KoSe5-I1e1000LAv500	Limestone & Andesite	3312.5	0.98655	1.669	2.242	1.343	1891-2938 & 2163-2812	1891-2534 & 2232-2793	0.96951	6.606	7.133	1.080
KoSe5-I1e1000Av500	Andesite	3312.5	0.53708	8.513	10.717	1.259	2041-2938	2179-2937	0.49750	8.567	11.781	1.375

Table 7. Inverse modelled gravity anomaly with a density shift of 500 kg/m3 in different layers compared to observed data. PCC is the Parson's Correlation Coefficient, MAE is the Mean Absolute Error [mGal] and RMSE stands for Root Mean Square Error [mGal]. RMSE/MAE ratio is a measure of error distribution, value of 1 (minima) implies a uniform error distribution. PCC, MAE and RMSE/MAE ratio have been included for two types of comparison, 'points' comparison stands for point to point comparison in which values for a given observed point and values in the model space corresponding to the same point are compared. 'Voxel' comparison stands for comparison of the whole model to the whole observed data map, and includes interpolation errors. Additionally, New Rho Range columns, for both all of the cells within a layer ('cells') and only the layer cells at a given depth ('depth'), give the new density range [kg/m3] as a result of data assimilation. Note: The depth specified in the table is the depth of the observation (measurement) and not the depth at which the density values have been extracted. New limestone density has been analysed at the depth of 687.5 meters.

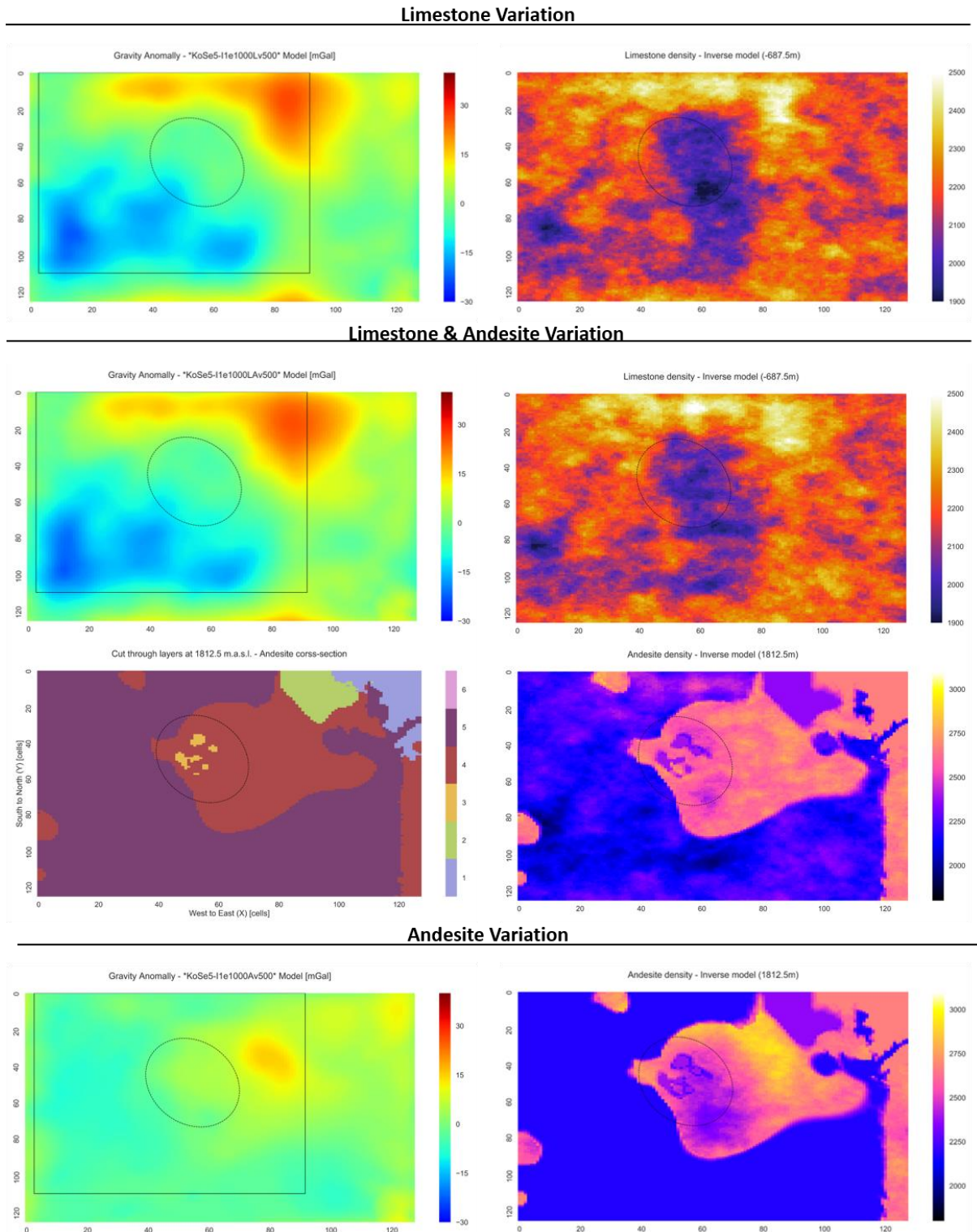


Figure 34. Inverse modelled gravity anomaly with a density shift of 500 kg/m³ in different layers (left) [mGal], and visualised density shift for limestone layer at depth of 687.5 meters [kg/m³] and for andesite layer at the elevation of 1812.5 meters (right). Top row: limestone variation. Second row: limestone and andesite variation. Third row: cross-section at the depth of 1812.5 meters showing the location of the andesite layer (left), and the corresponding density variation (right). Bottom row: andesite variation. Dashed line marks the outline of the Los Humeros caldera. Black solid line marks the outline of the observed data map that fits into the model space of the structural model by Evanno (2017).

All of the above scenarios, indicate, that the limestone layer is the key to achieving a better match of the modelled gravity anomaly to the observed gravity anomaly. Thus, following the suggestions of Verma in his papers (Verma et al., 1990; Verma, 2011; Verma, 2013), we will reduce the thickness of the limestone layer, and present the results in the next section. Some of the plots and data used in this analysis, not been included here, can be found in *Appendix II*.

Alternative Model of Los Humeros

Structural Model

The new structural model with a thinner limestone layer, retains the topography of the top four layers from the previous structural model. The top layers of 'air' (1), then top tuffs (2), ignimbrites (3) and andesite (4), remain unchanged in every aspect (Figure 35). However, the surface that forms the bottom of the limestone layer (5) and also the top of the crustal rock layer (6), changes its topography and depth. The new topography of the crustal layer (6) is the exact copy of the topography of the limestone layer, implying that the limestone layer (5) has now a uniform thickness of 2.5 km (cross-sections in Figure 36 and Figure 37). Thickness of the limestone layer has been set to 2.5 km based on the Verma's approach of 1 km (Verma et al. 2013), and on the more recent study by Arzate et al., who investigated the gravity anomaly of Los Humeros in his models, with the limestone layer set smaller or equal to 3 km (Arzate et al., 2018). The rest of the model below the limestone is filled with crustal rock (6). None of the densities have been changed in this new structural model, the cell discretisation has also been kept the same. Thus, the results of this new structural model can be compared to the results of the structural model by Evanno (2017).

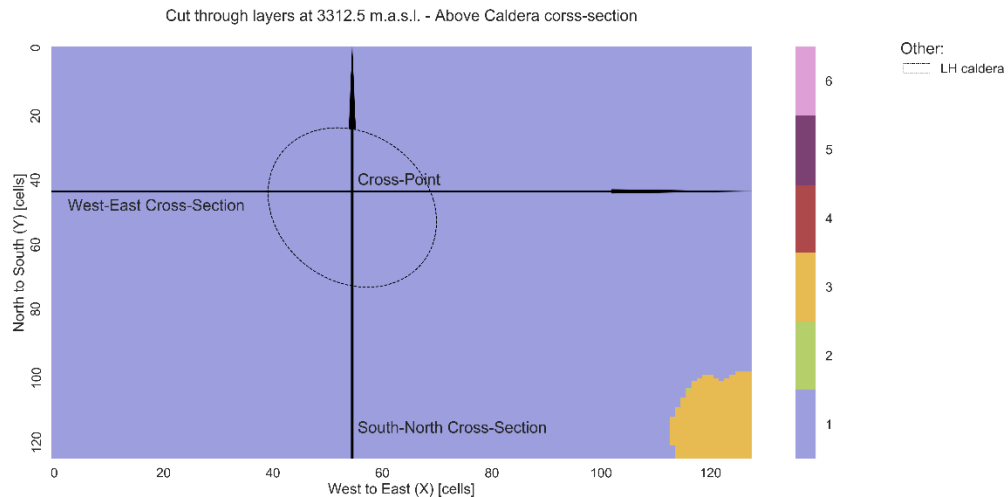


Figure 35. Alternative structural model of Los Humeros created in Basin3D Preprocess module. Layer 1 (violet) represents air layer above topography, Layer 2 (green) is the top tuffs also mentioned as pyroclastics, Layer 3 (yellow) is the ignimbrites, Layer 4 (dark red) is the Andesite, Layer 5 (purple) is the limestone basement which is now thinner and with uniform thickness, and Layer 6 (pink) is the crustal layer.

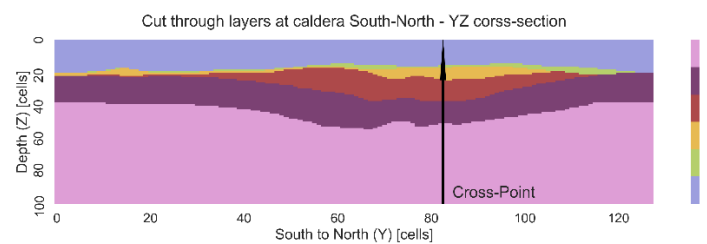


Figure 36. South to North cross-section of the Alternative Los Humeros structural model. Layer 1 (violet) represents air layer above topography, Layer 2 (green) is the top tuffs also mentioned as pyroclastics, Layer 3 (yellow) is the ignimbrites, Layer 4 (dark red) is the Andesite, Layer 5 (purple) is the limestone basement which is now thinner and with uniform thickness, Layer 6 (pink) is the crustal layer.

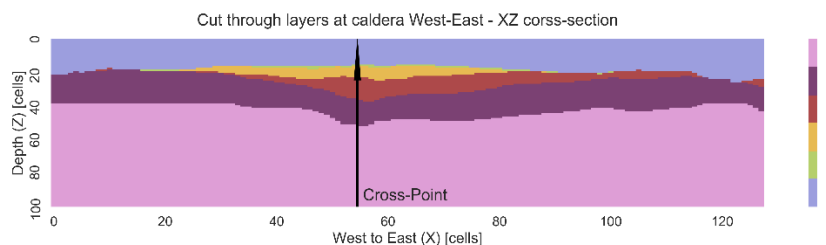


Figure 37. West to East cross-section of the Alternative Los Humeros structural model. Layer 1 (violet) represents air layer above topography, Layer 2 (green) is the top tuffs also mentioned as pyroclastics, Layer 3 (yellow) is the ignimbrites, Layer 4 (dark red) is the Andesite, Layer 5 (purple) is the limestone basement which is now thinner and with uniform thickness, and Layer 6 (pink) is the crustal layer.

Forward Model

The structural model with the thin limestone layer produces gravity anomalies (Figure 38), which are of a much smaller amplitude than the model with a thicker limestone layer (Figure 39). The high gravity anomaly in the south-east part of the Los Humeros caldera is reduced by around 10 mGal, and disappears almost completely. This is due to the thicker crustal layer that compensates the high density volume of andesite below the caldera, thus preventing the high anomaly in the south-east of the caldera from appearing. As expected, the correlation of the new model to the observed data is much improved, at 0.46740 for point to point comparison, in contrast to the correlation of 0.26128 in for the previous model. The error between the modelled values and the observed data points has also improved with MAE at 9.419 mGal, compared to 10.301 mGal. Yet, the error distribution is worse at 1.228, to previous 1.207. This is understandable, considering that the centre area of the model (around caldera) is now much better match to the observed data (Figure 40), than the south-west and north-east corners of the model (Figure 41). Next, we demonstrate the results of the inverse modelling and data assimilation of this structural model. Some of the plots and data used in this analysis, not included here, can be found in *Appendix II*.

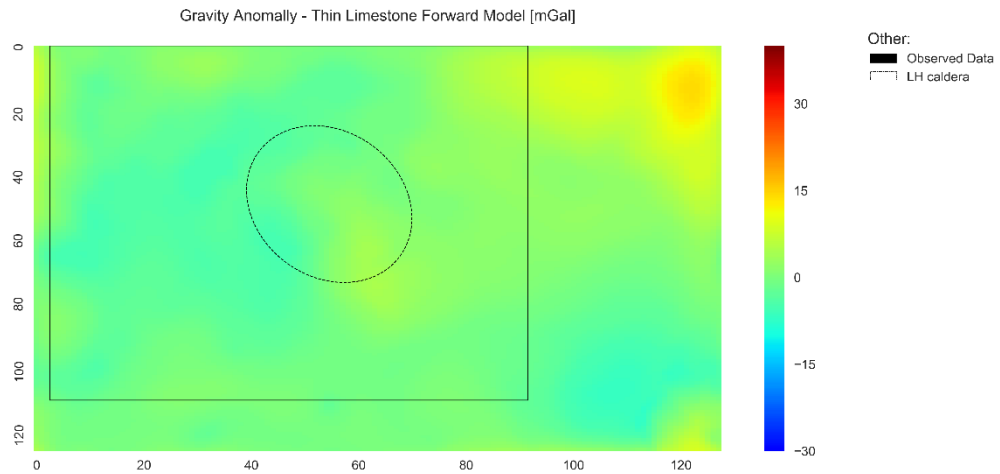


Figure 38. Forward modelled gravity anomaly. Modelling based on the alternative structural model of Los Humeros. Gravity values range from -7.06 mGal to 14.22 mGal. Dashed line marks the outline of the Los Humeros caldera. Black solid line marks the outline of the observed data map that fits into the model space of the structural model by Evanno (2017).

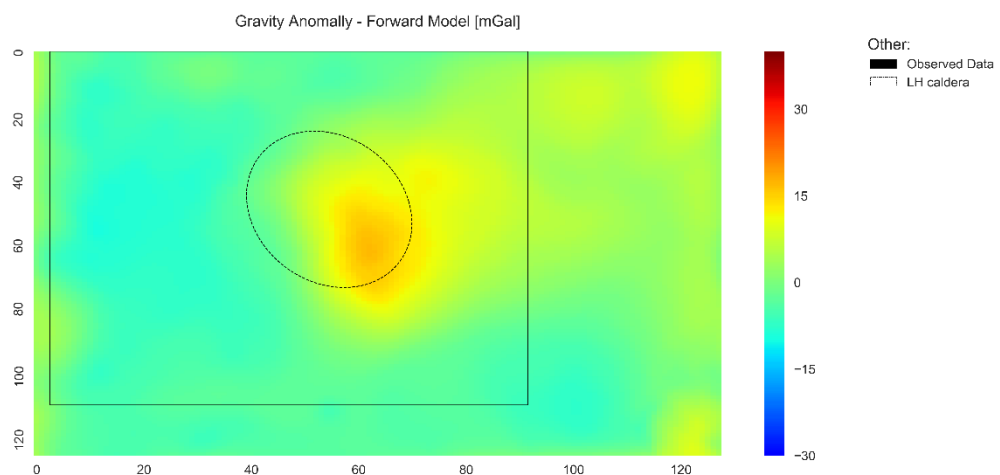


Figure 39. Forward modelled gravity anomaly. Modelling based on the structural model by (Evanno, 2017). Gravity values range from -9.38 mGal to 16.88 mGal. The data is identical to Figure 17, but the colour bar limits have been change to allow the comparison with observed gravity data in Figure 28. Dashed line marks the outline of the Los Humeros caldera. Black solid line marks the outline of the observed data map that fits into the model space of the structural model by Evanno (2017).

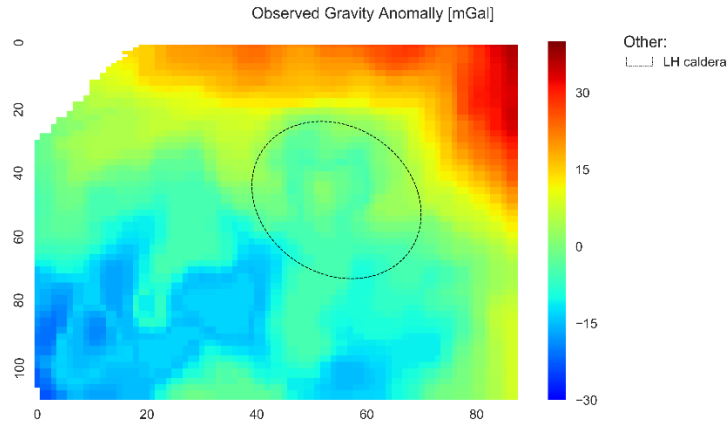


Figure 40. Gravity anomaly map (based on gravity map from García-Estrada, 1992) with units in mGal and discretised to match the forward model discretisation – allowing 9657 data points. Dashed line marks the outline of the Los Humeros caldera.

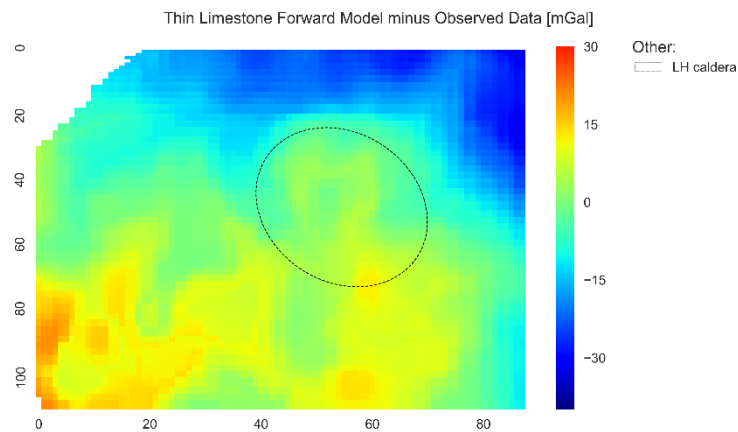


Figure 41. Difference between the forward modelled gravity anomaly (alternative structural model) and the observed data. High value areas are those in which the modelled anomaly is too high, and low value areas represent the sections of the model with too low gravity anomaly [mGal]. Dashed line marks the outline of the Los Humeros caldera.

Inverse Model

As for the previous structural model, we set the data assimilation parameters to 97 observed data. We calculate the observation error that will allow a shift equivalent to the shift for the previous structural model. In the previous structural model the maximum difference between the observed and modelled points was 28.06 mGal, and the set observational error was 5 mGal, this allowed a shift equivalent to 3.55 standard deviations of the modelled data. Thus, to allow the same shift of 3.55 standard deviations in the new model, we must first find the difference between the observed data

points and the corresponding points in the forward model. We find that the maximum difference is 28 mGal, and one standard deviation of the new forward model is equal to 2.42 mGal. Therefore, a shift of 3.55 standard deviations in the new model would be equal to the shift of 8.61 mGal. To allow for correct shift of the point with the highest difference between the observed and modelled value, the error would have to be at least 19.39 mGal. As a result, we set the error to 19.5 mGal to allow any rounding errors in the above computation. The model is set to run for one iteration and one thousand ensembles, with variation in both limestone and andesite layers and density shift of up to 500 kg/m³.

Results of inverse modelling of gravity anomaly for the structural model with thinner limestone layer are presented in Figure 42. The gravity anomaly map (top row, Figure 42) does not show any low or high amplitude anomalies, and based on images alone, looks almost identical to the forward model. The calculated correlation between the observed data points and modelled values for corresponding points is much higher than the correlation of the forward model, at 0.82542. Thus, although the anomalies have a low amplitude, they do have a similar distribution to that of the observed measurements. We calculate the MAE to be at 8.017 mGal, which is an improvement from the forward model error by around 1.4 mGal. Yet, the error distribution remains the same (to two decimal places) at a RMSE/MAE ratio of 1.221. A significant change from the inverse models of the previous structural model, is the smaller variation in the density. Both limestone layer and andesite layer, underwent a much smaller shift than in the former inverse modelling of a different structural model. The new density range for limestone is 2072 – 2718 kg/m³ including all of the cells of the layer, and 2119 – 2257 kg/m³ for limestone only at the depth of 687.5 meters. This shift is visualised in Figure 42 (second row), where a cross-section at the depth of 687.5 meters is shown (left) and a new density variation (right). As previously, the majority of the decrease in the density takes place in the south-east section of the Los Humeros caldera. The bulk of the increase in the density occurs in the north-east area, where the observed data forces the high gravity anomaly. Similarly, andesite layer range is also smaller, it has changed to 2329 – 2726 kg/m³ for all of the layer cells, and to 2349 – 2712 kg/m³ for the andesite at elevation of 1812.5 meters. The cross-section at the elevation of 1812.5 meters shows the andesite layer surrounded by limestone (third row, left column, Figure 42), and it has not been affected by the change in the thickness of the limestone layer. The visualised variation of the density within the andesite layer (third row, right column, Figure 42), displays a slight decrease in density below the south-east area of the Los Humeros caldera, but an overall increase in

the other parts of the layer. Some of the plots and data used in this analysis, not been included here, can be found in *Appendix II*.

