

---

# Exploratory Multivariate Statistical Analysis of Curvature of the Lunar Surface

Special Project

---

Jesús F. Almansa G.

# EXPLORATORY MULTIVARIATE STATISTICAL ANALYSIS OF CURVATURE OF THE LUNAR SURFACE

PREPARED, WRITTEN, AND TYPESET BY: Jesús F. Almansa G. (s131905)

SUPERVISED BY: Allan Aasbjerg Nielsen

---

Project period:	June 2015
Type:	Special Course
Study programme:	M.Sc. Earth and Space Physics and Engineering Mapping and Navigation
Comments:	This document was typeset using L <sup>A</sup> T <sub>E</sub> X, and contains color images. All reported computations and images were programmed using MATLAB.

DTU Compute  
Technical University of Denmark  
DK-2800 Kgs. Lyngby  
Denmark

<http://www.compute.dtu.dk/english>

# Contents

Preface . . . . .	i
<b>1 Introduction</b> . . . . .	<b>1</b>
<b>2 Data Set</b> . . . . .	<b>3</b>
2.1 Topography . . . . .	3
2.2 Geoid Anomalies . . . . .	5
2.3 Crustal Thickness . . . . .	8
<b>3 Surface Curvatures</b> . . . . .	<b>11</b>
3.1 Vertical Curvatures . . . . .	11
3.2 Horizontal Curvatures . . . . .	13
3.3 Difference Curvature . . . . .	15
3.4 Gauss Curvature . . . . .	15
3.5 Mean Curvature . . . . .	15
<b>4 Moonform Classification</b> . . . . .	<b>16</b>
4.1 Gauss Classification . . . . .	20
4.2 Efremov-Krcho Classification . . . . .	20
4.3 Shary Classification . . . . .	20
<b>5 Considerations on Discriminant Analysis</b> . . . . .	<b>22</b>
5.1 Quality of Discrimination with k-means . . . . .	22
5.2 Data Standardization and other Metrics . . . . .	24
<b>6 Principal Component Analysis</b> . . . . .	<b>25</b>
6.1 Analysis . . . . .	25
6.2 Discrimination . . . . .	25
<b>7 Conclusions</b> . . . . .	<b>28</b>
<b>Bibliography</b> . . . . .	<b>30</b>
<b>A Statistical Summaries of Curvatures</b> . . . . .	<b>31</b>

# Chapter 1

## Introduction

This work deals with the use of Differential Geometry to study spatial surfaces (i.e., manifolds in the three-dimensional Euclidean space) in general, and the use of Geomorphometry to study topographic surfaces in particular, but the emphasis is made on some statistical tools that could be employed in such studies.

The description of a topographic surface can be done with the help of classical notions of Differential Geometry. The application of these notions to study real-life surfaces is general enough so that anything else that is not the surface itself is abstracted out from the study.

However, this is not sufficient for the study of landforms, since from the viewpoint of Geomorphometry there must be considered a context within which the surface actually exists and that may explain the changes it is subject to. Gravity, among other geophysical forces, provides such a context.

In [3], it is provided a systematic account of several techniques that can be used to study landforms on topographic surfaces (such as the surface of the Earth or other celestial bodies), as well as geological horizons, and thus the techniques are applicable in particular to the lunar surface. However, in this work we apply the same techniques upon two other lunar spatial surfaces: Geoid Anomalies and Crustal Thickness. The author is not aware of literature regarding studies applying the same techniques on such spatial surfaces.

The surfaces representing Geoid Anomalies and Crustal Thickness are indeed not topographic surfaces or geological horizons, but mathematical surfaces obtained as the graphs of functions of the form  $z = f(x, y)$  where  $z$  is the value corresponding to a given latitude-longitude coordinate  $(x, y)$  in a selenographic coordinate system.

Nonetheless, under the conditions that they are orientable and smooth, they can be studied with the general tools of Differential Geometry, and under the condition of verticality, they can be studied with the particular tools of Geomorphometry.

