Prediction of free air anomaly obtained by geophysical multivariate data in Amazonia: an approach to missing data imputation with Self-Organizing Maps

Bruno Monserrat Perillo, Cleyton de Carvalho Carneiro, Ana Cristina de Matos, Denizar Blitzkow, Stephen Fraser

This paper was prepared for presentation during the 15th International Congress of the Brazilian Geophysical Society held in Rio de Janeiro, Brazil, 31 July to 3 August, 2017.

Abstract

This work deals with the creation of synthetic database values on multivariate analysis of Self-Organizing Maps (SOM). Terrestrial and airborne gravimetric data were used to compare synthetic with real values. Gap areas were generated. As a result, three analysis were developed, with different map sizes, which showed a high correlation between original and synthetic data produced from the Best Matching Units (BMUs). The synthetic values of anomaly terrestrial Free Air showed a correlation of 84%, while the terrestrial Bouguer anomaly was 87% correlated to the originals. Suggesting a high correlation between the data produced and the data generated. The variation in the map size did not influence abruptly, but we suggest that other analysis be done to explore these dimensions.

Introduction

Gravimetric data are widely used in several studies and investigations related to theoretical and applied models.

Acquisition of gravimetric data can be developed from satellite, airborne or terrestrial surveys. Current sampling in satellite or airborne surveys, has lower spatial resolutions than terrestrial surveys. However, despite the higher sampling density, the difficulty of access in terrestrial surveys can generate irregular samplings or coverage of sparse or non-existent data.

The Amazon region in Brazil, characterized by vast vegetation cover and few access roads, has aerial and terrestrial gravimetric surveys at different sampling scales. The Amazonian land surveys, despite the better sample density than the available aerial surveys, have many areas of missing data.

There are several predictive methods that can synthetically create values in areas with missing data. However, most of these methods are statistical, and processed in a single variable, such as neighborhood-based interpolations. Self-Organizing Maps (SOM) consists of a non-linear and non-statistical method, based on the principles of vector quantization and the measures of the vector similarity (Kohonen, 2001). SOM applications can be used in classifications, predictions, noise reduction, among others.

In a multivariate database, the development of SOM analyzes provides the treatment of each sample as a vector unit. After creating an n-dimensional space, where n is the number of variables involved in the analysis, the samples elect a Best Matched Unit (BMU). The BMU vector provides the corresponding value for each variable. Thus, in a multivariate database, it becomes possible to create synthetic values of samples with missing values from their respective BMUs. The researches presented in Samad & Harp (1992) and Folguera et al. (2015) show case studies involving the use of BMUs to create synthetic values in regions with missing data.

In the Amazon region, there are available aerial and terrestrial gravimetric surveys, which can be used to evaluate the accuracy of the BMUs' synthetic values. In addition, omission of part of the existing land data for later comparison with the generated synthetic values, as in Schafer & Graham (2002), provides the quantitative evaluation of these results.

This manuscript uses multivariate gravimetric data, with different spatial resolution of samples, to produce synthetic values in terrestrial gravimetric variables in the Amazon region. In order to do so, the BMUs of SOM analysis will be used in different sizes of 2-D maps, which simplify the representation of n-dimensional space.

Method

A synthesis of the methodological sequence can be seen in Figure 1. The variables were initially analyzed to observe correlations that may generate spurious tendencies to the results. Thus, variables with correlations close to -1, 0 and 1 were excluded. The remaining variables were adjusted to ensure spatial matching between the samples. Then, gap areas were created in the original data, for later comparison with the synthetic values. The SOM training was developed in three different map sizes: (A) 24x24; (B) 36x36; And (C) 50x50. The results were compared and interpreted using statistical evaluations.

Data-set

The data covers a region located between 2N to 5S in latitude and 44W to 58W in longitude. The original data-set used has variables and was built with a grid of samples with spatial resolution of 5' composed by the following data:

- digital terrain model (SAM3s_v2);
- gravity disturbance of global geopotential model and geoid heights (EIGEN_6C4; degree and order = 200);