(Thin) Limestone and Andesite Variation

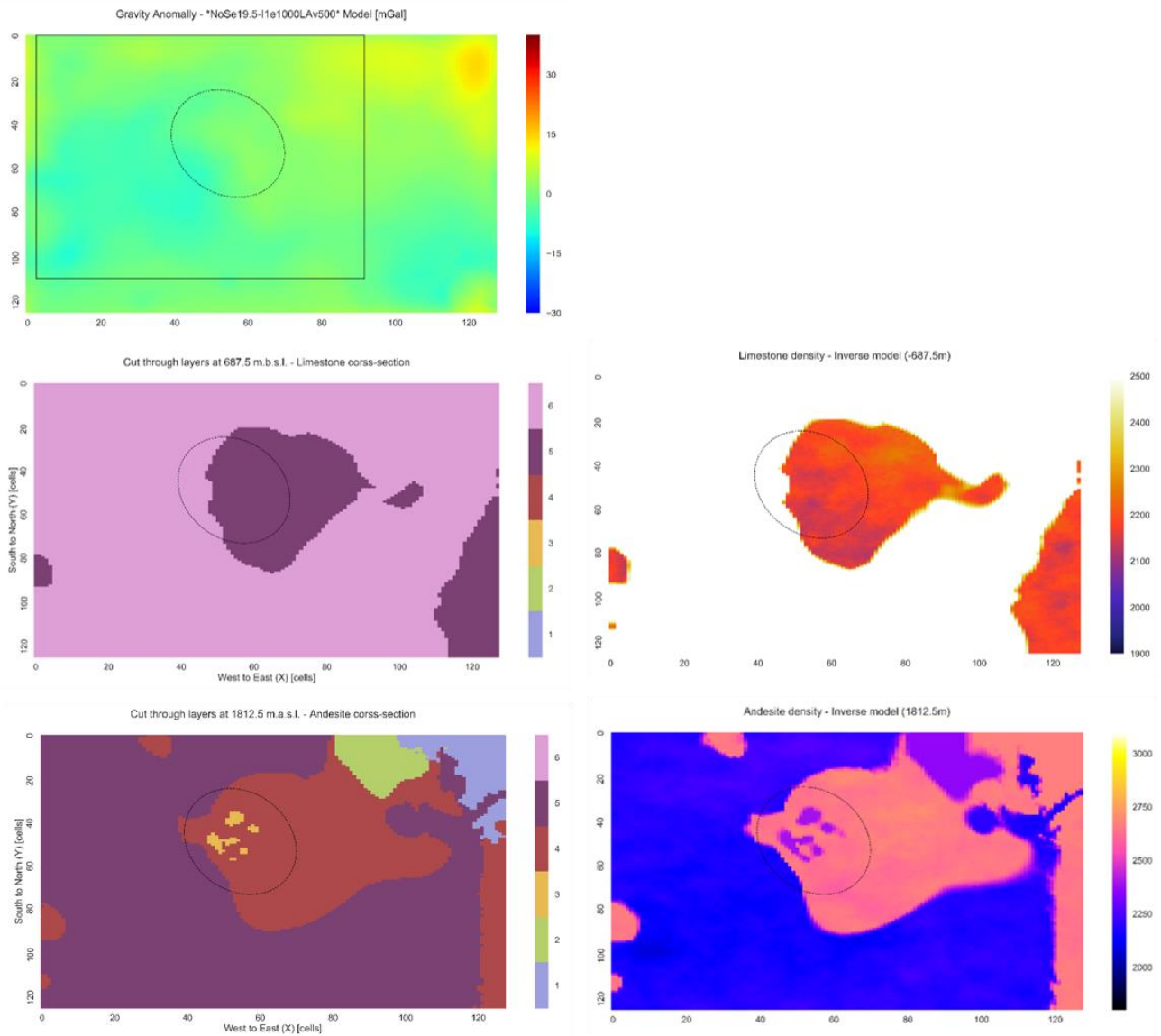


Figure 42. Inverse modelled gravity anomaly with a density shift of 500 kg/m³ in both limestone and andesite layers (top row, left) [mGal]. Second row: cross-section at the depth of 687.5 meters showing the location of the limestone layer (left) and visualised density shift for limestone layer at depth of 687.5 meters [kg/m³] (right). Bottom row: cross-section at the depth of 1812.5 meters showing the location of the andesite layer (left) and visualised density shift for andesite layer at the elevation of 1812.5 meters (right). Dashed line marks the outline of the Los Humeros caldera. Black solid line marks the outline of the observed data map that fits into the model space of the structural model by Evanno (2017).

Inverse Modelling Discussion

Forward model of the Los Humeros is based on the structural model by Evanno (2017) and density values commonly used for that area (Verma et al., 1990; Verma, 2011; Verma, 2013). These densities are 2360 kg/m³ for the top layer of top tuffs, 2394 kg/m³ for the underlying layer of ignimbrites, 2650 kg/m³ (Hantschel & Kauerauf, 2009) for the third layer of andesite and 2180 kg/m³ for the limestone basement, and additionally a bottom crustal layer density of 2680 kg/m³. We assume that the chosen densities are a good approximation of the densities of the Los Humeros geothermal complex. Based on this assumption, we assess the forward modelled gravity anomaly of the area, which to certain degree reflects the thickness of the denser layer, andesite. The high volume of andesite below the south-east part of the Los Humeros caldera (Figure 14 and Figure 15) creates a high gravity anomaly (Figure 39) and its extension towards the north-east increases the gravity in that area to small amplitude high gravity anomaly. These findings do not match the observed measurements of Los Humeros gravity anomaly. The digitised observations map (Figure 24) is significantly different to the forward model, both in terms of distribution of the anomalies and their amplitudes. Even when considering the limitations of this study, which include the low reliability of the observations map and the errors arising from its digitisation process (including interpolation), we find the low correlation (0.24370) between the modelled and observed gravity anomaly unsatisfactory. We find that the alternative structural model of the area (Figure 36 and Figure 37), with a thinner layer of limestone of a uniform thickness, generates a gravity anomaly map that correlates much better to the observed values (0.46740). In this new model the high gravity anomaly in the south-east part of the Los Humeros caldera is significantly reduced in amplitude. This suggests that the thicker crustal layer (Figure 36 and Figure 37), a result of uniform thinner limestone, acts as a 'density buffer' by balancing out the changing volumes of andesite. Thus, based on the forward models alone, we already have an indication that the large volume of andesite below the Los Humeros caldera must be offset, by either a geometry change or a density change. The geometry change must be that the higher density layer, one underlying the limestone basement, must have a lower topography below the caldera, and higher topography elsewhere.

Alternatively, a different density in the layers below the caldera would be required to achieve the same result. Therefore, we inversely model the gravity anomaly, using data assimilation of the observed data, and allowing the densities in limestone and andesite layers to shift by maximum of 500 kg/m³. The result of inverse modelling for the first structural model achieves a very good

correlation with the observed data at 0.98655, and relatively small error of 1.67 mGal. Yet, to achieve this correlation the density is shifted to a large new range of 1891-2534 kg/m³ for limestone and 2232-2793 kg/m³ for andesite. On the contrary, the new density range for the inverse modelling results of the alternative structural model is much more narrow, 2119 – 2257 kg/m³ for limestone and 2349 – 2712 kg/m³ for the andesite. Although, the correlation between the gravity anomaly and the observed data is much lower at 0.82542. In both cases the variation of density occurs in the same areas and the shift takes place in the same direction. As mentioned above, the primary area of change is the south-east part of the caldera, in which densities always decrease in value. Another part of the model, in which the density becomes lower is the south-west region where the observed values show a spatially large low gravity anomaly. A density increase takes place in the north-east section of the model to match the high gravity anomaly of large amplitude and spatial extend. Hence, on the assumption that the observed data is representative and the initial densities have been selected correctly, we identify three main areas of the structural model that require geometrical change or density value change. South-east part of the Los Humeros caldera must be offset by lower density, the same of the south-west section of the model. The north-east area of the model must be adjusted to a higher density to reflect the high gravity anomaly in the observed data.

Therefore, we suggest a further study on the gravity anomaly of Los Humeros geothermal field, consisting of forward and inverse modelling based on two new structural models. First of the proposed new models (left column, Figure 43), has a limestone layer much thicker in the south and west of the model (maximum thickness in south-west), but thinner towards the north and east (minimum thickness in the north-east). We expect that this irregular thickness of the limestone layer will generate a low gravity anomaly in the south-west and a high anomaly in the north-east. The effect on the south-east part of the caldera should compensate for the thickness of the andesite layer, and result in neither high nor low gravity anomaly. Assuming that the ultimate goal would be to adjust the structural model, so that the forward gravity model is in a very good correlation to the observed data, and the density adjustment is not necessary, we suggest a second structural model with an additional low density layer (right column, Figure 43). In the case of Los Humeros, a known volcanic system and a geothermal complex, this low density layer (layer 7, Figure 43) would represent a potential magma chamber, or partially melted hot material. Based on the location of gravity anomalies and the locations of density changes during data assimilation, we expect to find such a layer at its minimum depth just off-centre of the Los Humeros caldera (off-centre to the north-west),

where a small amplitude low gravity anomaly is shown in the observed data (Figure 40). We further suggest the layer to remain shallower in the west and south than in the east and north, additionally increasing the amplitude of the gravity anomalies in those areas.

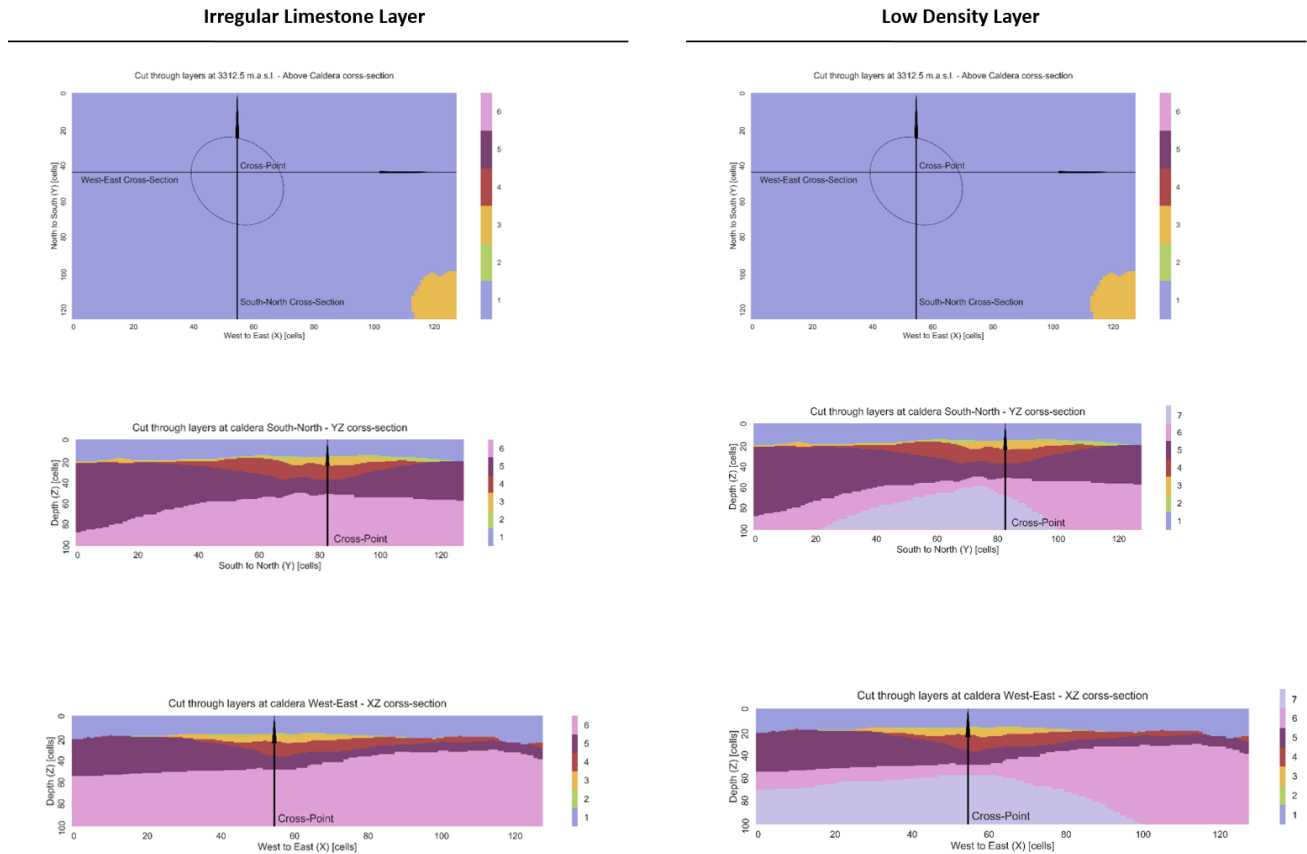


Figure 43. New suggested structural models for Los Humeros study area. Left: irregular limestone layer thicker in the south-west and thinner in the north-east. Right: new additional layer of low density with a minimum depth just below the caldera. Layer 1 (violet) represents air layer above topography, Layer 2 (green) is the top tuffs also mentioned as pyroclastics, Layer 3 (yellow) is the ignimbrites, Layer 4 (dark red) is the Andesite, Layer 5 (purple) is the limestone basement which is now thinner and with uniform thickness, Layer 6 (pink) is the crustal layer and Layer 7 (grey-blue) is the new proposed low density layer.

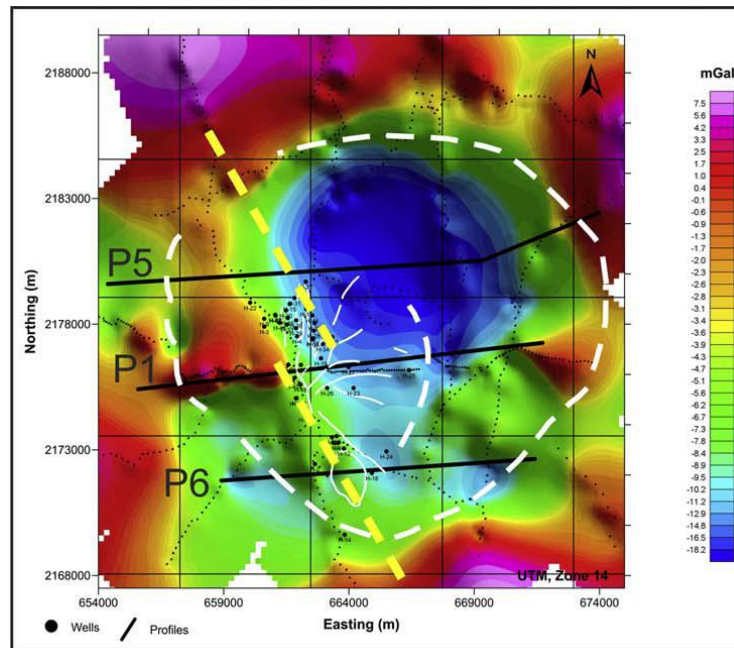


Figure 44. Measured gravity anomaly from Arzate et al. (2018). Gravity anomaly ranges from -11 mGal to 12 mGal.

In a recent study of the of the Los Humeros geothermal complex restricted to the Los Humeros caldera, Arzate et al. (2018) utilised an astonishing number of 718 measurements of gravity from accurately levelled stations in the Los Humeros area. Generated map of gravity anomaly varies from -11 mGal to 12 mGal, with a high amplitude, almost circular, low gravity anomaly in the centre of the Los Humeros caldera (Figure 44). As an explanation, authors suggest a highly fractured intrusive body, with density similar to crustal density, but substantially heated from below. This is in contrast to the model by Verma et al. (2011), in which the basaltic magma chamber is extensive and located at shallow depths of around 5 km. Therefore, we suggest that the proposed new structural models will be used to forward and inverse model the gravity anomaly, and the results will be compared to other structural and gravity studies. Furthermore, we advise that such study would be performed with newly measured observed gravity data, which undoubtedly would be more precise than the currently available map.

CONCLUSIONS

This study aimed to benchmark the forward modelling of gravity anomaly in Basin3D software versus available comparable methodology. This has been achieved by comparing the results of forward modelling in Basin3D and Geomodeler (Geophysics, 2015), based on an identical structural model and matching modelling parameters. Used measures of difference were Mean Absolute Error (MAE) for error description, and Root Mean Square Error (RMSE) to MAE ratio for error distribution description. Furthermore, Pearson's Correlation Coefficient (PPC) was used as a measure of equivalence. Results have been compared with a great degree of correlation, exceeding three decimal places of correlation, and minimal error below 0.1 mGal. Thus, the conclusion of the benchmarking tests, is that the Basin3D software performs well against a commercially available equivalent software. Moreover, it is a reliable tool for forward modelling of gravity anomaly and can be successfully utilised in the geothermal applications.

Another aim of this study was to investigate the gravity anomaly of the Los Humeros geothermal complex, based on an available structural model of the area (Evanno, 2017). Potential constraints on the structural model and the densities of the layers have been explored by inverse modelling with data assimilation. As a result, we identify three major areas of interest, in which either the densities or the geometry of the structural model are inconsistent with the measured gravity anomaly. We point out the required changes and suggest new structural models for further studies. These new suggested models are based on an assumption that the used observed data, even if not very precise, were somewhat accurate. Another assumption was made, that the density values widely cited in the literature (Verma et al., 1990; Verma, 2011; Verma, 2013) are accurate for the area of Los Humeros. We expect that a follow up study, using suggested new structural models, will be able to provide more accurate evidence on a low density anomaly below the Los Humeros caldera. Thus, we conclude that the results of this study were a good initial investigation into the gravity response of the structural model by Evanno (2017), that also provided constraints for further studies.

REFERENCES

- Arzate, J., Corbo-Camargo, F., Carrasco, G., Hernández, J. and Yutsis, V., 2018. The Los Humeros (Mexico) geothermal field model deduced from new geophysical and geological data. *Geothermics*, 71, pp.200-211.
- Bhattacharyya, B.K., 1967. Some general properties of potential fields in space and frequency domain: a review. *Geoexploration*, 5(3), pp.127-143.
- Buda, A. and Jarynowski, A., 2010. *Life-time of correlations and its applications vol. 1*, Wydawnictwo Niezależne: 5–21, December 2010. ISBN 978-83-915272-9-0.
- Burger, H.R., Sheehan, A.F. and Jones, C.H., 2006. *Introduction to applied geophysics: Exploring the shallow subsurface*. WW Norton.
- Calcagno, P., Chilès, J.P., Courrioux, G. and Guillen, A., 2008. Geological modelling from field data and geological knowledge: Part I. Modelling method coupling 3D potential-field interpolation and geological rules. *Physics of the Earth and Planetary Interiors*, 171(1-4), pp.147-157.
- Caratori Tontini, F., Cocchi, L. and Carmisciano, C., 2009. Rapid 3-D forward model of potential fields with application to the Palinuro Seamount magnetic anomaly (southern Tyrrhenian Sea, Italy). *Journal of Geophysical Research: Solid Earth*, 114(B2).
- Carrasco-Núñez, G., López-Martínez, M., Hernández, J. and Vargas, V., 2017. Subsurface stratigraphy and its correlation with the surficial geology at Los Humeros geothermal field, eastern Trans-Mexican Volcanic Belt. *Geothermics*, 67, pp.1-17.
- Cedillo, F.R., 1999. Modelo hidrogeológico de los yacimientos geotérmicos de Los Humeros, Puebla, México: Geotermia. *Revista Mexicana de Geoenergía*, 15(3), pp.159-170.
- Chilès, J.P., Aug, C., Guillen, A. and Lees, T., 2004, November. Modelling the geometry of geological units and its uncertainty in 3D from structural data: the potential-field method. In *Proceedings of international symposium on orebody modelling and strategic mine planning, Perth, Australia* (pp. 313-320).
- Contreras, E., Domínguez, B. and Rivera, M.O., 1990. Mediciones petrofísicas en núcleos de perforación del campo geotérmico Los Humeros: Geotermia. *Revista Mexicana de Geoenergía*, 6, pp.9-42.
- Emerick, A.A. and Reynolds, A.C., 2013 [a]. Investigation of the sampling performance of ensemble-based methods with a simple reservoir model. *Computational Geosciences*, 17(2), pp.325-350.
- Emerick, A.A. and Reynolds, A.C., 2013 [b]. Ensemble smoother with multiple data assimilation. *Computers & Geosciences*, 55, pp.3-15.
- Evanno, G., 2017. 3D Preliminary Geological Modelling of the Los Humeros Geothermal Area (Mexico). [Unpublished Final Dissertation of Master Degree] BRGM. ENAG/MFE-088-GB-2017.
- Evensen, G., 1994. Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. *Journal of Geophysical Research: Oceans*, 99(C5), pp.10143-10162.
- Ferriz, H. and Mahood, G.A., 1984. Eruption rates and compositional trends at Los Humeros volcanic center, Puebla, Mexico. *Journal of Geophysical Research: Solid Earth*, 89(B10), pp.8511-8524.
- García-Estrada, G.H., 1992. Study for the development of new interpretative criteria for gravity studies of geothermal areas- case study of Los Humeros, Pue., Mexico. *Transactions- Geothermal Resources Council*.
- Geophysics, I., 2015. GeoModeller.
- Giordano, G., Calzolari, G., Lucci, F., 2017. Regional structural and tectonic synthesis of the Acoculco and Los Humeros geothermal fields. *GEMex Project*. [Unpublished paper] Università Roma Tre.
- Glegola, M.A., 2013. Gravity observations for hydrocarbon reservoir monitoring. [Unpublished Final Dissertation of PhD Degree] Technical University Delft.