The first condition, orientability, is inherent to scalar fields defined on geographic (or, in our case, selenographic) coordinate systems. The second condition, smoothness, can always be assumed as a form of approximation to reality. The third condition, that we call verticality, needs always a justification. Such condition refers to the possibility of associating a vertical vector to every point of the surface that is normal to a zero-level surface of reference. In the case of topographic surfaces the reference surface is the geoid and the vertical vector is the gravity vector.

It turns out that both the lunar Geoid Anomalies surface and the lunar Crustal Thickness surface are scalar fields that are defined in terms of other reference surfaces, the lunar spheroid and the lunar geoid, and thus the verticality condition is satisfied.

In this work we aim specifically at applying the landform classification techniques as presented in [3] to all three lunar surfaces, and to assess statistical characteristics of classification using a simple form of Discriminant Analysis.

The data set that we use consists of elevation and gravity data obtained by the 1991-1994 joint NASA/DoD mission to the Moon dubbed Clementine.

**Acknowledgement:** The author would like to thank especially to Steen Markvorsen for helping in understanding many notions of Differential Geometry and for providing several suggestions that improved this work.

# Chapter 2

## Data Set

The Clementine mission performed several experiments that produced a variety of data, but in this work only three types of data were used: Topography, Geoid Anomalies, and Crustal Thickness. Next, we describe the data sets as obtained from the PDS Geosciences Node, Washington University in St. Louis, USA [5]. In addition to depicting the data, we also provide statistical summaries, and comment on some features of the Moon that are of interest.

### 2.1 Topography

The Clementine Laser Image Detection And Ranging (LIDAR) experiment was designed to measure the slant range from the spacecraft to the lunar surface in order to derive elevations on the surface of the Moon. Data are available from [8] under the `topo` directory as file `topogr1.dat` in ASCII format.

The file contains a 2D regular grid of the topography of the Moon relative to a spheroid of radius 1738 km at the equator, with a flattening of 1/3234.93. The grid is a  $180 \times 360$  grid at  $1 \times 1$  degree resolution with latitudes in the range  $[-89.5, 89.5]$  and longitudes in  $[0.5, 359.5]$ . The coordinate system is selenocentric, center of mass, longitude positive east, and the location of the pole and the prime meridian are defined according to [7, 2].

Observations are topography in meters in the format F8.1, and are arranged as 10 observations per line in 6480 lines. Every 360 observations read downward and rightward correspond to a whole latitude. The first observation is the topography at 89.5 degrees of latitude and 0.5 degrees of longitude.

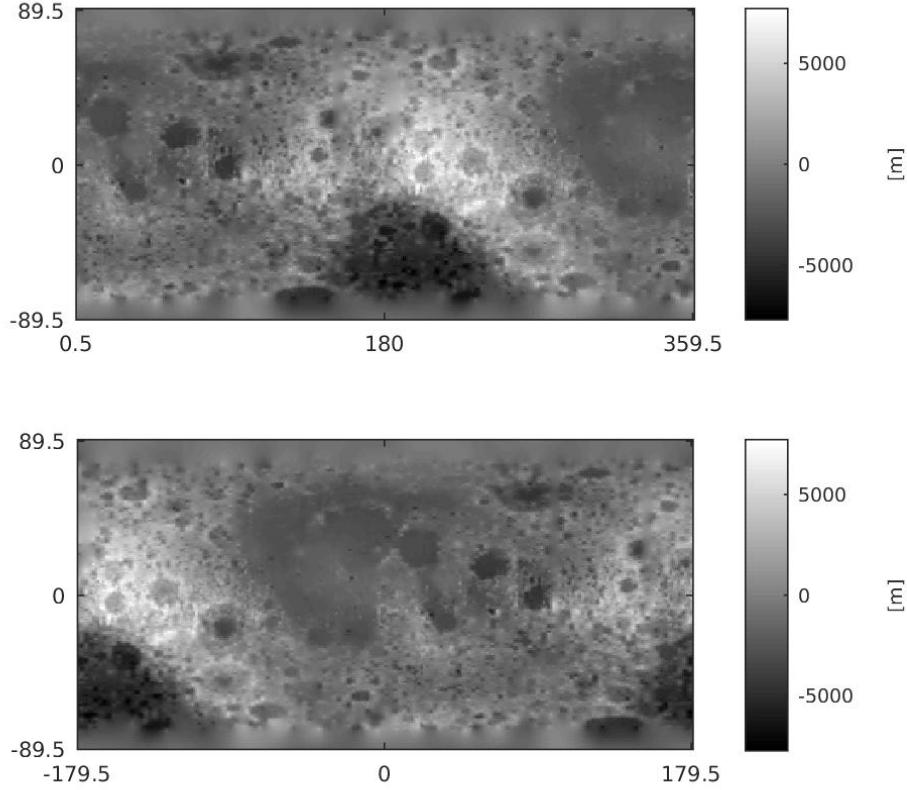
Figure 2.1 depicts the data using a Plate Carrée projection and a gray color scale in meters. The plot on the top depicts the data as provided, where the central longitude is at  $180^\circ$ , whereas in the plot on the bottom the central longitude is at  $0^\circ$ . The latter is the conventional representation of the Moon, where the Near Side (i.e. the hemisphere of the Moon that always faces Earth) is the central part of the scene; this is the representation that will be adopted in this work.