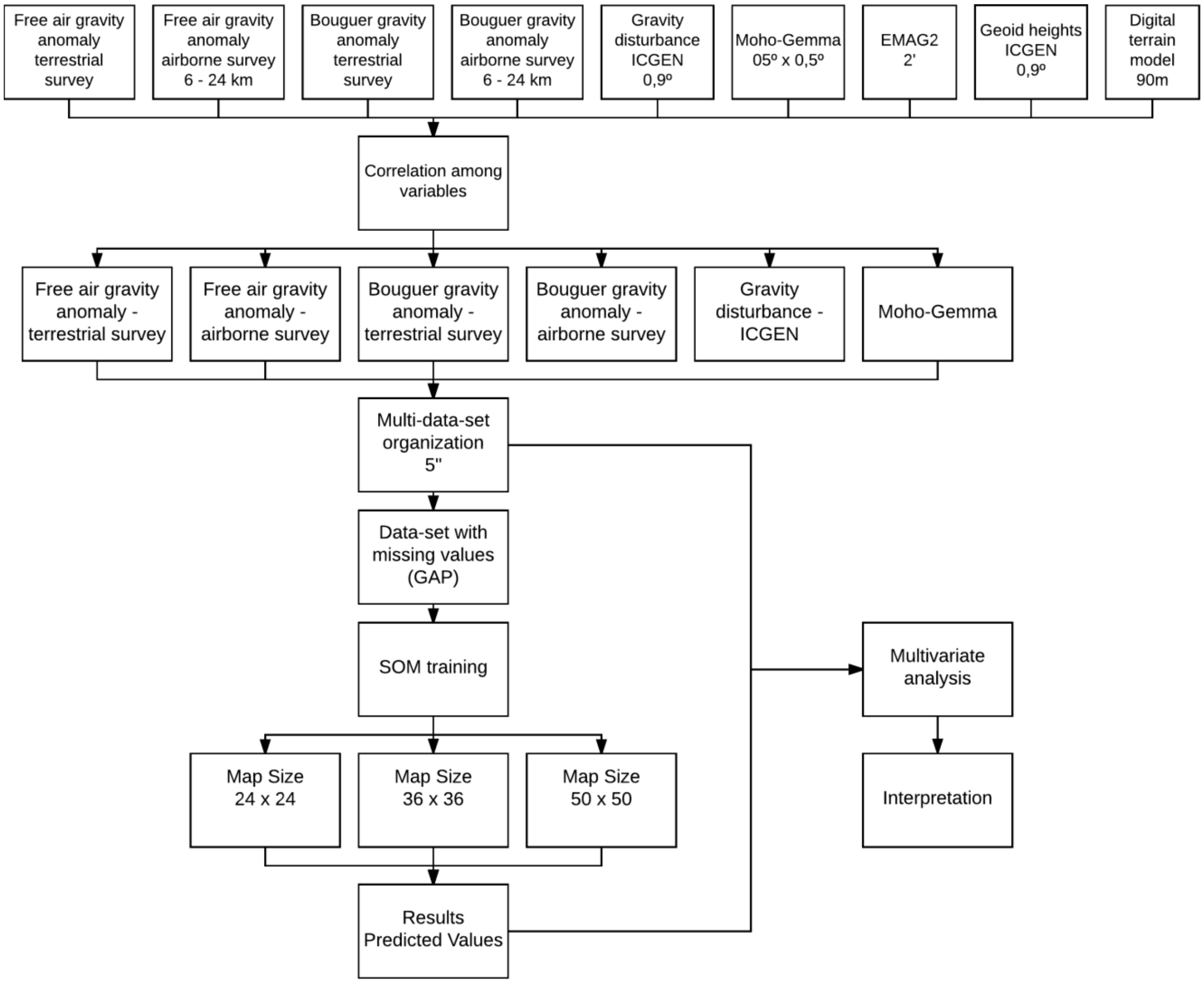


Figure 1 – Schematic of SOM data prediction

- free air gravity anomaly obtained from aero and terrestrial gravity;
- Bouguer gravity anomaly obtained from aero and terrestrial gravity;
- Moho Gemma;
- EMAG2

Terrestrial gravimetry survey was performed by Petrobras in the 1950s close to the Amazon River and affluents. A total of 43605 points with different spacing patterns were collected. The airborne geophysical survey covered the entire area analyzed in this research (Figure 2). It consists of 6km and 24km spaced gravimetric lines (Andrade and Konzen, 2008).

GFZ German Research Centre (GFZ) for Geosciences Potsdam, German, and Groupe de Recherche de Géodésie Spatiale (GRGS)/CNES, Toulouse, produced the global geopotential model EIGEN-6C4 (degree and order 2190), which is a global combined gravity field model (Shako et al, 2014; Förste et al, 2014). This model is based on the Gravity field and steady-state data of the Ocean Circulation Explorer (GOCE) satellite and Gravity Recovery and Climate Experiment (GRACE) satellite.