- Guillen, A., Calcagno, P., Courrioux, G., Joly, A. and Ledru, P., 2008. Geological modelling from field data and geological knowledge: Part II. Modelling validation using gravity and magnetic data inversion. *Physics of the Earth and Planetary Interiors*, 171(1-4), pp.158-169.
- Guillen, A., Courrioux, G., Calcagno, P., Lane, R., Lees, T. and McNerney, P., 2004. Constrained gravity 3D litho-inversion applied to Broken Hill. *ASEG Extended Abstracts*, 2004(1), pp.1-6.
- Hantschel, T. and Kauerauf, A.I., 2009. *Fundamentals of basin and petroleum systems modeling*. Springer Science & Business Media.
- Heagy, L.J., Cockett, R. and Oldenburg, D., 2016, December. GeoSci: Practices to Collaboratively Build Online Resources for Geophysics Education. In *AGU Fall Meeting Abstracts*.
- Holstein, H., 2003. Gravimagnetic anomaly formulas for polyhedra of spatially linear media. *Geophysics*, 68(1), pp.157-167.
- Husson, E., Guillen, A., Séranne, M., Courrioux, G. and Couëffé, R., 2018. 3D Geological modelling and gravity inversion of a structurally complex carbonate area: application for karstified massif localization. *Basin Research*.
- Lajaunie, C., Courrioux, G. and Manuel, L., 1997. Foliation fields and 3D cartography in geology: principles of a method based on potential interpolation. *Mathematical Geology*, 29(4), pp.571-584.
- López-Hernández, A., 1995. Estudio regional volcánico y estructural del campo geotérmico de Los Humeros, Pue., Mexico. Geotermia. *Revista Mexicana de Geoenergía*, pp.11-1.
- Norini, G., Groppelli, G., Sulpizio, R., Carrasco-Núñez, G., Dávila-Harris, P., Pellicoli, C., Zucca, F. and De Franco, R., 2015. Structural analysis and thermal remote sensing of the Los Humeros Volcanic Complex: Implications for volcano structure and geothermal exploration. *Journal of Volcanology and Geothermal Research*, 301, pp.221-237.
- Okabe, M., 1979. Analytical expressions for gravity anomalies due to homogeneous polyhedral bodies and translations into magnetic anomalies. *Geophysics*, 44(4), pp.730-741.
- Oliver, D.S., Reynolds, A.C. and Liu, N., 2008. *Inverse theory for petroleum reservoir characterization and history matching*. Cambridge University Press.
- Parker, R.L., 1973. The rapid calculation of potential anomalies. *Geophysical Journal of the Royal Astronomical Society*, 31(4), pp.447-455.
- Peiffer, L., Carrasco-Núñez, G., Mazot, A., Villanueva-Estrada, R.E., Inguaggiato, C., Romero, R.B., Miller, R.R. and Rojas, J.H., 2018. Soil degassing at the Los Humeros geothermal field (Mexico). *Journal of Volcanology and Geothermal Research*, 356, pp.163-174.
- Plouff, D., 1976. Gravity and magnetic fields of polygonal prisms and application to magnetic terrain corrections. *Geophysics*, 41(4), pp.727-741.
- Quezadas-Flores, A., 1961. Las rocas del basamento de la cuenca Tampico-Mizantla: México, DF, Bachelor Thesis. *Fac. Ing., UNAM*, p.72.
- Rodríguez, F.C., 2000. Hydrogeologic Model of the Geothermal Reservoir from Los Humeros, Puebla Mexico. *TRANSACTIONS-GEOTHERMAL RESOURCES COUNCIL*, pp.433-436.
- Struijk, M., 2016. An improved method for modeling the subsurface temperature of the onshore Netherlands. [Unpublished report] Netherlands Organisation for Applied Scientific Research (TNO).
- Van Wees, J.D., 2016. IMAGE-D6.01: Catalogues and European models. *IMAGE project*. [Confidential report] Netherlands Organisation for Applied Scientific Research (TNO).
- Verma, S.P. and López, M., 1982. Geochemistry of Los Humeros caldera, Puebla, Mexico. *Bulletin Volcanologique*, 45(1), pp.63-79.
- Verma, S.P., 1985. On the magma chamber characteristics as inferred from surface geology and geochemistry: examples from Mexican geothermal areas. *Physics of the Earth and Planetary Interiors*, 41(2-3), pp.207-214.

REFERENCES

- Verma, M.P., Verma, S.P. and Sanvicente, H., 1990. Temperature field simulation with stratification model of magma chamber under Los Humeros caldera, Puebla, Mexico. *Geothermics*, 19(2), pp.187-197.
- Verma, S.P., 2000. Geochemical evidence for a lithospheric source for magmas from Los Humeros caldera, Puebla, Mexico. *Chemical Geology*, 164(1), pp.35-60.
- Verma, S.P., Gómez-Arias, E. and Andaverde, J., 2011. Thermal sensitivity analysis of emplacement of the magma chamber in Los Humeros caldera, Puebla, Mexico. *International Geology Review*, 53(8), pp.905-925.
- Verma, S.P. and Gómez-Arias, E., 2013. Three-dimensional temperature field simulation of magma chamber in the Los Humeros geothermal field, Puebla, Mexico. *Applied Thermal Engineering*, 52(2), pp.512-515.
- Viniegra, F., 1965. Geología del macizo de Teziutlán y la cuenca Cenozoica de Veracruz. *Boletín de la Asociación Mexicana de Geología y Petrología*, 17, pp.7-12.
- Willmott, C.J. and Matsuura, K., 2005. Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance. *Climate research*, 30(1), pp.79-82.

APPENDIX I: ADDITIONAL BENCHMARKING FIGURES AND TABLES

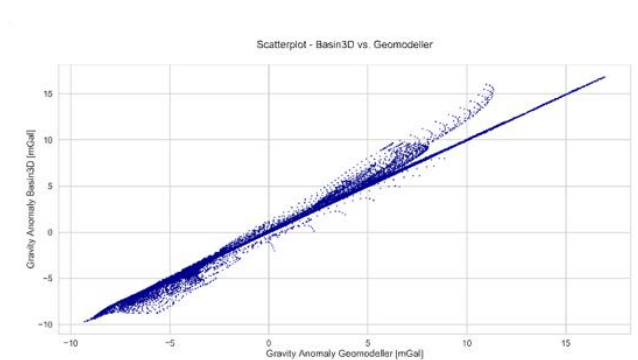
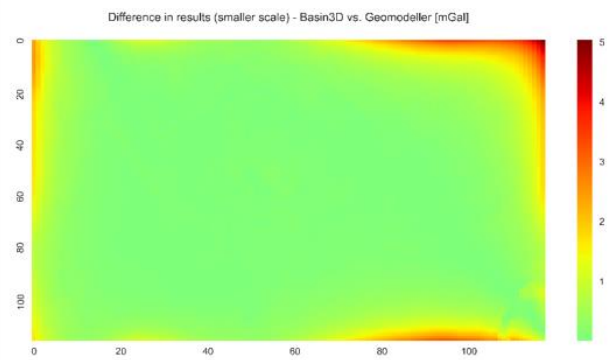
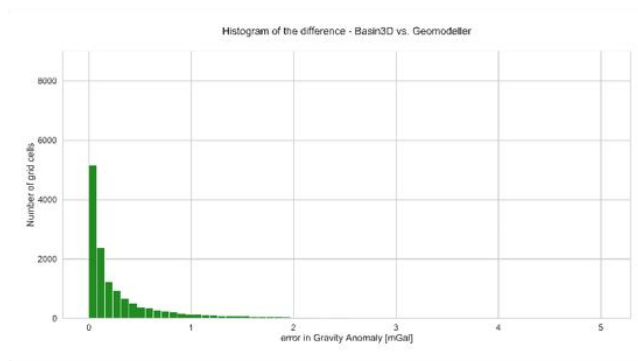
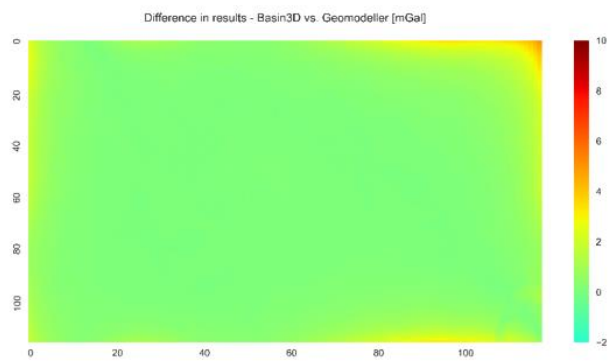
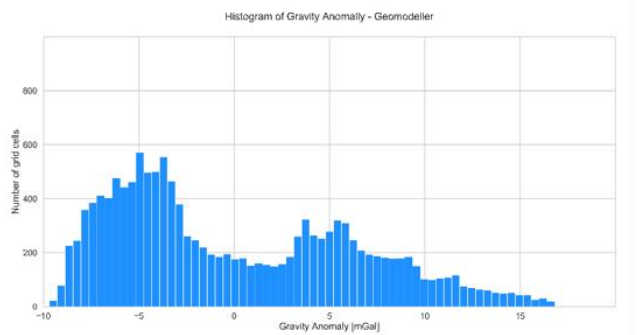
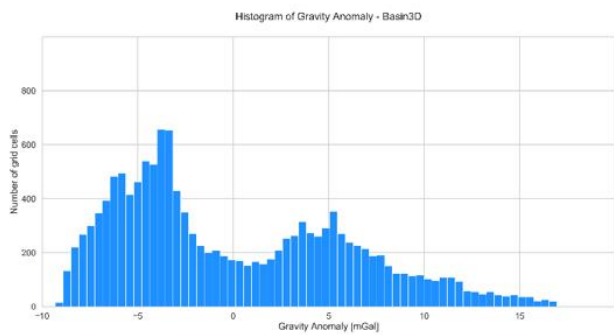
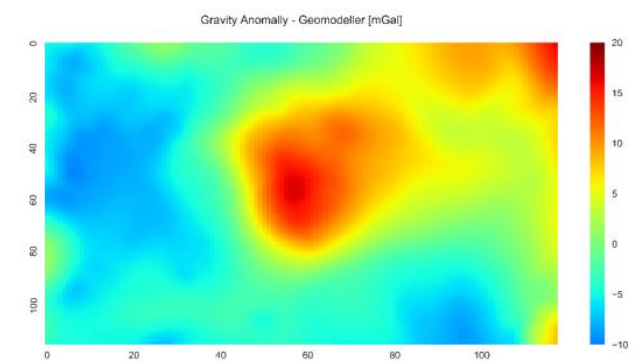
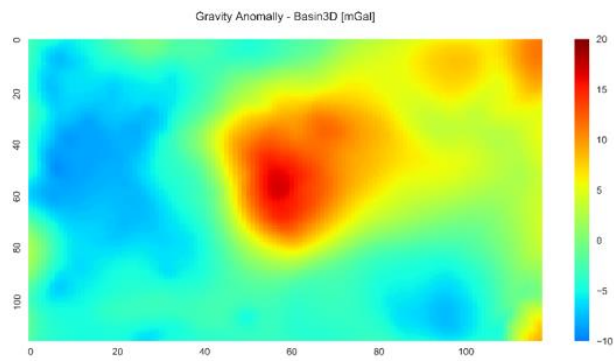
In this appendix we include the additional figures, maps, plots and tables used to assess the results of the benchmarking of Basin3D versus Geomodeller.

Edge cut-off Table

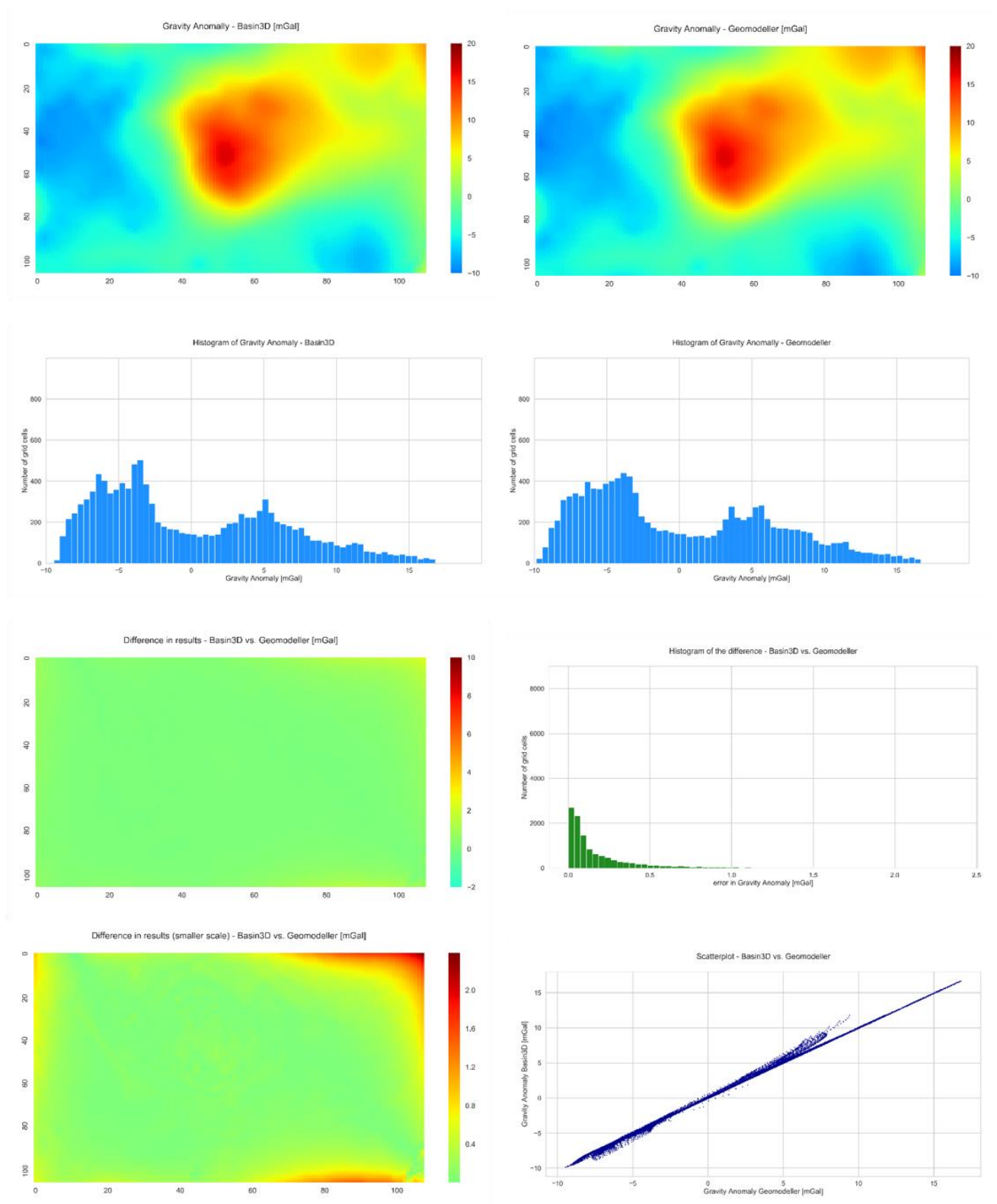
Comparison of Basin3D Model vs. Geomodeller Model with different edge thicknesses removed											
Model	Depth [m]	FFT [%]	Cells Removed	Distance Removed [km]		Percentage of Depth Removed [%]		PCC	MAE [mGal]	RMSE [mGal]	RMSE to MAE Ratio
				x-dimension	y-dimension	x-dimension	y-dimension				
LH	3312.5	100	0	0.000	0.000	0.00	0.00	0.972274	1.53477	0.74138	2.07015
LH	3312.5	100	5	2.188	1.406	19.34	12.43	0.995806	0.63011	0.34686	1.81662
LH	3312.5	100	10	4.375	2.813	38.67	24.86	0.998869	0.32889	0.19539	1.68321
LH	3312.5	100	15	6.563	4.219	58.01	37.29	0.999592	0.19605	0.12195	1.60764
LH	3312.5	100	20	8.750	5.625	77.35	49.72	0.999830	0.12565	0.08129	1.54569
LH	3312.5	100	25	10.938	7.031	96.69	62.15	0.999919	0.08610	0.05842	1.47391
LH	3312.5	100	30	13.125	8.438	116.02	74.59	0.999956	0.06301	0.04604	1.36861
LH	3312.5	100	35	15.313	9.844	135.36	87.02	0.999972	0.04840	0.03801	1.27343
LH	3312.5	100	40	17.500	11.250	154.70	99.45	0.999977	0.04302	0.03435	1.25234

Table 1. Different edge cut-off by number of cells. Basin3D model versus Geomodeller model. FFT stands for Fast Fourier Transform expansion and is expressed in the percent of the model size below $z=0$. PCC is the Pearson's Correlation Coefficient, MAE is the Mean Absolute Error [mGal] and RMSE stands for Root Mean Square Error [mGal]. RMSE/MAE ratio is a measure of error distribution, value of 1 (minima) implies a uniform error distribution.

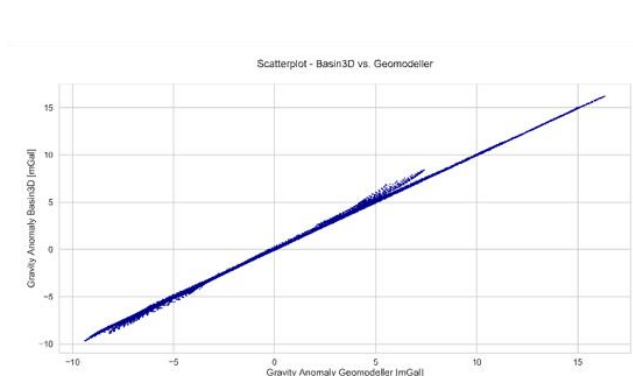
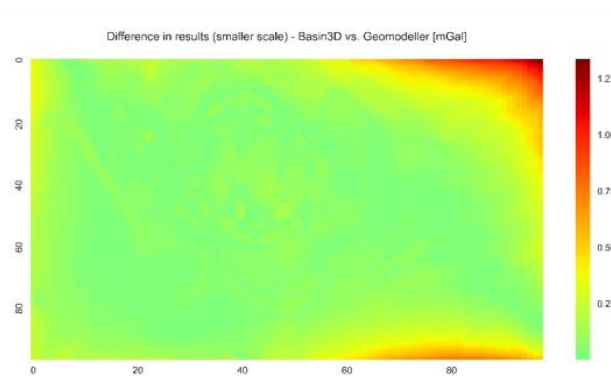
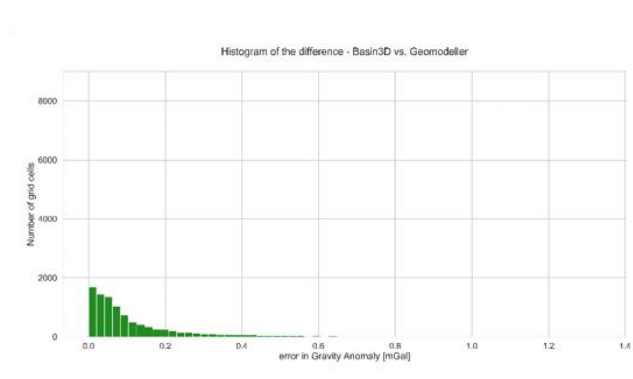
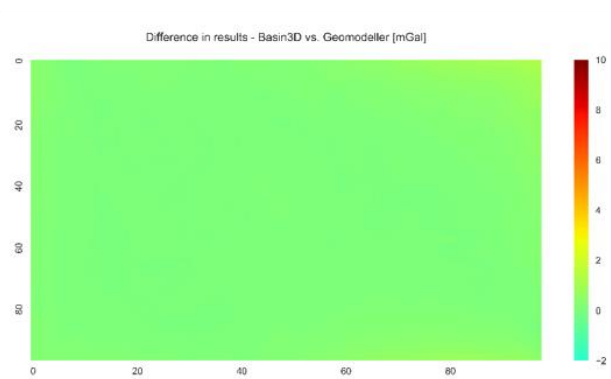
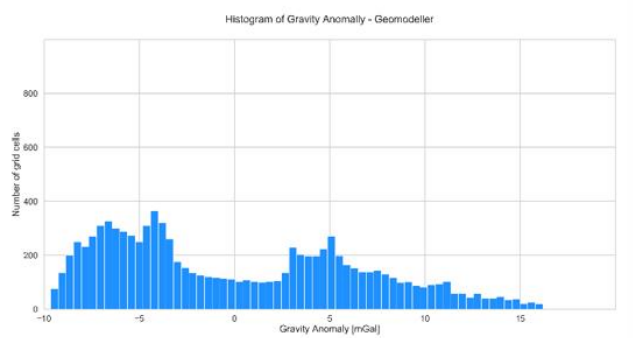
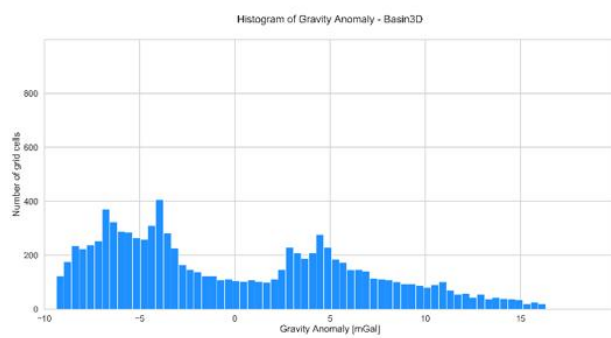
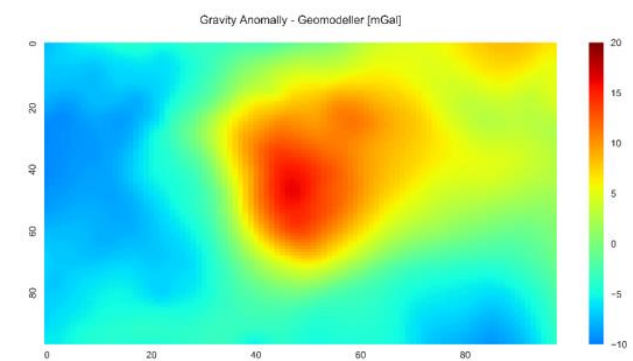
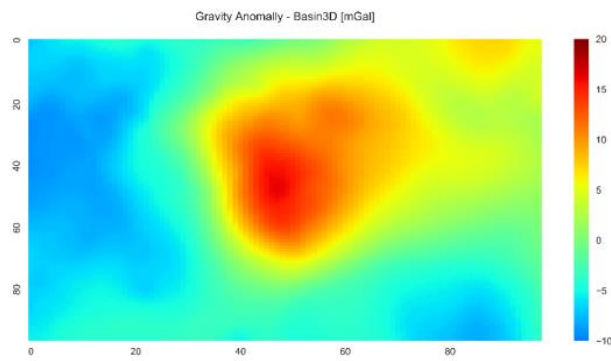
Edge cut-off: 5 cells