The relative errors in Clementine topography are governed by the orbital accuracy, of order 10 m, and the LIDAR noise, of order 40 m.

### Features of Interest

Because of the low data-resolution, the identification of features on the Moon is difficult. In order to ease this task, we applied on the Topography data a symmetric Gaussian smoothing [cite] of size 3 followed by a subsampling of size 1, with the purpose of reducing noise and obtaining an image of higher detail. The results appear in Figure 2.2.

In the figure, we applied the process twice, ending with an image of  $45 \times 90$  pixels. Even though at a proper scale the image seems sharper than the original, it is too small to enable feature identification. We then proceeded to obtain a  $0.25 \times 0.25$  degree resolution 2D grid for



*Figure 2.1: Topography of the Moon. Central longitudes at 180° (top) and 0° (bottom).*

Topography [8] and applied a Gaussian smoothing. With the result, as shown in Figure 2.3, we were able to identify some features of interest that we will refer to along this document. They can be found by scanning rightward the map, and have been indicated using markers:

**Arrows:** These are mounts. Starting with Mount Archimedes, and following with two mounts which appear to form a single geological feature that has a chevron (spær) -like shape, Mount Apenninus and Mount Haemus.

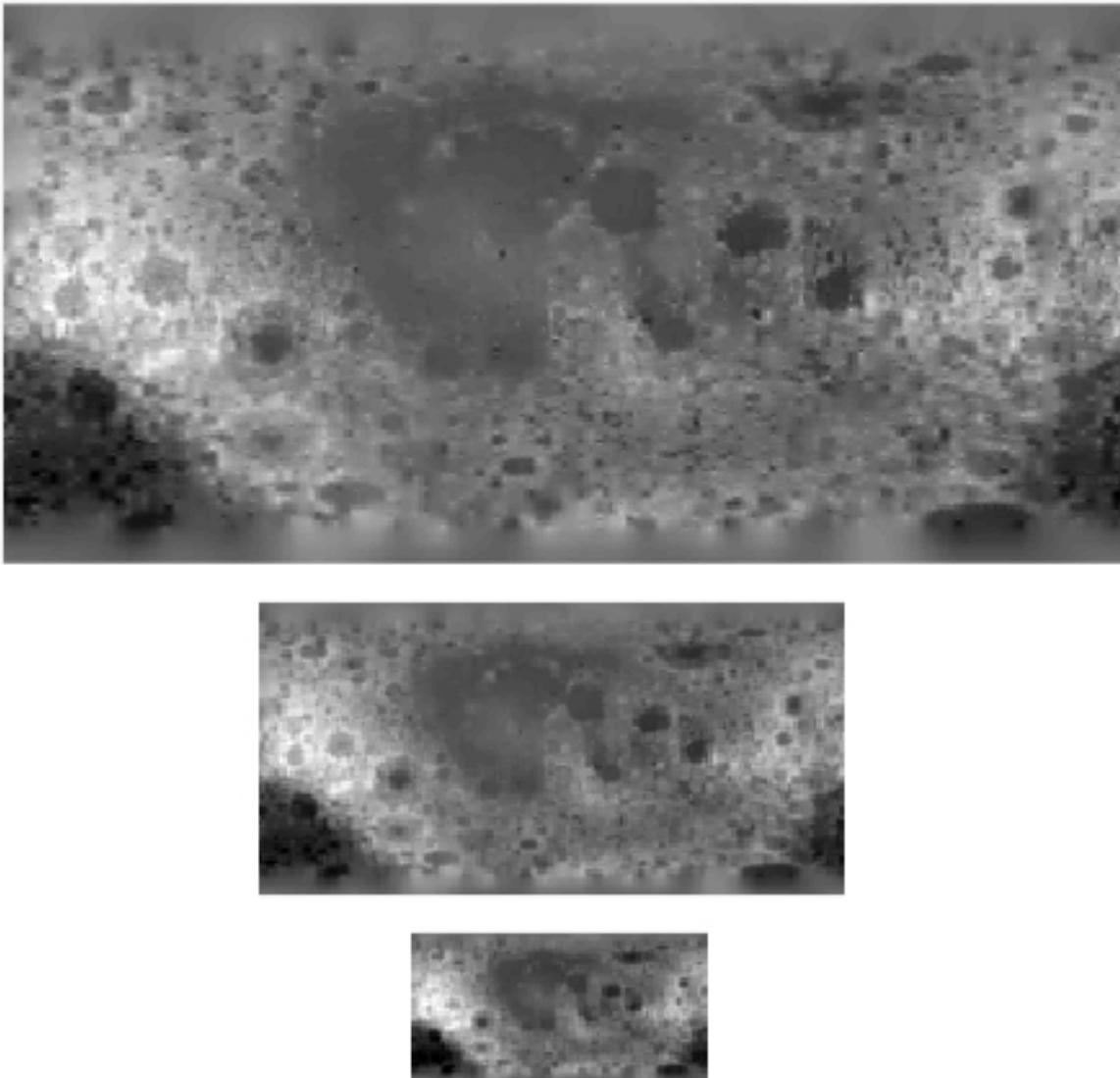
**Ellipses:** These are all moon seas (lunar maria). Starting with Mare Orientale, and continuing with Mare Imbrium, Mare Serenitatis, Mare Crisium, and Mare Smythii.

**Rectangles:** These are craters. Starting with Apollo, and continuing with Hertzsprung, Pythagoras, Copernicus, and Langrenus.