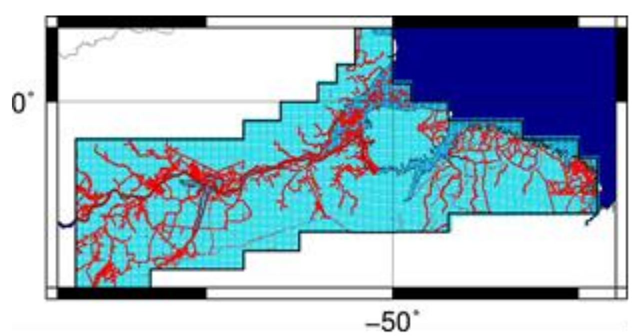


Figure 2 – Aerogravity (blue) and terrestrial gravity (red) located in Amazon region.

The GEMMA project (GOCE Exploitation for Moho Modeling and Applications), funded by the European Space Agency and Politecnico di Milano, aimed at estimating the Moho depth from GOCE data in key regions of the world.

The digital terrain model used in this work (SAM3s_v2), has a resolution of 3" of arc, comprising a total of 1983 grid cells of 1º x 1º. This model is based on SRTM (Shuttle Radar Topography Mission). The altitude of this model is linked to the EIGEN-GL04C, presenting better consistency in the results of the geoidal model for South America than the original SRTM (EGM96 was used).

EMAG2 is a global earth magnetic anomaly grid, it has been compiled from satellites, ships and airborne magnetic measurements (Maus et al, 2009).

Sampling

Primarily, the free air anomaly was calculated using the acquired data of the terrestrial gravimetry surveys (coordinates, orthometric altitude and acceleration of gravity). With this information and after correction of the terrain, Bouguer anomalies were obtained. Afterward, a grid with coordinates in the center of the 5’ resolution grid was generated. Bouguer anomalies were obtained by a simple mean, using the points belonging to each grid. The Generic Mapping Tools computational package is used for this calculation (WESSEL et al., 2013). Subsequently, the mean free air anomalies in 5’ grid cells were reconstructed, using the digital elevation model SAM3s_v2.

Results

The SOM method was applied to predict missing values in the missing data (or gap) areas. Than, the results were compared with the known values. The information was omitted in two areas within the coordinates 2S-3S/55W

-56W and 1S-2S/46W-47W (6% of the total area). The mean, standard deviation, maximum, minimum and coefficient of correlation (R^2) values of the difference between the data-set and the results of SOM method using a data-set with missing values.

This research evaluated three experiments to verify the influence of the map size parameter. For a SOM arrangement, the U-Matrix will have map size (MS) = Nx × Ny dimension, usually determined by the heuristic relationship of MS = $5\sqrt{N}$, defined by Vensato et al. (2000) where N = 6912, is the amount of data. The three experiments are (A) 24x24; (B) 36x36; e (C) 50x50.

In the gap areas, after SOM analysis using MS 24 x 24, R^2 values showed 84% and 87% for free air and Bouguer anomaly, respectively. Using a map size 36 x 36, the coefficient correlation pointed 84% and 86%, respectively. Finally, in the third experiment, R^2 showed 81% and 83% (table 1).

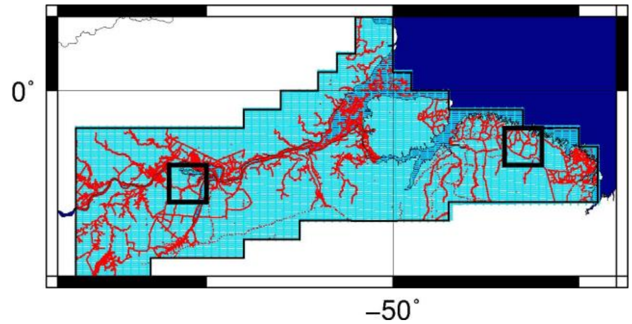


Figure 3 –Two GAPs were the SOM method was applied to predict free air anomaly and Bouguer anomaly.

Conclusions

A high correlation was observed between original and synthetic data produced from the BMUs. The synthetic values of anomaly terrestrial Free Air showed a correlation of 84%, while the terrestrial Bouguer anomaly were 87% correlated to the originals.

The different sizes of maps applied in SOM analyzes show little discrepancy in the absolute values generated between the correlation of the synthetic data with the original data.