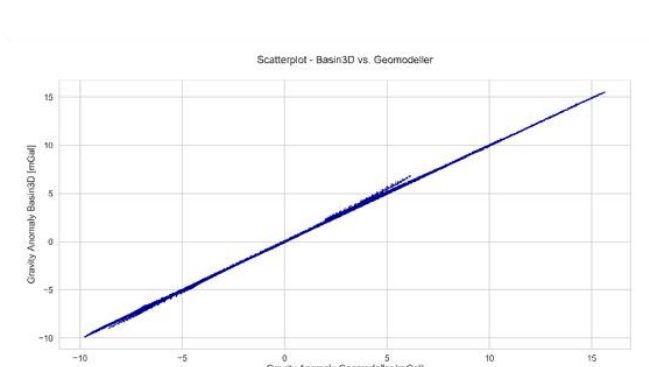
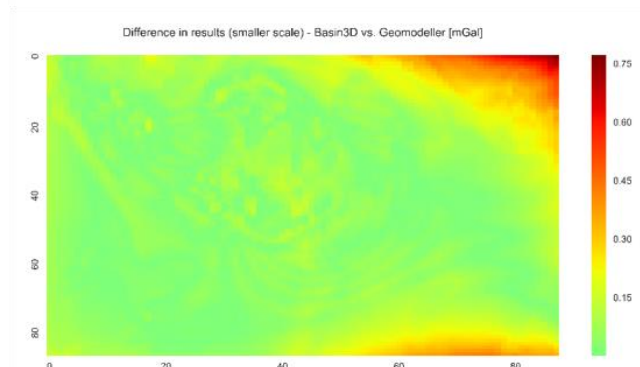
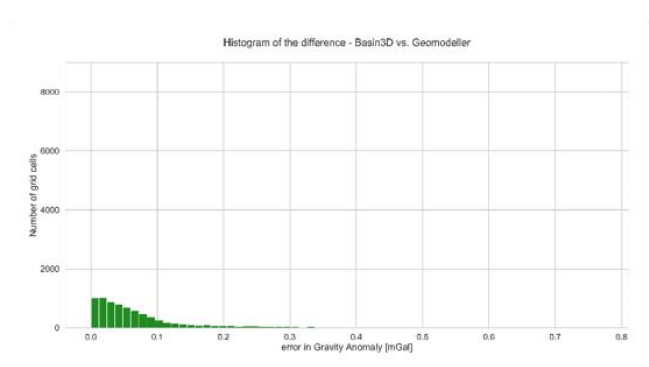
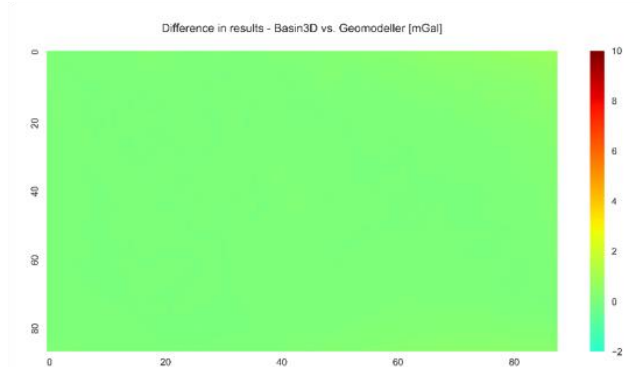
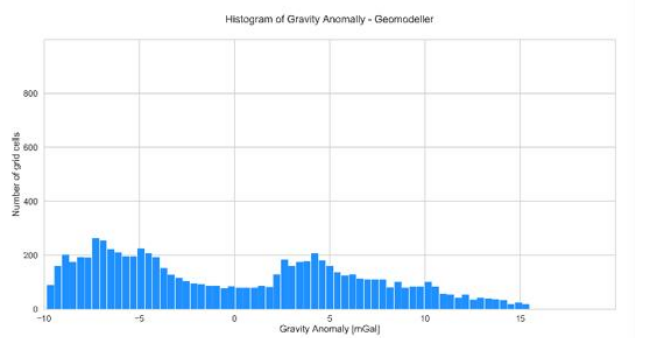
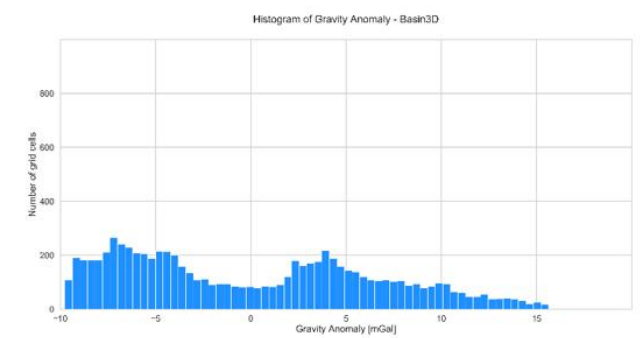
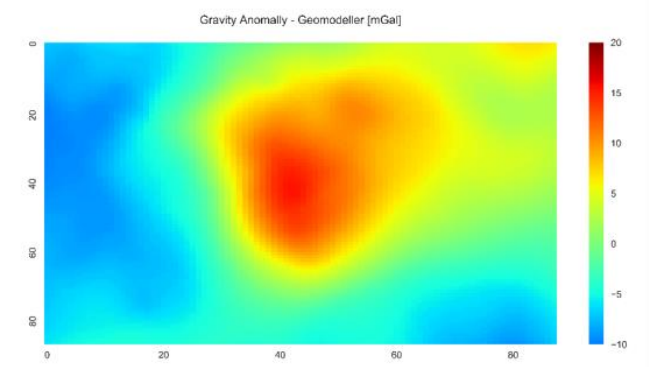
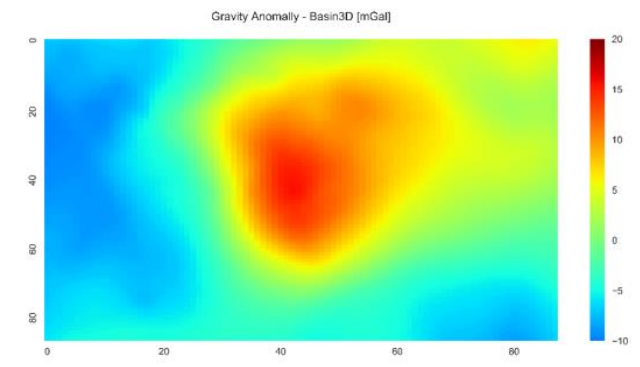
Edge cut-off: 10 cells



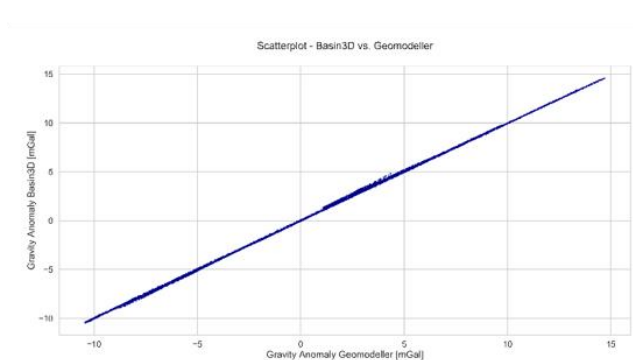
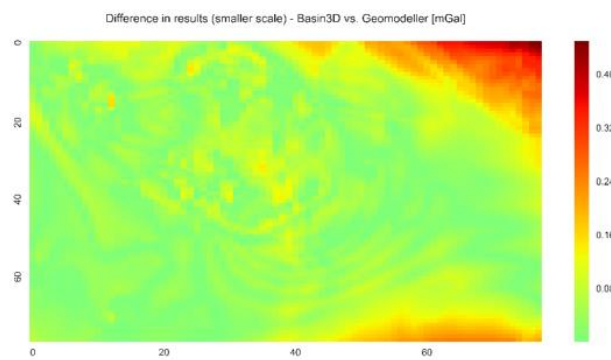
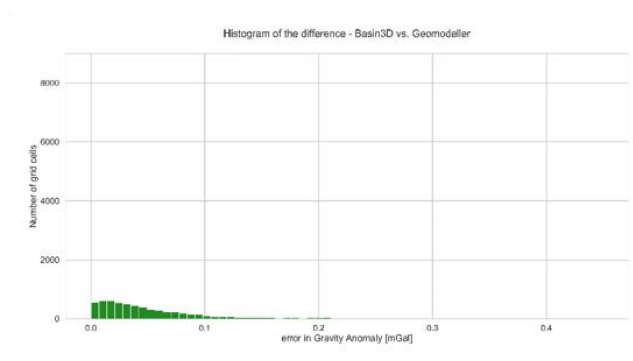
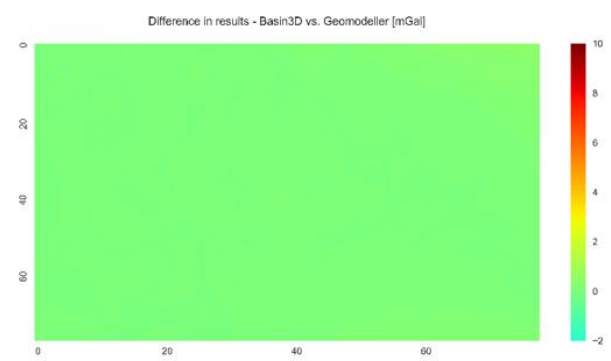
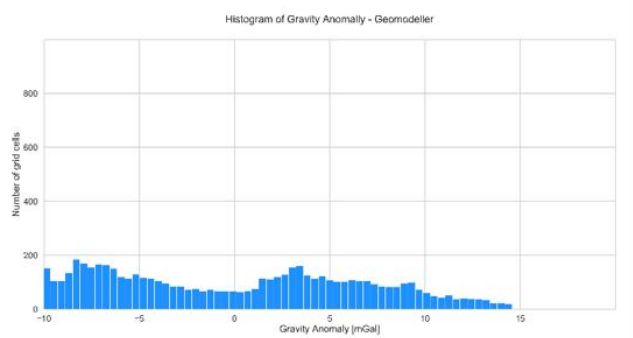
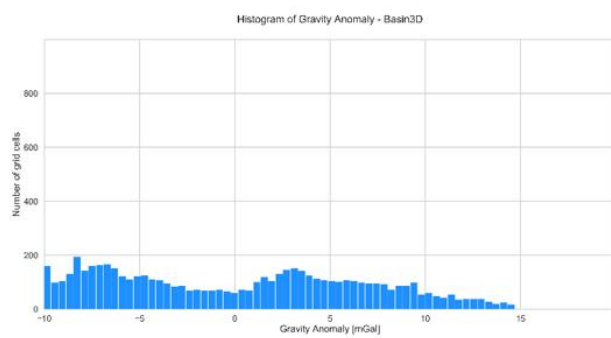
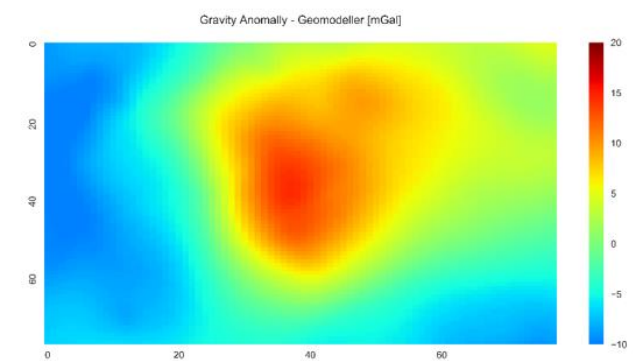
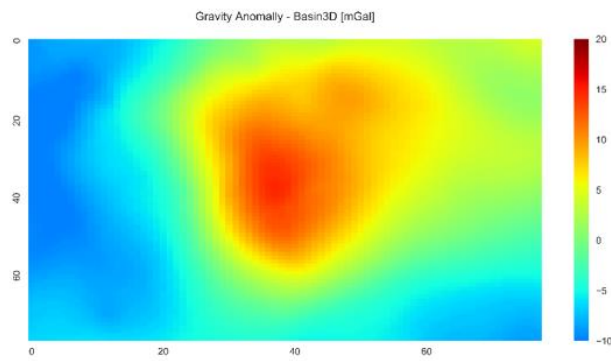
Edge cut-off: 15 cells



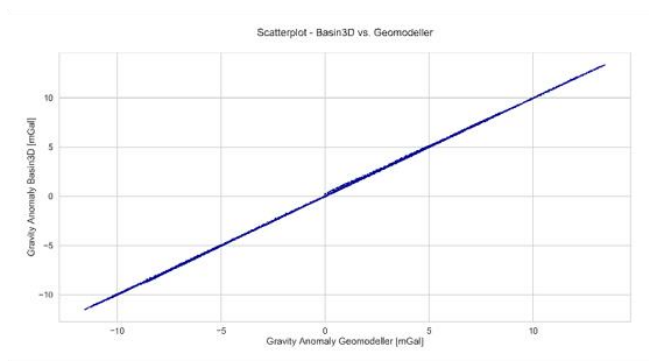
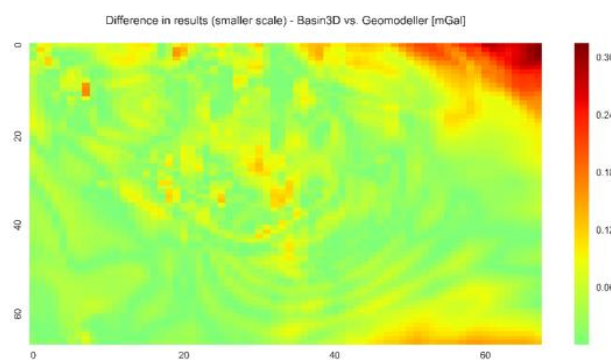
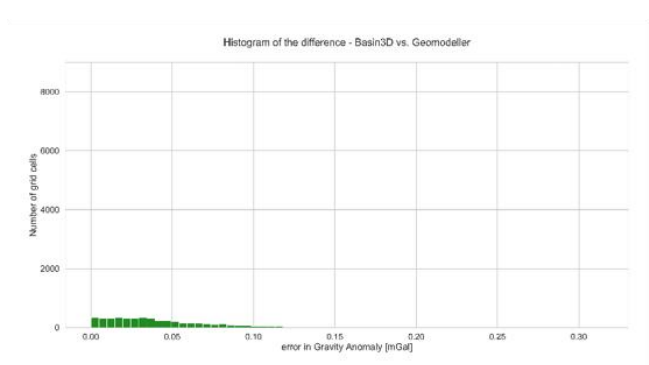
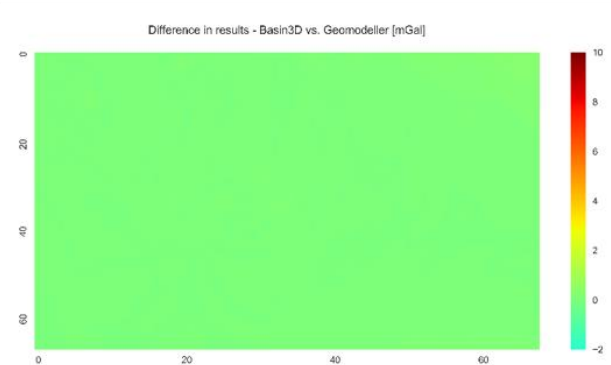
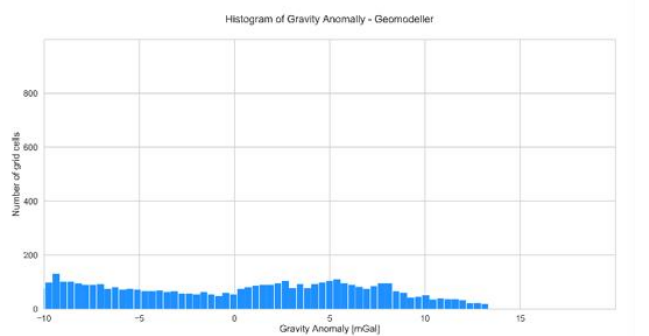
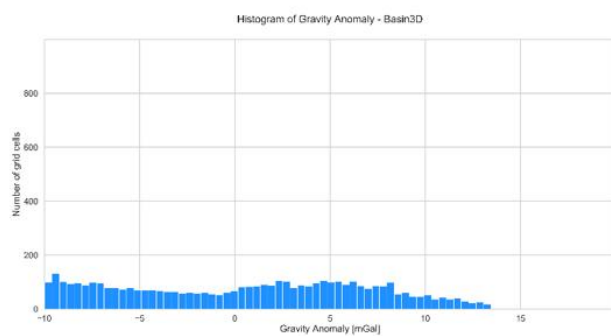
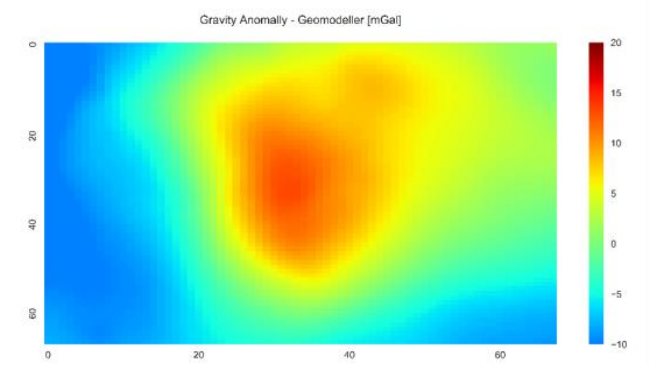
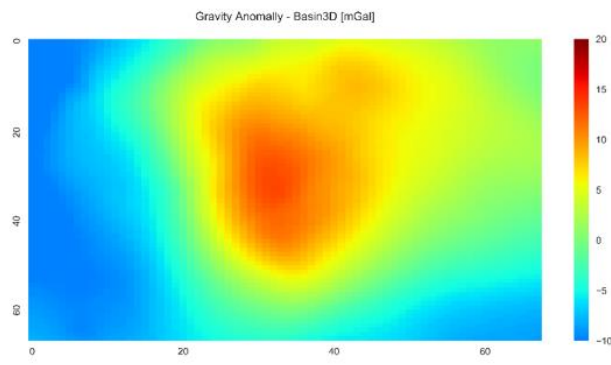
Edge cut-off: 20 cells