## Statistical Summary

Variable Topography has 64800 observations in units of meter, with a mean of -881.98 m and a standard deviation of 2021.8 m. Figure 2.4a shows a box plot of the variable. It ranges in the interval [-7729.9, 8403.1] in meters, with first, second, and third quartiles equal to -2114.7 m, -1059.9 m, and 86.8 m, respectively. With a whisker length of 1.5, values seem to be symmetrically distributed, and yield a wider range for outliers in higher elevations.

Figure 2.4b shows a histogram with ca.  $\sqrt{64800}$  bins and a normal density function fit. The data distribution seems however somewhat skewed to the right and has a larger peak than the normal distribution. This is confirmed by finding that the skewness, 0.54078, is positive, and that the



*Figure 2.2: Gaussian smoothing and subsampling on image for Topography.*

kurtosis, 4.112, is greater than 3. The lack of symmetry is indicated too by the fact that the mean is bigger than the median.

Finally, Figure 2.4c shows the deviation of data (dashed line) from normality (crosses).

## 2.2 Geoid Anomalies

The Clementine S-band Transponder Doppler Gravity experiment was designed to measure perturbations in the motion of the spacecraft in order to infer the lunar gravity field. Data are available from [4] under the **gravity** directory as file **geoidgrd.dat** in ASCII format.

The file contains a 2D regular grid of geoid anomalies of the Moon relative to a 70th degree and order spherical harmonic gravitational field model, designated Goddard Lunar Gravity Model 2 (GLGM-2), on a spheroid of radius 1738 km at the equator, with a flattening of 1/3234.93. The grid is a  $180 \times 360$  grid at  $1 \times 1$  degree resolution with latitudes in the range  $[-89.5, 89.5]$  and longitudes in  $[0.5, 359.5]$ . The coordinate system is selenocentric, center of mass, longitude positive east, and the location of the pole and the prime meridian are defined according to [1].