SOM analyzes can be widely used for the generation of synthetic data in gravimetric surveys, especially in regions with difficult access and dense forest, such as the Amazon region. However, more in-depth studies on the variables involved, as well as other ways of calculating vector relations, can contribute to the improvement of the correlations of these synthetic data to real measures.

References

ANDRADE, F., KONZEN, L. 2008. Levantamento Aerogeofísico para a Identificação de Áreas com Ocorrência Potencial de Petróleo e Gás nas Bacias do Amazonas, Marajó e Adjacências. Relatório técnico - ANP

FOLGUERA, I.; ZUPAN, J.; CICERONE, D.; MAGALLANES, J. 2015.Self-organizing maps for imputation of missing data in incomplete data matrices. Chemometrics and Intelligent Laboratory Systems 143, 146-151 p.

FÖRSTE, C.; BRUINSMA, S. L.; ABRIKOSOV, O.; LEMOINE, J-M.; SCHALLER, T.; GÖTZE, H. J.; EBBING, J.;MARTY, J-C.; FLECHTNER, F.; BALMINO, G.; BIANCALE, R. 2014. EIGEN-6C4:The latest combined global gravity field model including GOCE data up to degree and order 2190 of GFZ Potsdam and GRGS Toulouse. 5th GOCE User Workshop, November, 25–28.11, Paris.

MAUS, S., et al. (2009), EMAG2: A 2 –arc min resolution Earth Magnetic Anomaly Grid compiled from satellite, airborne, and marine magnetic measurements, Geochem. Geophys. Geosyst., 10,

SAMAD, T. & HARP, S. 1992. Self-organization with partial data. Network 3, 205-212 p.

SHAKO, R.; FÖRSTE, C.; ABRYKOSOV, O.; BRUINSMA, S.; MARTY, J.-C.; LEMOINE, J.-M.; FLECHTNER, F.; NEUMAYER, K.-H.; DAHLE, C. 2014. EIGEN-6C: A High-Resolution Global Gravity Combination Model Including GOCE Data - In: Flechtner, F., Sneeuw, N., Schuh, W.-D. (Eds.), Observation of the System Earth from Space - CHAMP, GRACE, GOCE and future missions, (GEOTECHNOLOGIEN Science Report; No. 20; Advanced Technologies in Earth Sciences), Berlin.

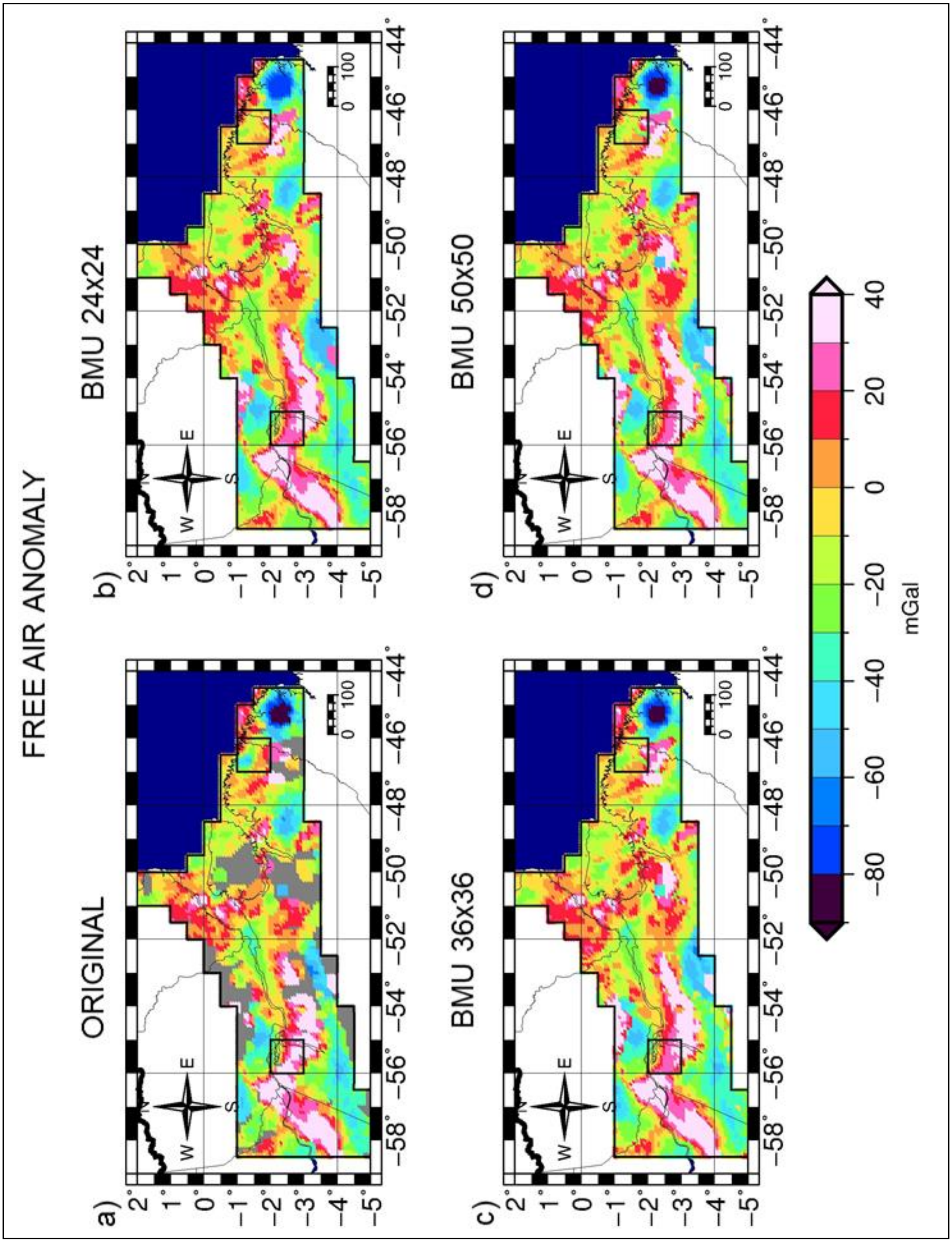
WESSEL, P., SMITH, W. H. F., SCHARROO , R., LUIS, J., WOBBE, F. 2013. Generic Mapping Tools: Improved Version Released. EOS Trans. AGU, 94, 409-410,