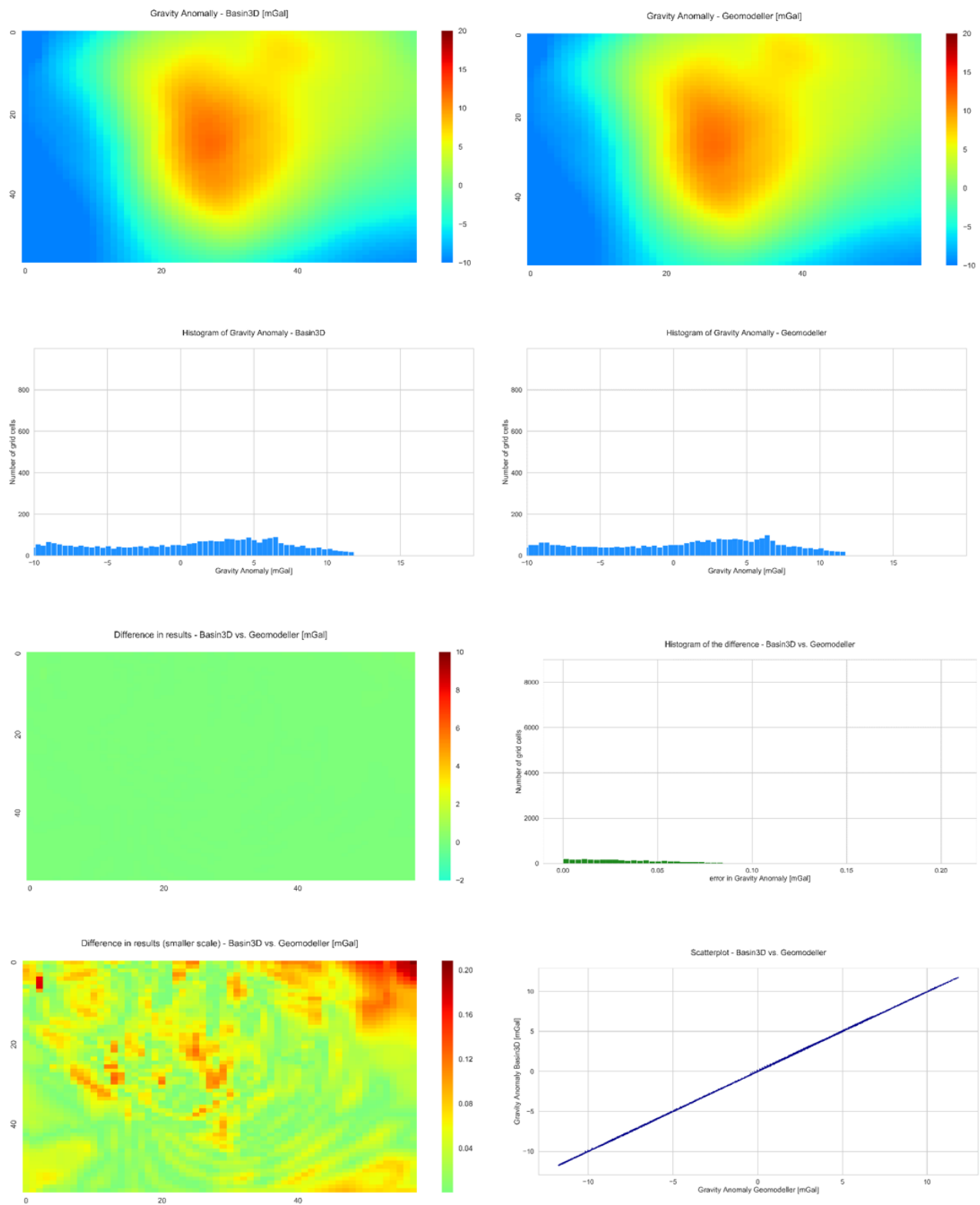
Edge cut-off: 25 cells



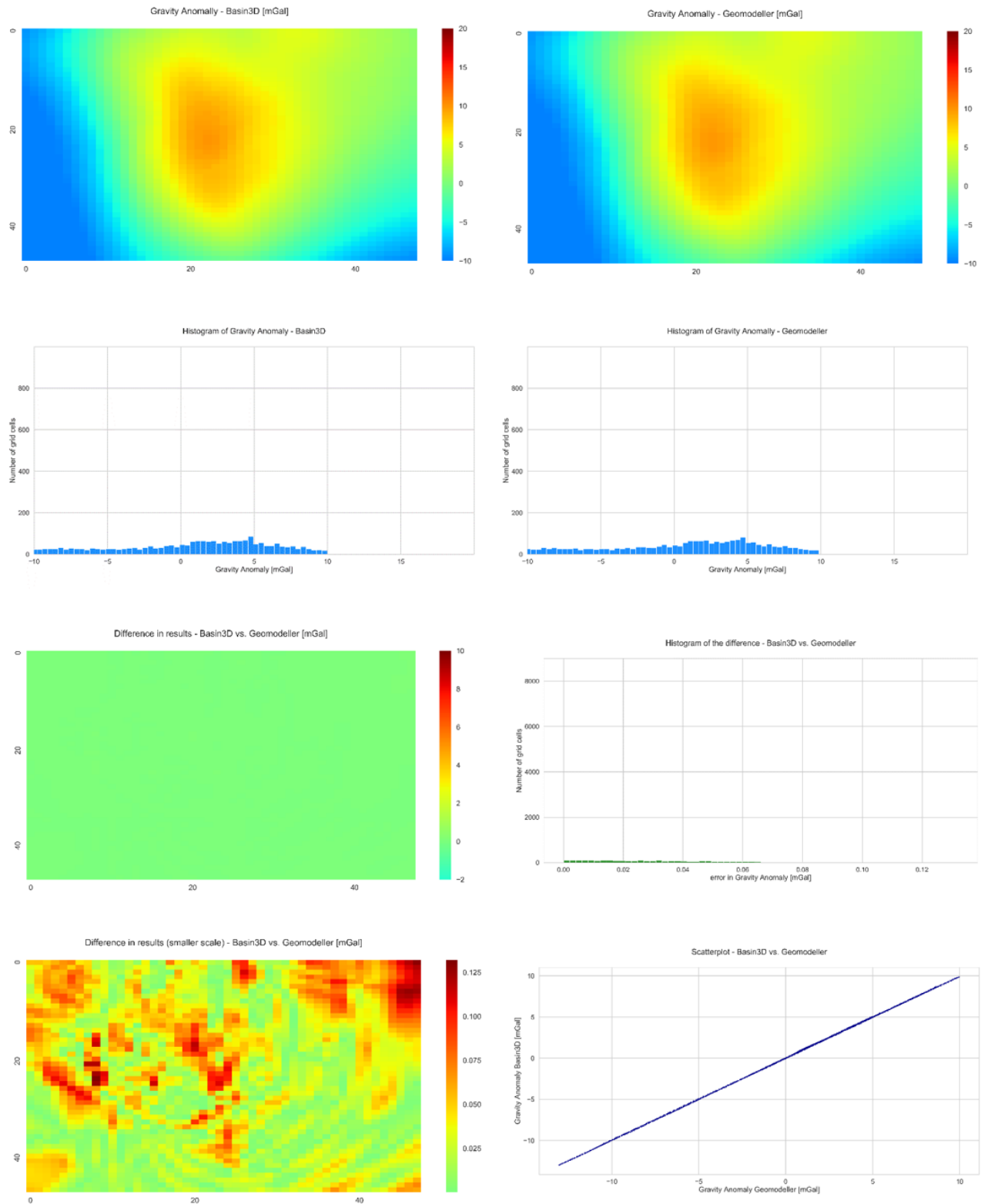
Edge cut-off: 30 cells



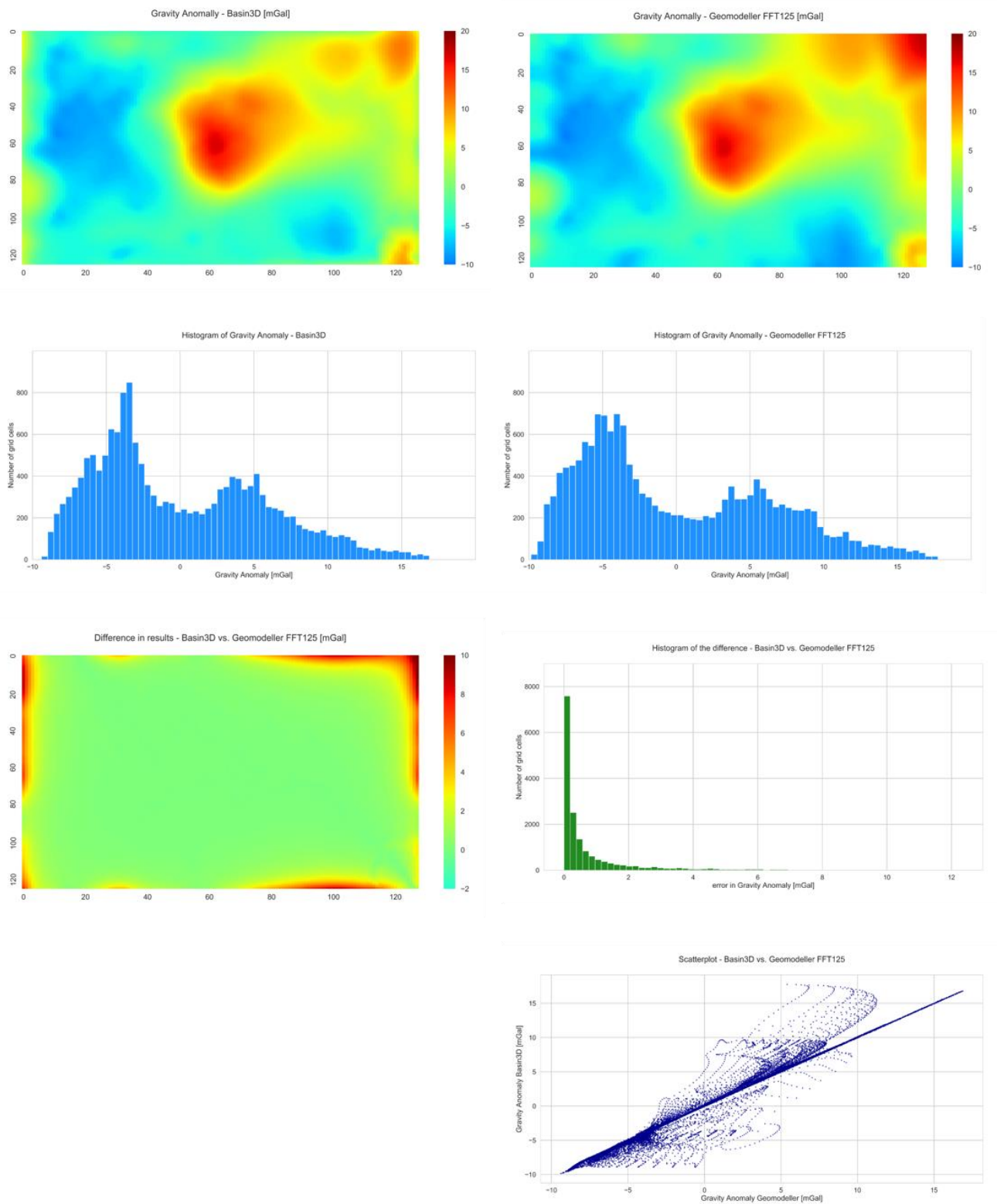
Edge cut-off: 35 cells



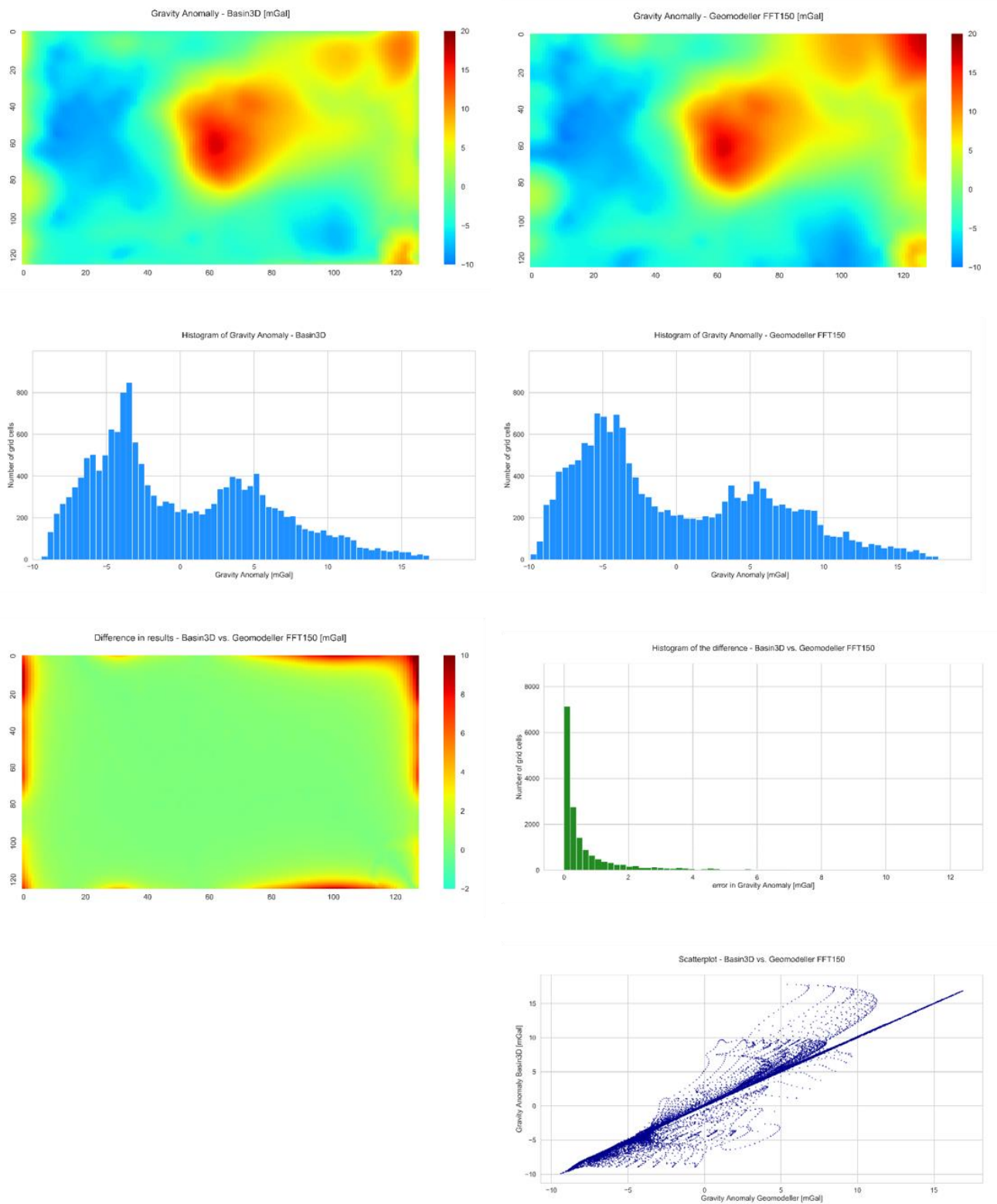
Edge cut-off: 40 cells



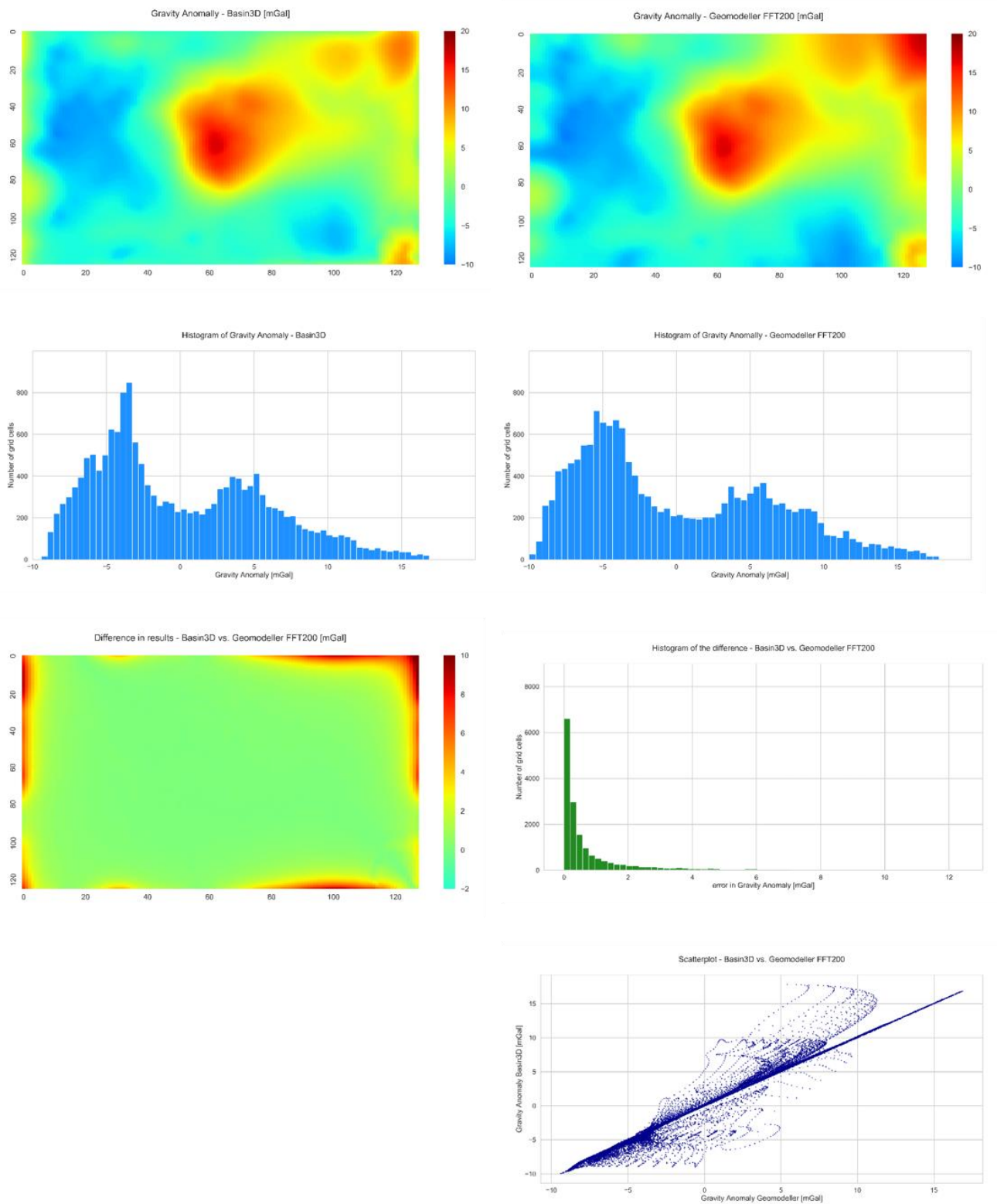
FFT expansion change in Geomodeller vs. Basin3D at 100%: 125%



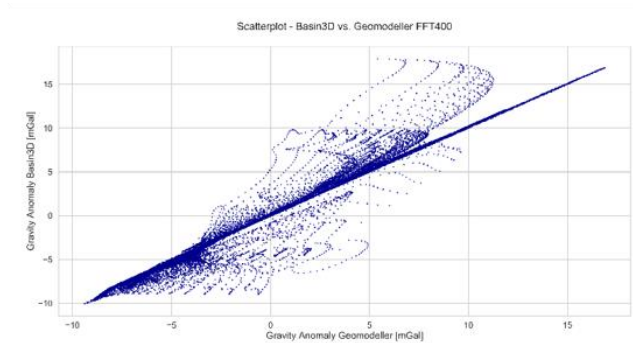
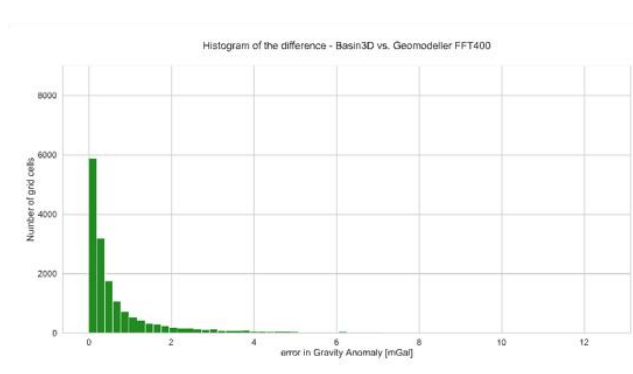
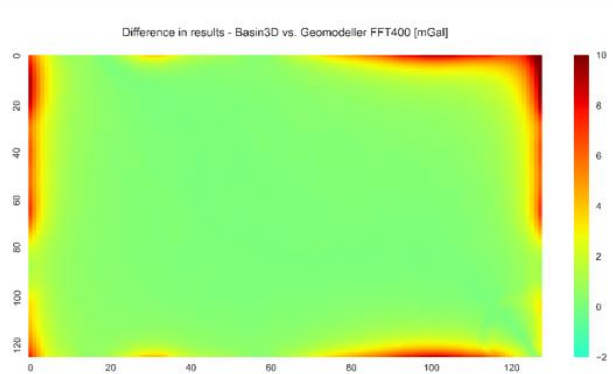
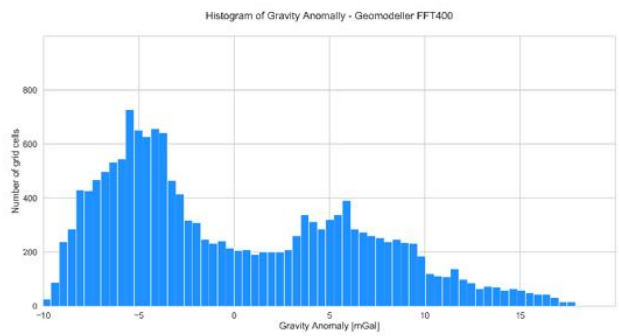
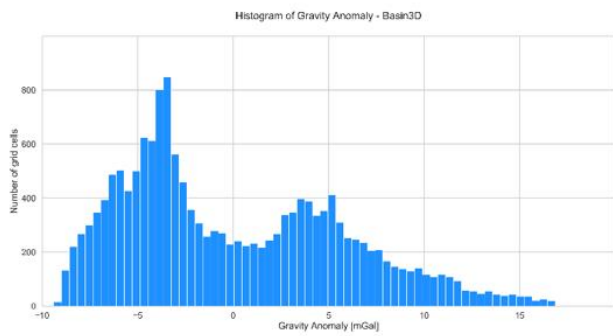
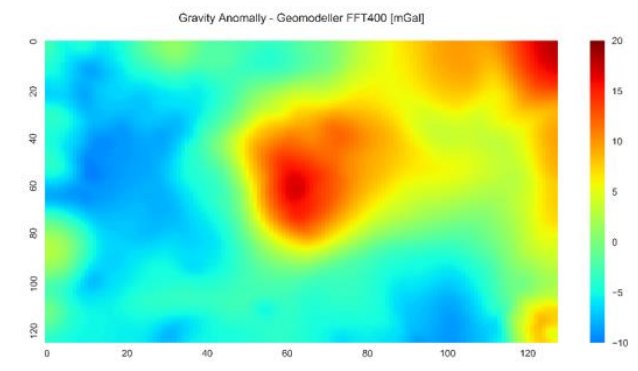
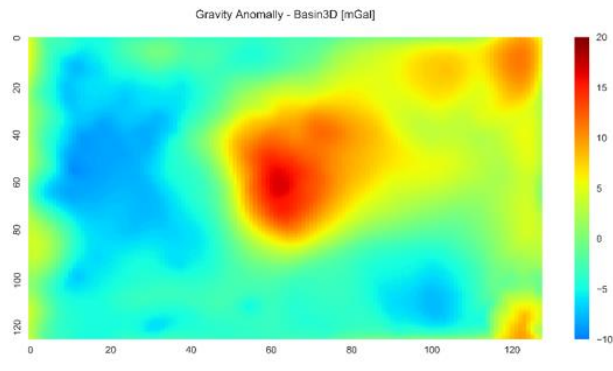
FFT expansion change in Geomodeller vs. Basin3D at 100%: 150%