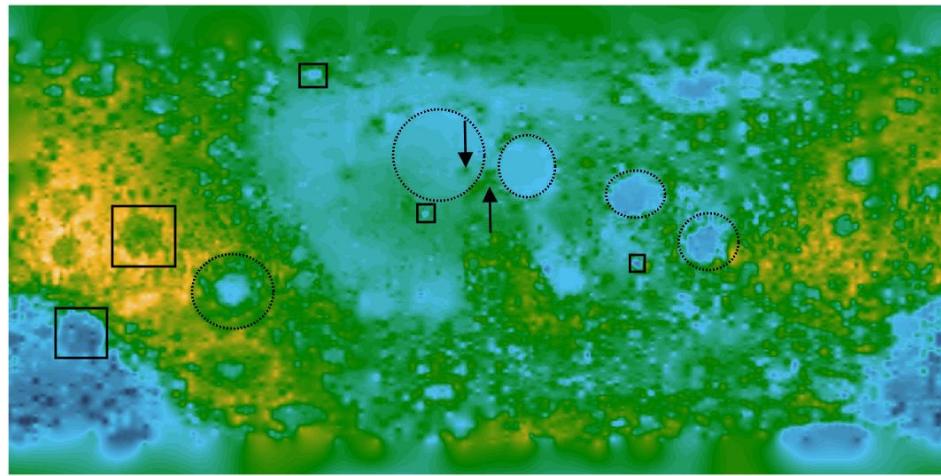


Figure 2.3: Some features of the Moon.