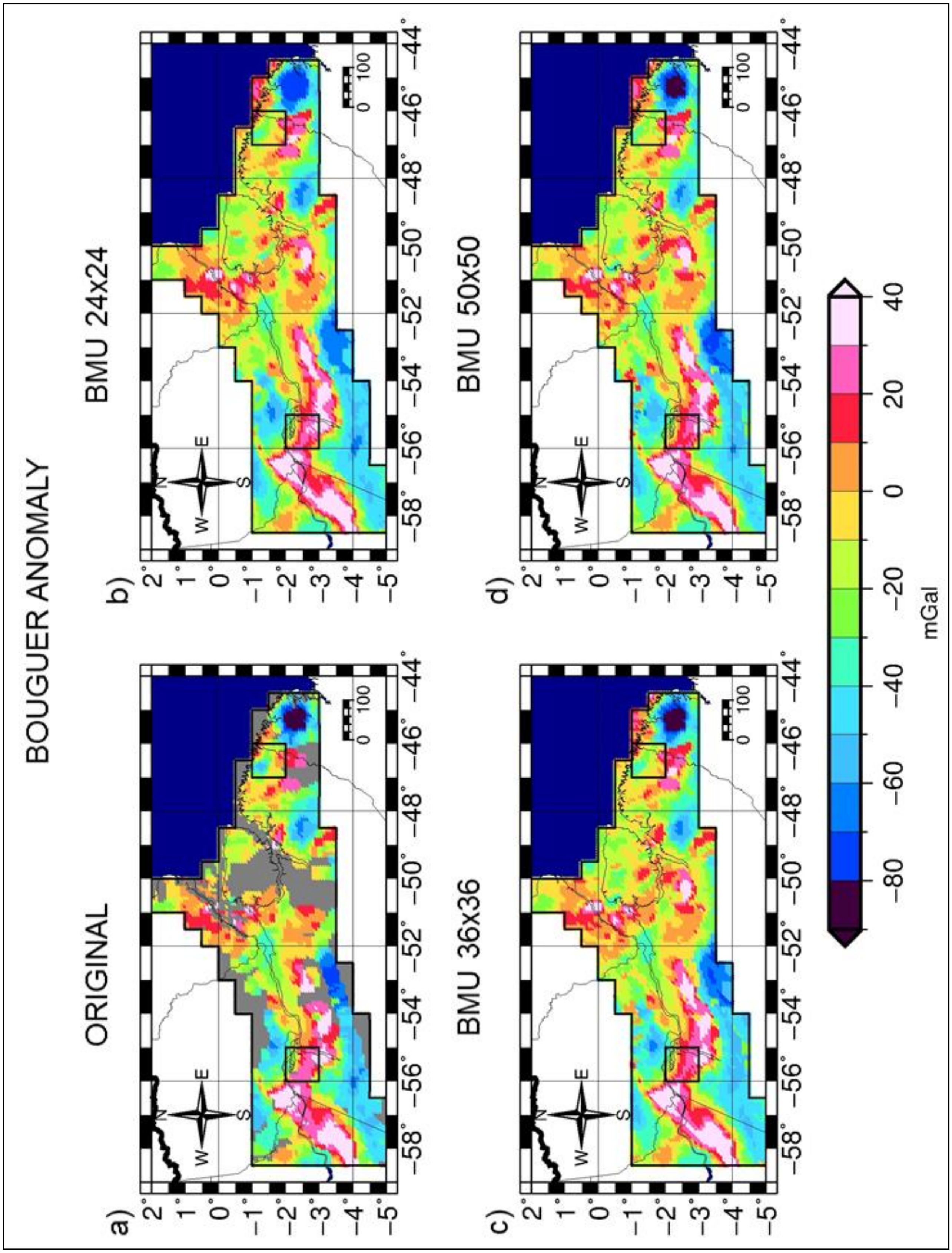
Table 1 – Comparison among different map size dimension: (A) Matrix 24 x 24; (B) Matrix 36 x 36; and (C) Matrix 50 x 50.