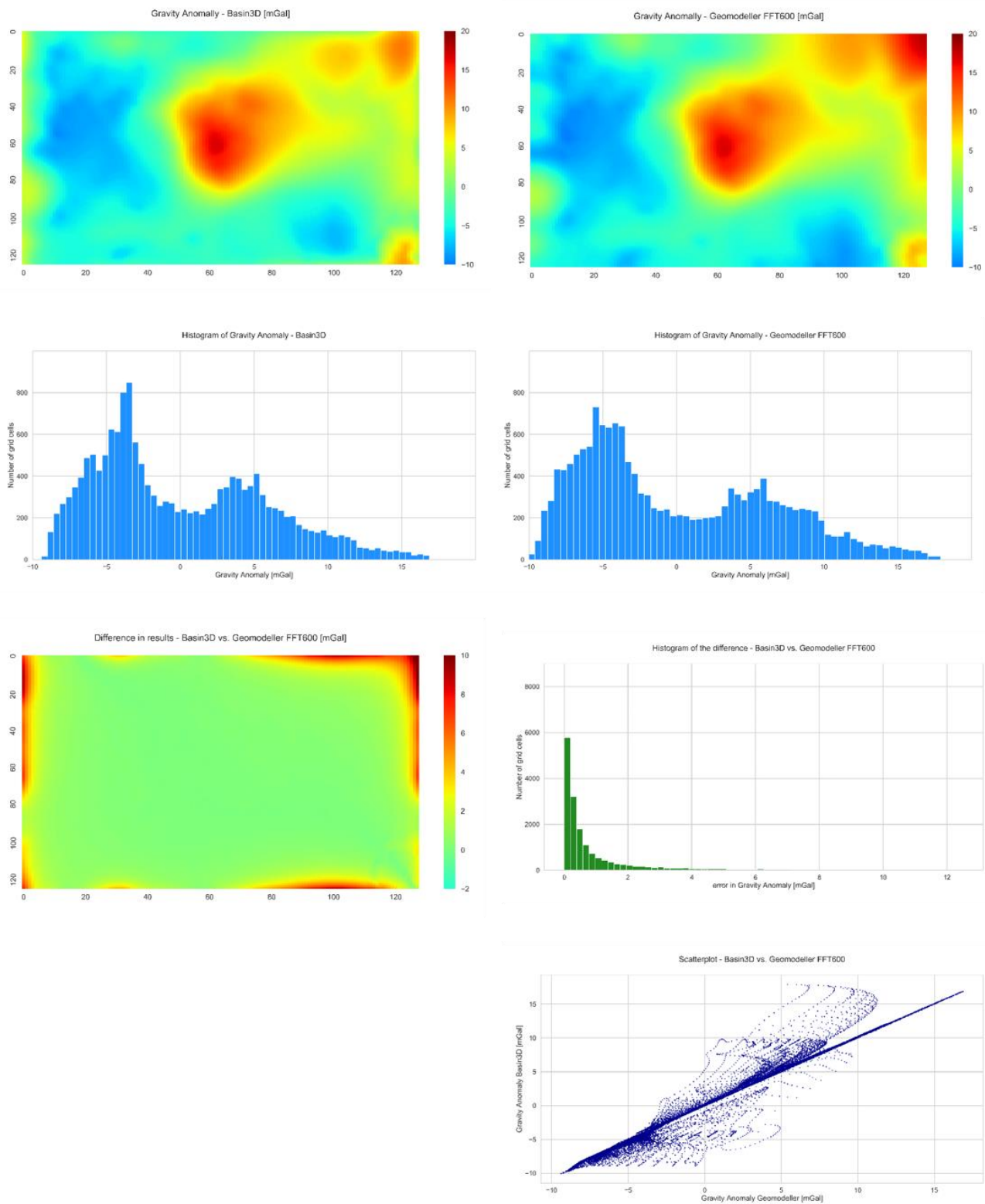
FFT expansion change in Geomodeller vs. Basin3D at 100%: 200%



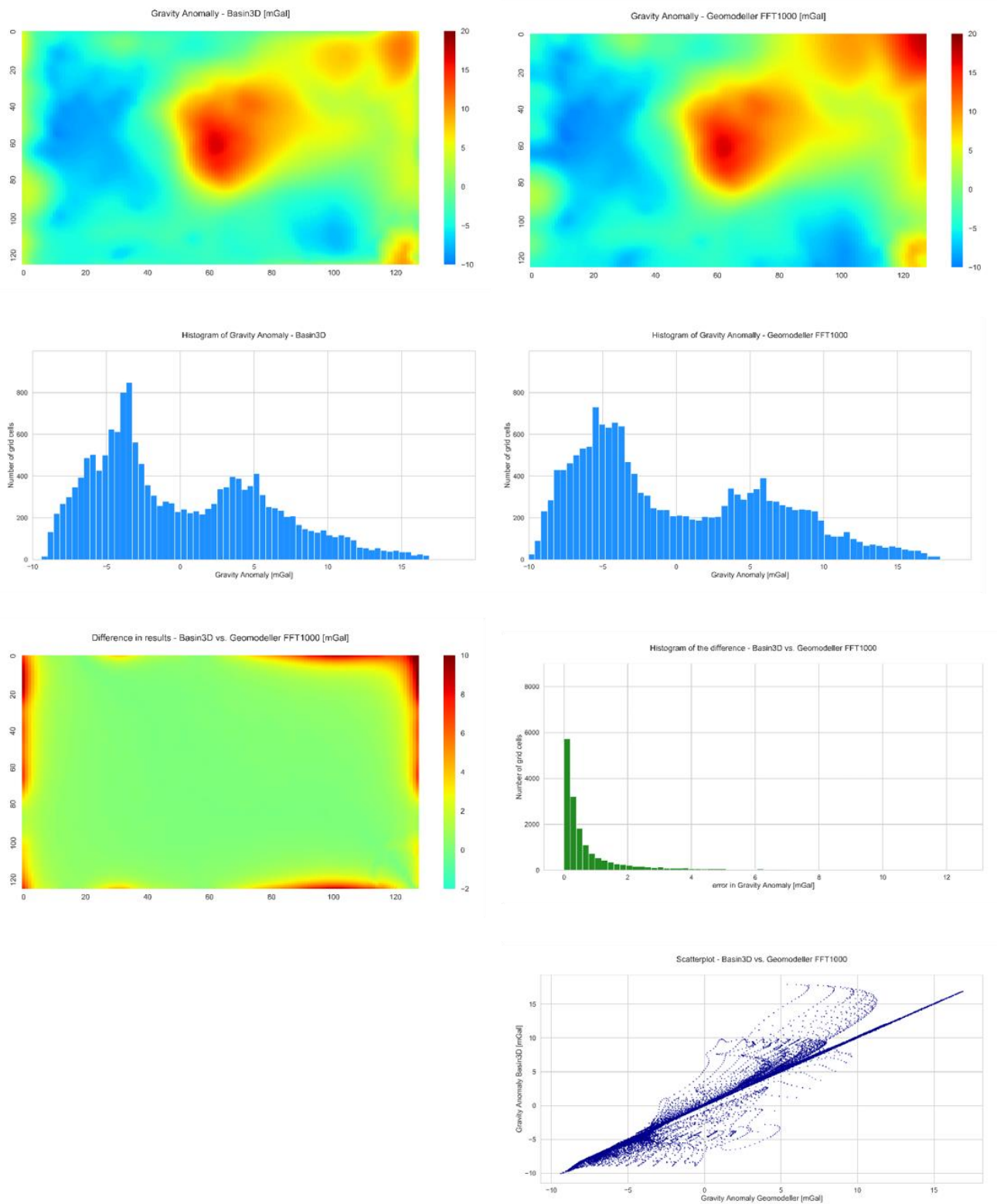
FFT expansion change in Geomodeller vs. Basin3D at 100%: 400%



FFT expansion change in Geomodeller vs. Basin3D at 100%: 600%



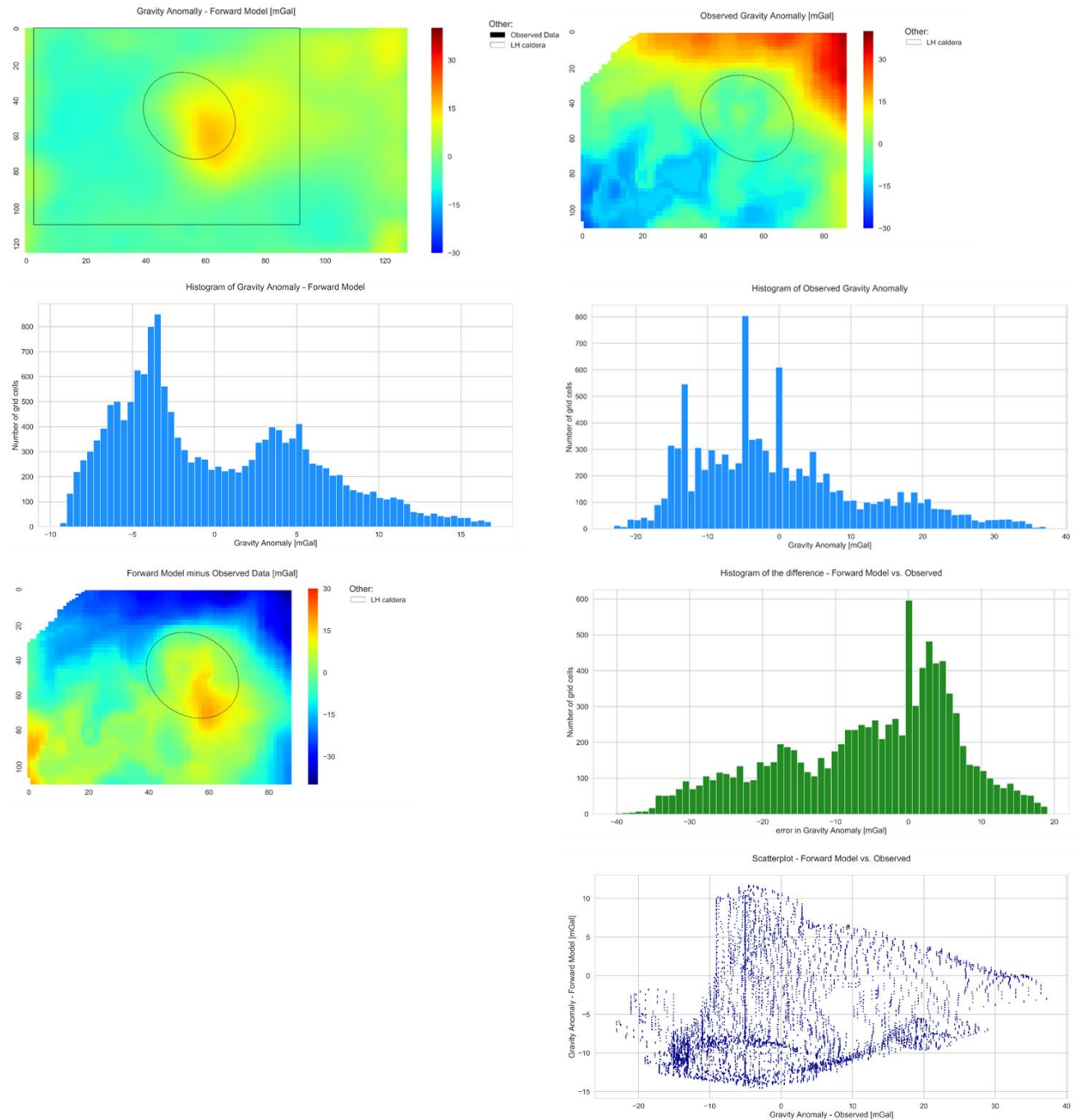
FFT expansion change in Geomodeller vs. Basin3D at 100%: 1000%



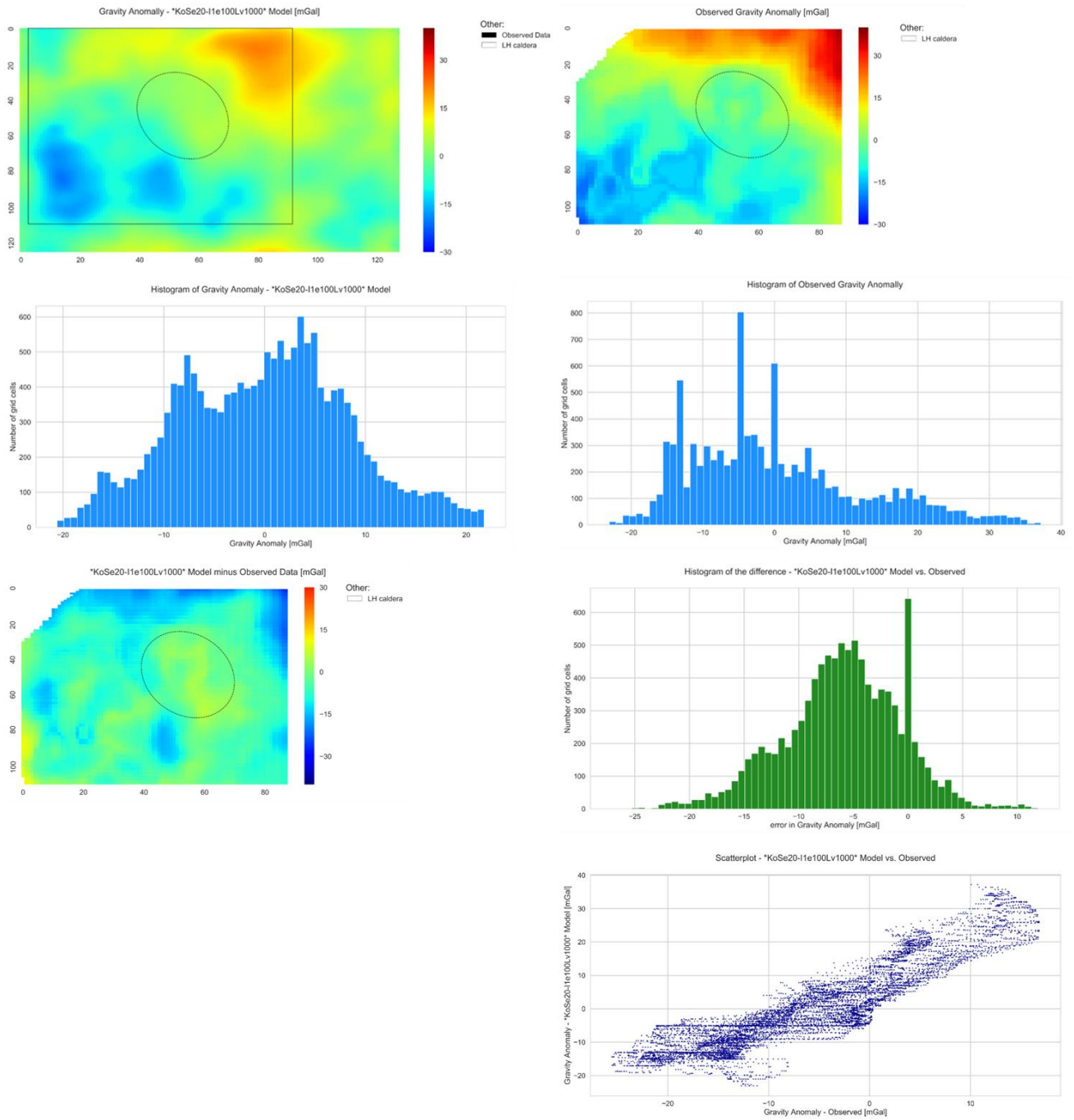
APPENDIX II: ADDITIONAL MODELLING FIGURES AND TABLES

In this appendix we include the additional figures, maps, plots and tables used to assess the results of the inverse modelling of the Los Humeros geothermal complex.

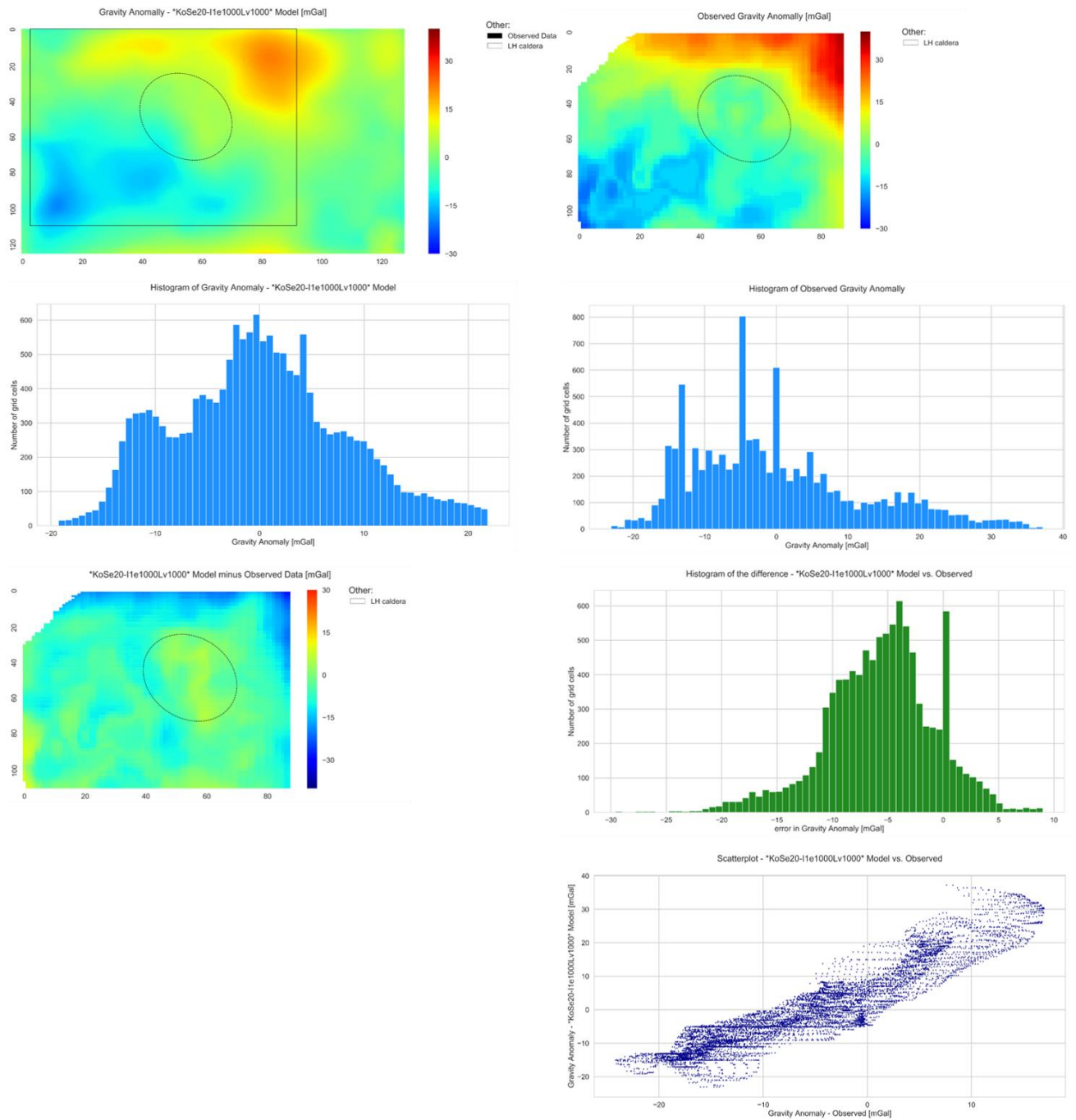
Forward vs. Observed gravity anomaly



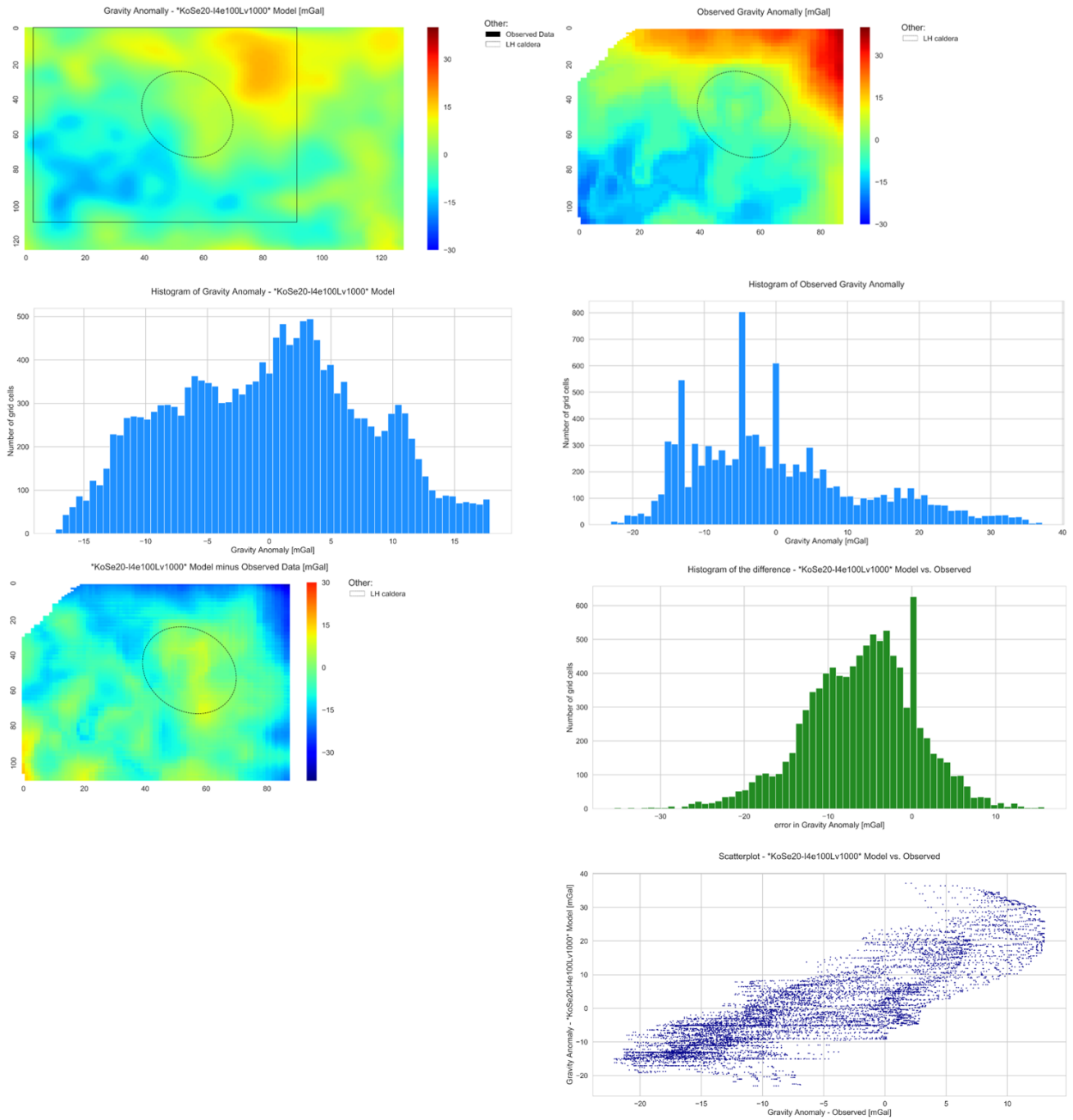
Inverse vs. Observed gravity anomaly: 1 iteration 100 ensembles



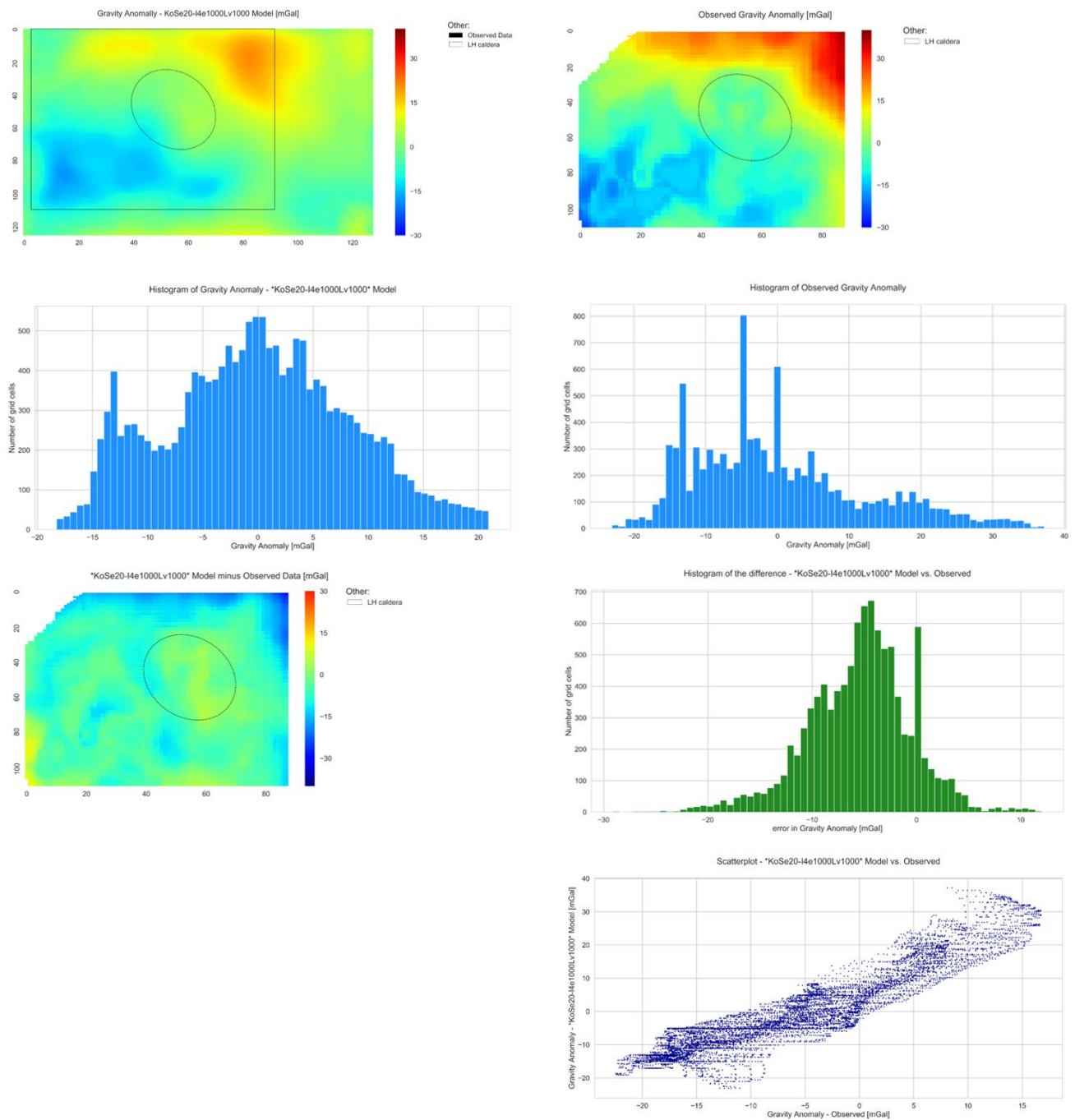
Inverse vs. Observed gravity anomaly: 1 iteration 1000 ensembles



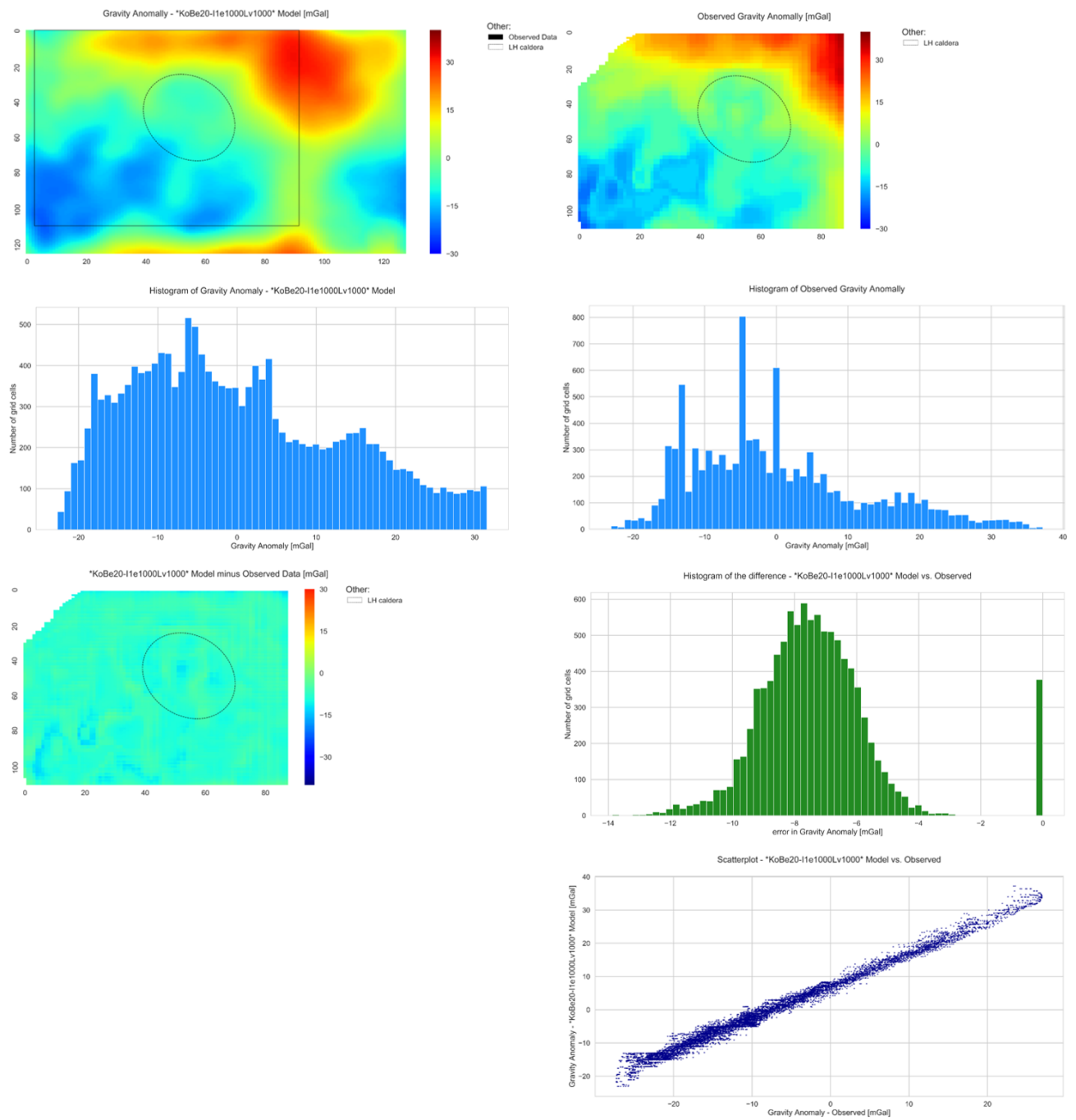
Inverse vs. Observed gravity anomaly: 4 iterations 100 ensembles



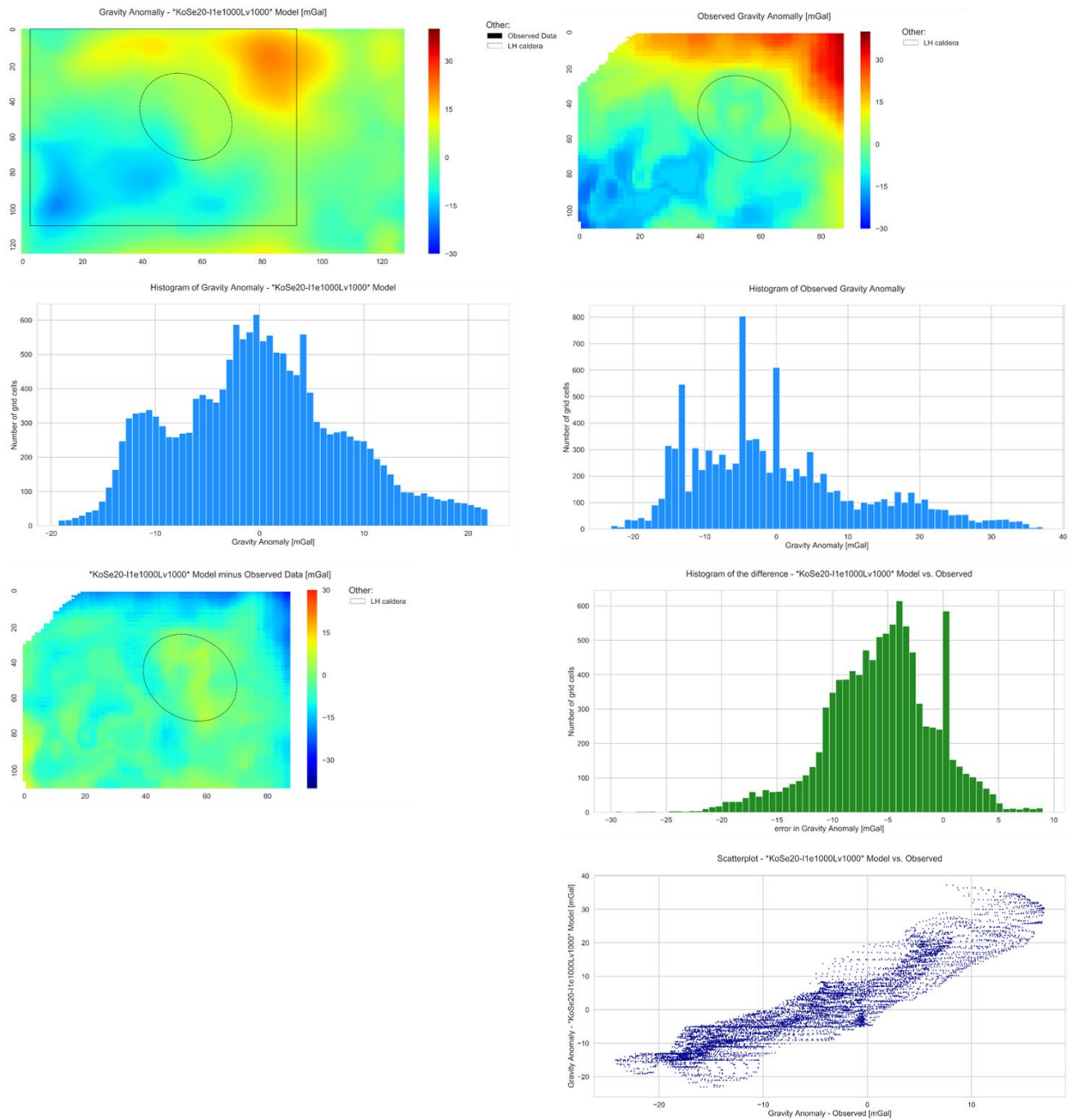
Inverse vs. Observed gravity anomaly: 4 iterations 1000 ensembles



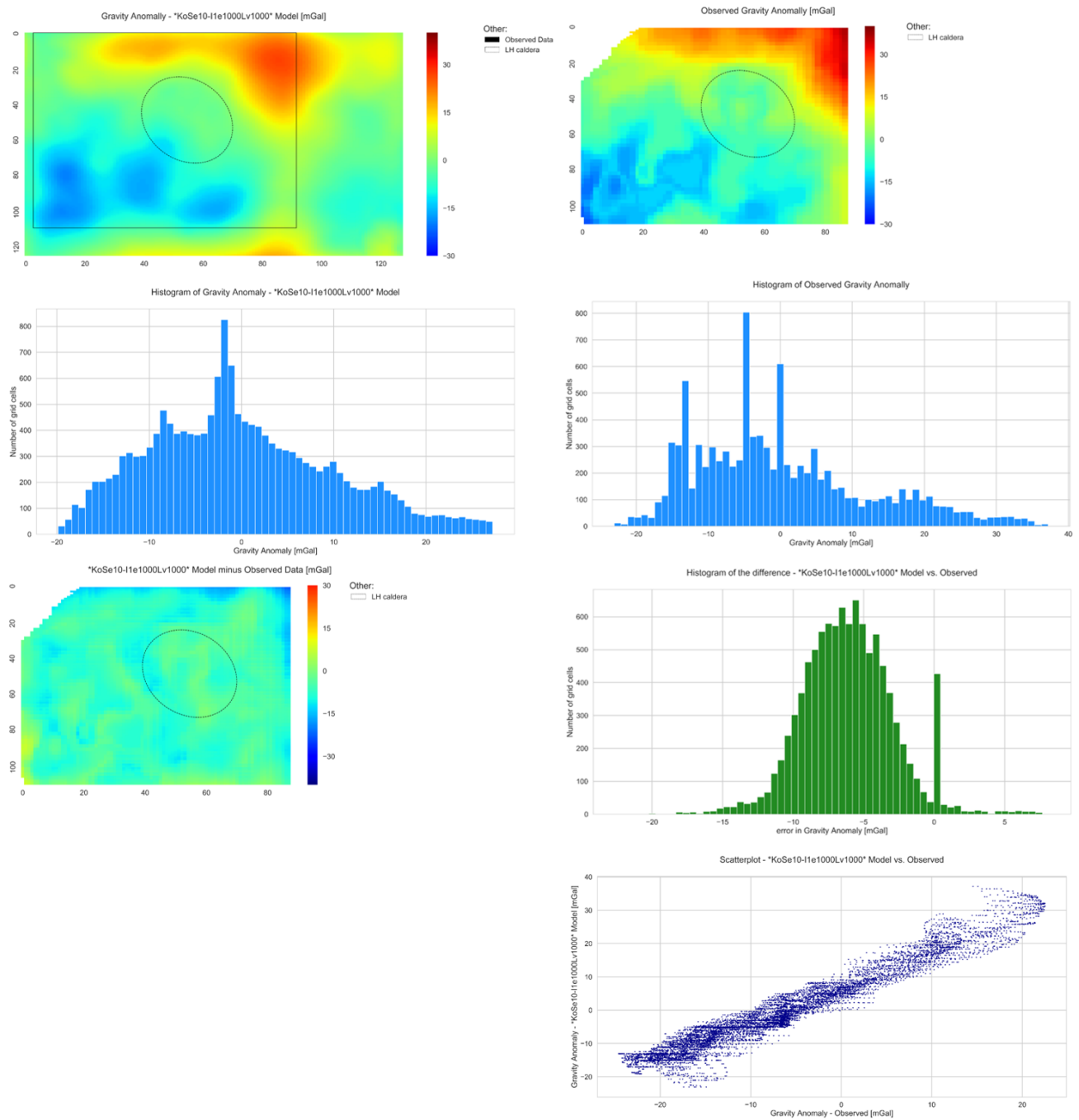
Inverse vs. Observed gravity anomaly: 9657 observed data points



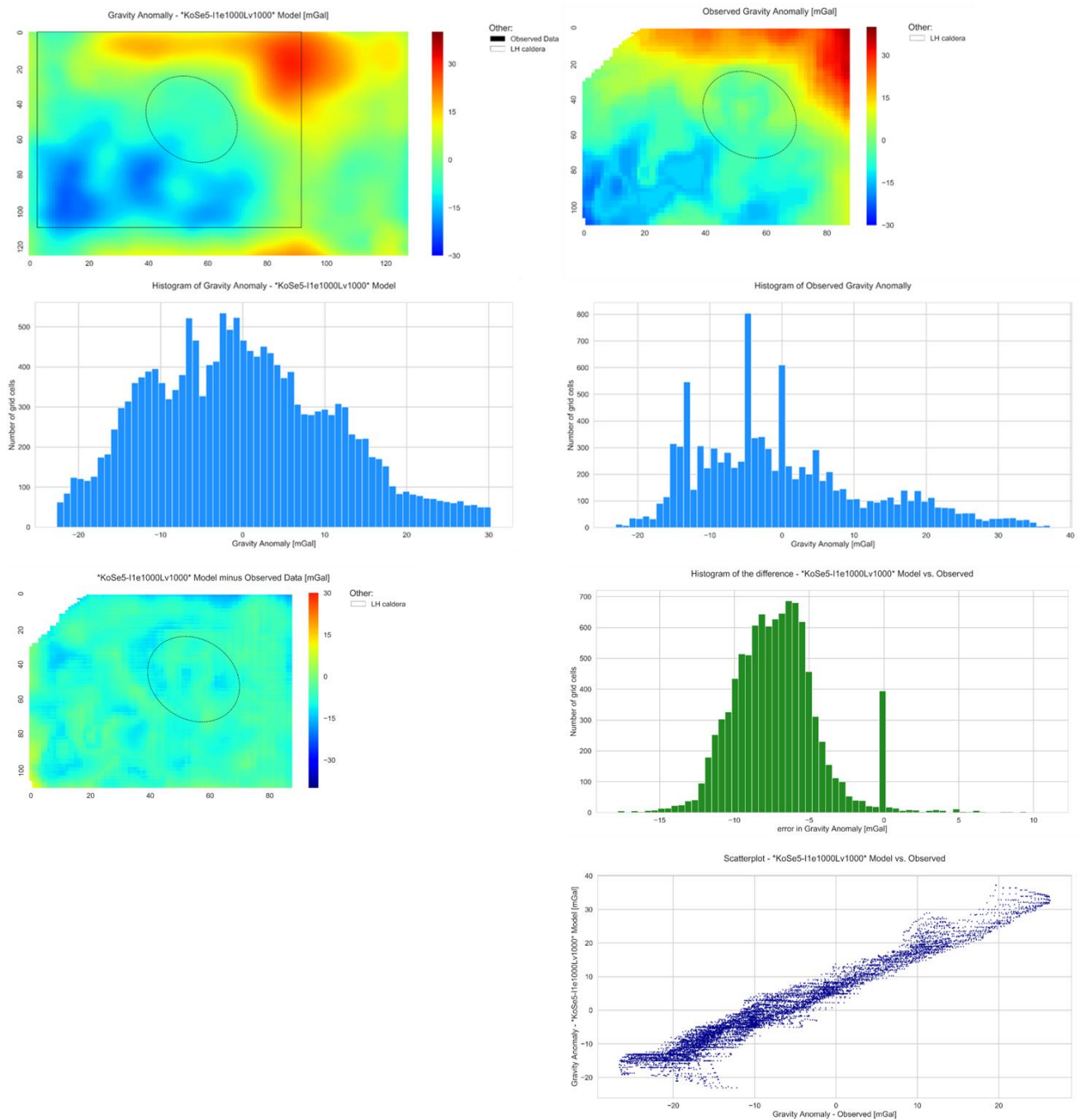
Inverse vs. Observed gravity anomaly: Observed error 20