(A) Analysis A	**Spatial Resolution of Total Area (original 5' x BMU X' values)**						**Spatial Resolution of GAP Area**	
Map Size 24 x 24	Free Air ANP	Bouguer ANP	Disturb Grav	Moho Gemm	Free Air Ter	Bouguer Ter	Free Air Ter	Bouguer Ter
Average (dif)	-0,01	-0,02	-0,02	0,00	-0,17	-0,18	-0,17	-0,17
Std dev (dif)	4,23	4,05	4,16	0,90	5,91	5,26	5,91	5,26
Max (dif)	28,37	25,01	18,81	4,45	49,30	58,45	49,30	58,45
Min (dif)	-40,77	-32,29	-27,44	-4,70	-43,76	-35,02	-43,76	-35,02
R^2	**0,97**	**0,98**	**0,96**	**0,97**	**0,95**	**0,97**	**0,84**	**0,87**

(B) Analysis B	**Spatial Resolution of Total Area (original 5' x BMU X' values)**						**Spatial Resolution of GAP Area**	
Map Size 36 x 36	Free Air ANP	Bouguer ANP	Disturb Grav	Moho Gemm	Free Air Ter	Bouguer Ter	Free Air Ter	Bouguer Ter
Average (dif)	0,00	0,00	-0,02	0,00	-0,12	-0,13	-0,12	-0,13
Std dev (dif)	3,28	3,13	3,03	0,66	4,67	4,40	4,67	4,40
Max (dif)	25,98	18,87	15,10	3,62	45,68	55,81	45,68	55,81
Min (dif)	-39,23	-30,53	-23,19	-4,41	-31,04	-31,30	-31,04	-31,30
R^2	**0,98**	**0,99**	**0,98**	**0,98**	**0,97**	**0,98**	**0,84**	**0,86**

(C) Analysis C	**Spatial Resolution of Total Area (original 5' x BMU X' values)**						**Spatial Resolution of GAP Area**	
Map Size 50 x 50	Free Air ANP	Bouguer ANP	Disturb Grav	Moho Gemm	Free Air Ter	Bouguer Ter	Free Air Ter	Bouguer Ter
Average (dif)	0,00	0,01	-0,01	0,00	-0,10	-0,09	-0,10	-0,11
Std dev (dif)	2,69	2,55	2,40	0,54	4,16	3,96	4,20	4,00
Max (dif)	25,00	16,06	9,91	2,89	47,39	45,88	45,88	45,88
Min (dif)	-36,06	-29,90	-22,87	-3,27	-37,69	-37,37	-37,37	-37,37
R^2	**0,99**	**0,99**	**0,99**	**0,99**	**0,98**	**0,98**	**0,81**	**0,83**


FREE AIR ANOMALY
a)
ORIGINAL
b)
BMU 24x24
c)
BMU 36x36
d)
BMU 50x50
mGal
−80
−60
−40
−20
0
20
40


BOUGUER ANOMALY
a)
ORIGINAL
b)
BMU 24x24
c)
BMU 36x36
d)
BMU 50x50
mGal