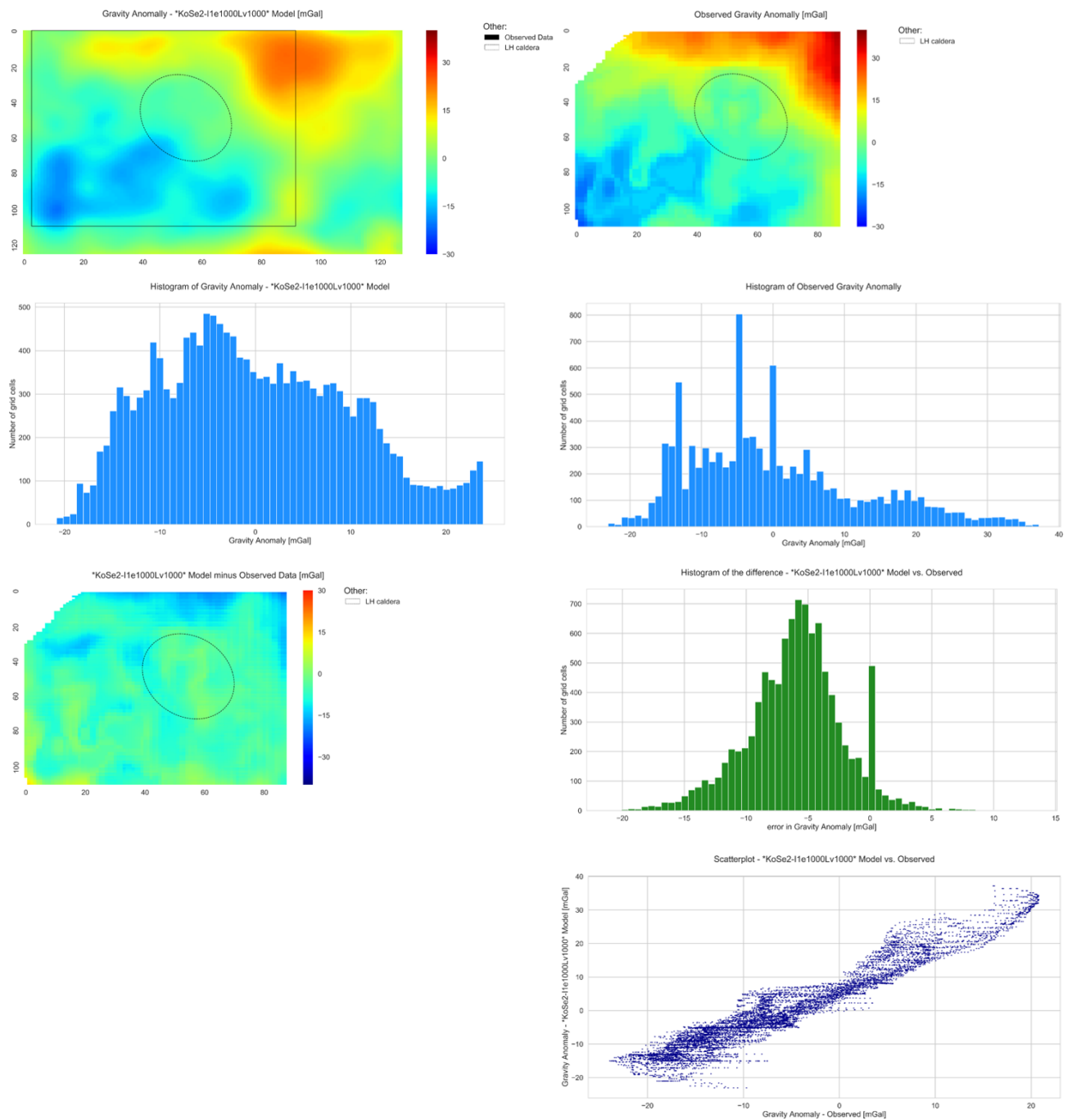
Inverse vs. Observed gravity anomaly: Observed error 10



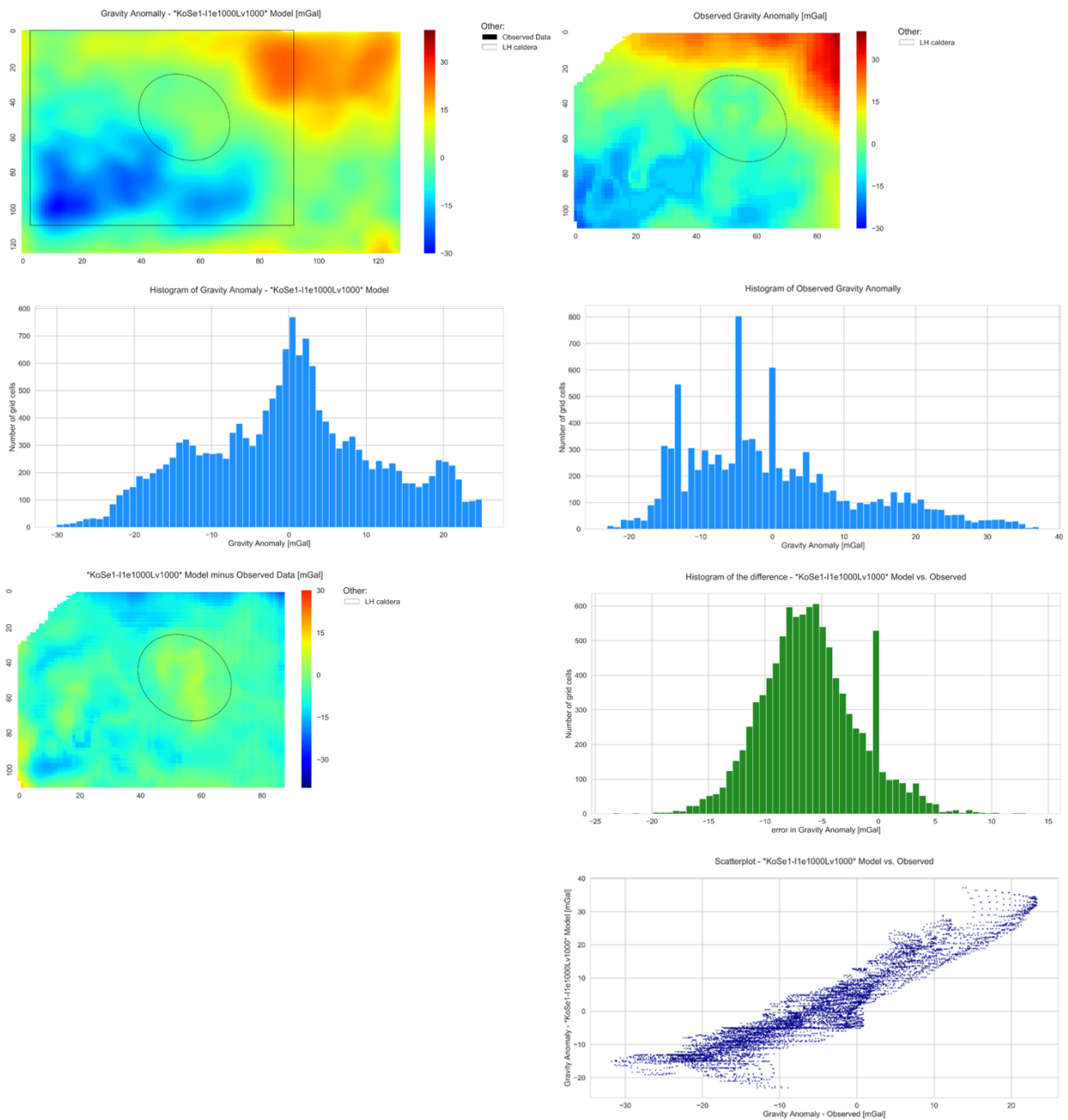
Inverse vs. Observed gravity anomaly: Observed error 5



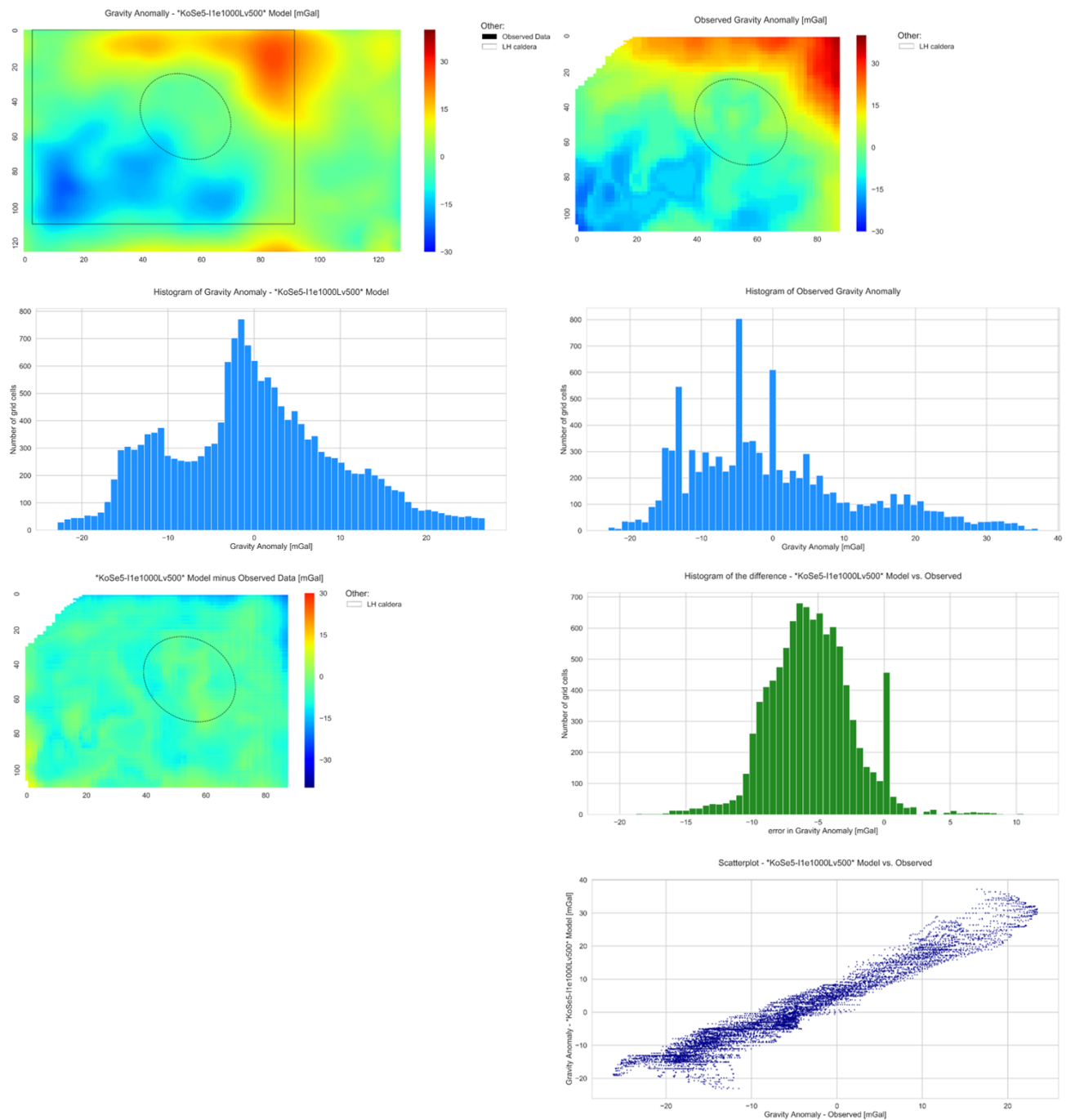
Inverse vs. Observed gravity anomaly: Observed error 2



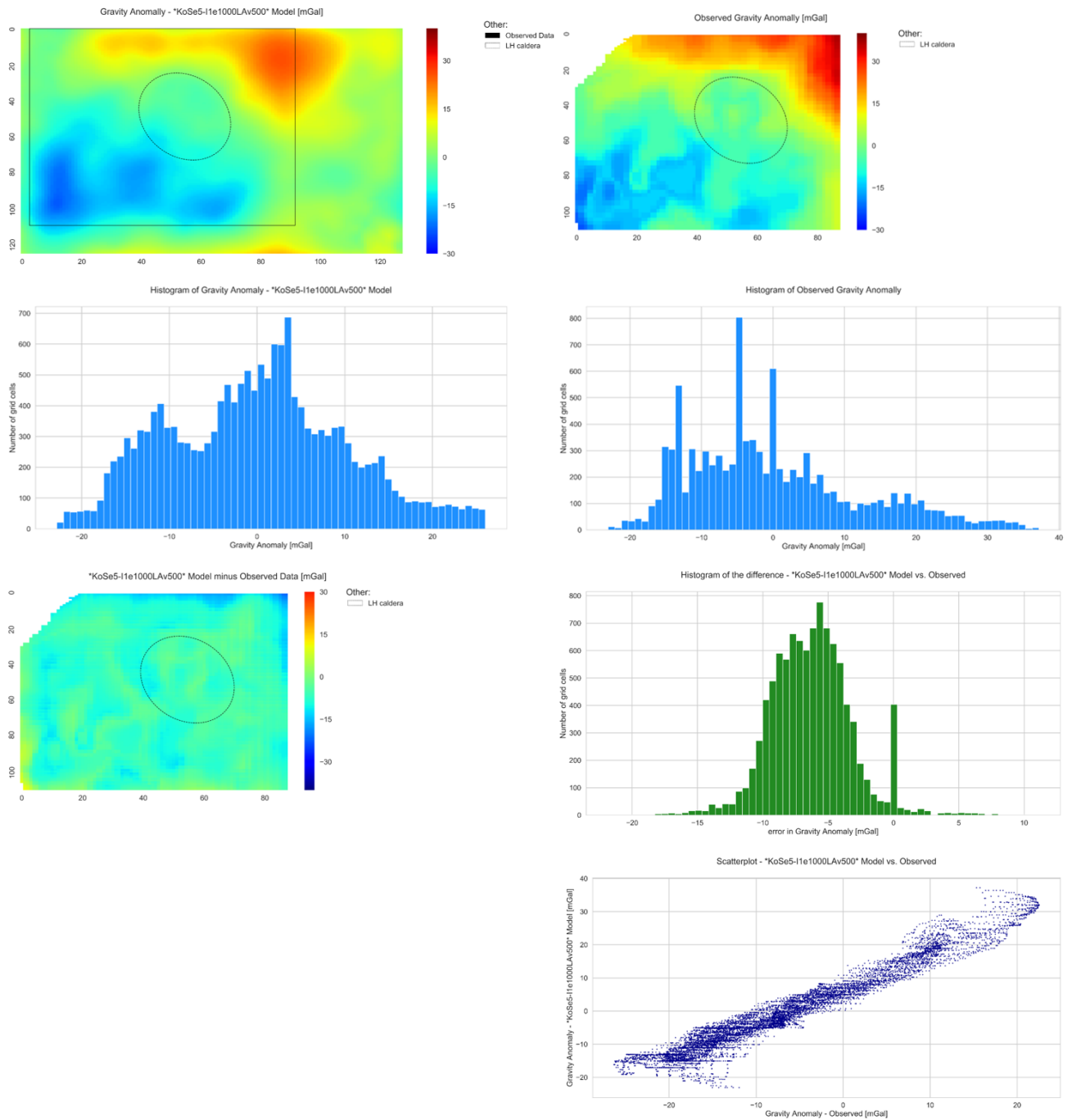
Inverse vs. Observed gravity anomaly: Observed error 1



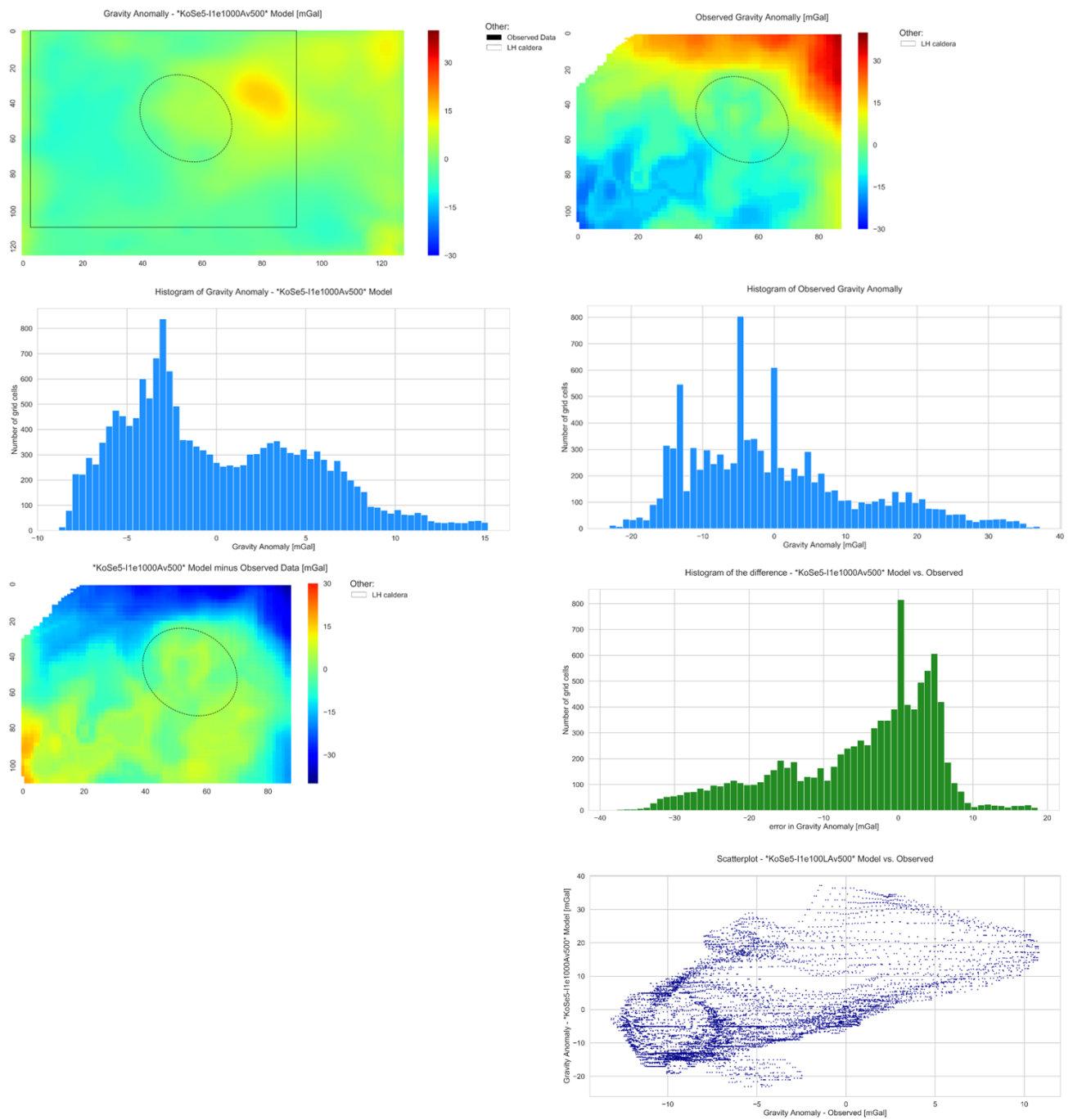
Inverse vs. Observed gravity anomaly: Limestone Variation



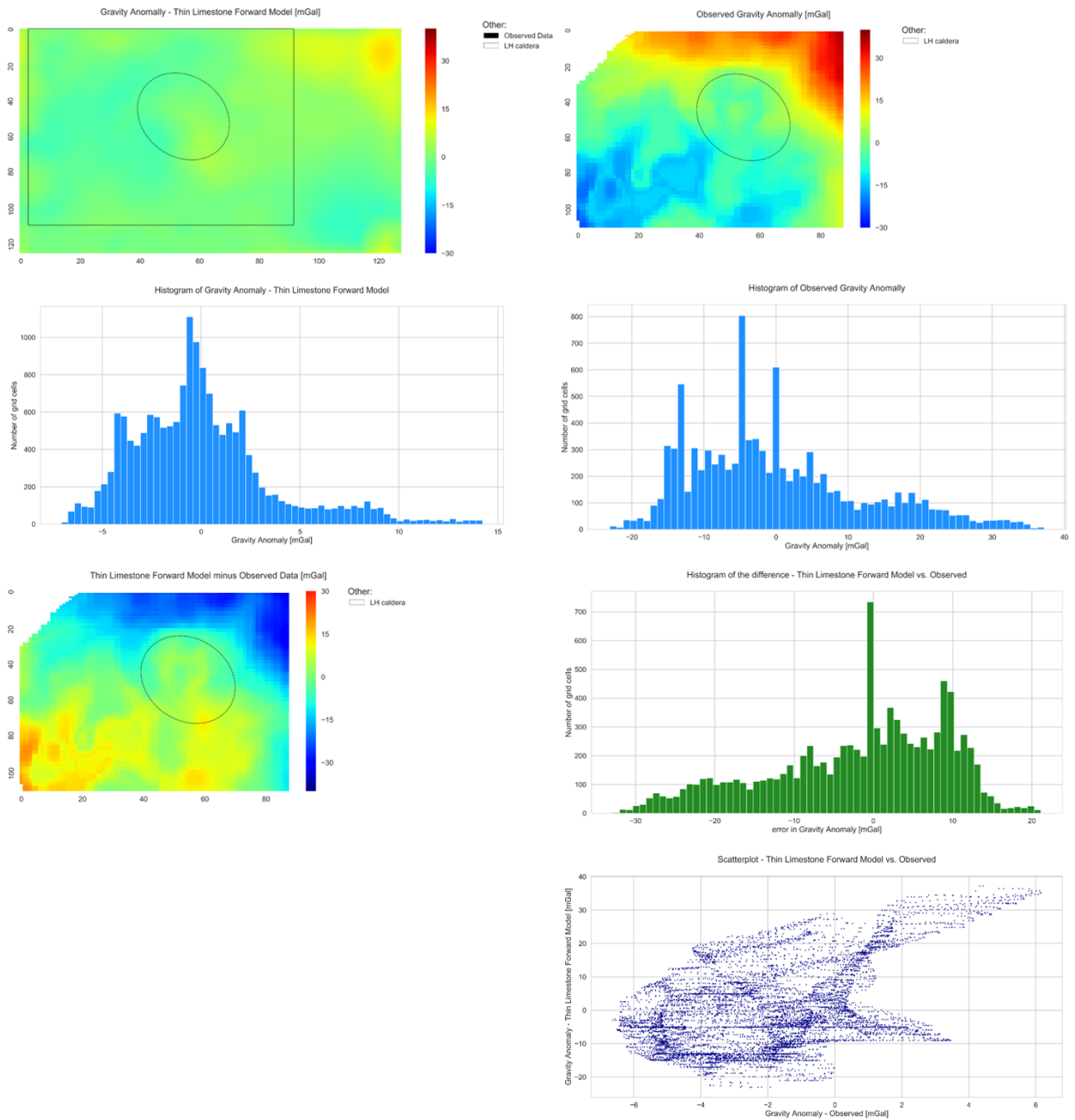
Inverse vs. Observed gravity anomaly: Limestone and Andesite Variation



Inverse vs. Observed gravity anomaly: Andesite Variation



Thin Limestone Structural model – Forward vs. Observed gravity anomaly



Thin Limestone Structural model – Inverse vs. Observed gravity anomaly: Limestone and Andesite Variation