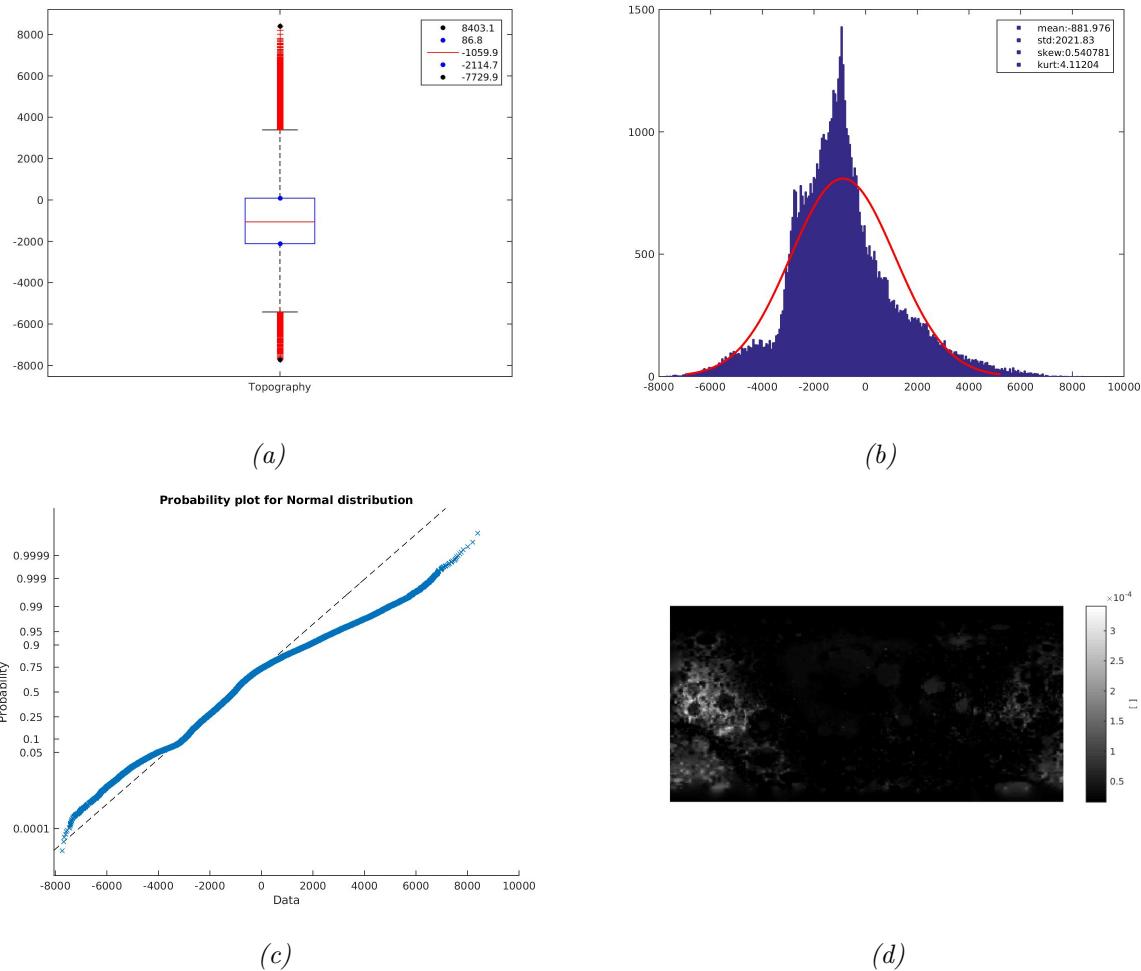
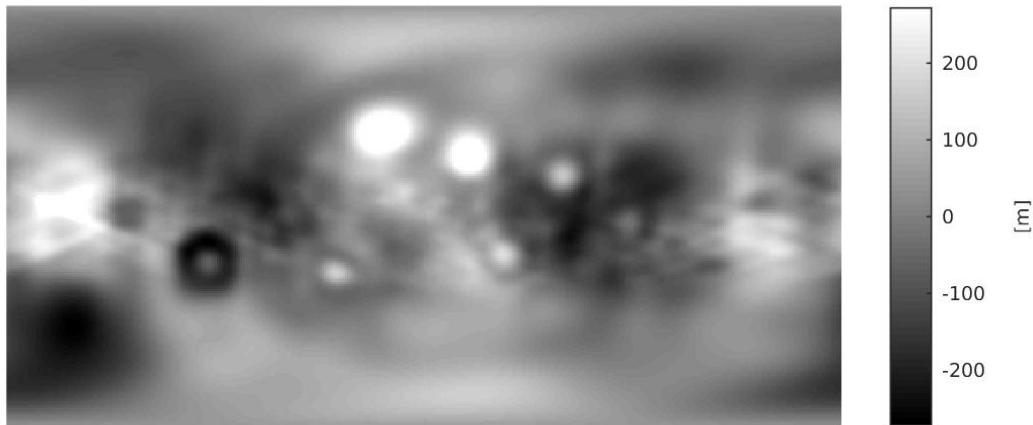


Figure 2.4: Summary for Topography: (a) boxplot; (b) histogram; (c) probability; (d) leverages.

Observations are geoid anomalies in meters in the format F8.1, and are arranged as 10 observations

per line in 6480 lines. Every 360 observations read downward and rightward correspond to a whole latitude. The first observation is the geoid anomaly at 89.5 degrees of latitude and 0.5 degrees of longitude.

Figure 2.5 depicts the data using a Plate Carrée projection and a gray color scale in meters.



*Figure 2.5: Geoid anomalies of the Moon.*

We observe that at all maria display centers of higher anomalies inscribed in rings of lower anomalies. High latitude maria display centers of highest anomalies, whereas maria closer to the equator have centers of lowest anomalies. Notable are maria Orientale and Smythii, which show rings of very low anomalies relative to their centers. Geoid anomalies at other features are doubtful, except for craters Apollo and Hertzsprung, which in contrast to maria show centers of lower anomalies inscribed in rings of higher anomalies.

Errors in the geoid range from 4-24 meters.

### Statistical Summary

Variable Geoid Anomalies has 64800 observations in units of meter, with a mean of 5.426185 m and a standard deviation of 99.584340 m. Figure 2.6a shows a box plot of the variable. It ranges in the interval [-271.9, 473.21] in meters, with first, second, and third quartiles equal to -63.32 m, 8.47 m, and 73.32 m, respectively. With a whisker length of 1.5, values seem to be symmetrically distributed, and yield outliers almost only of positive sign.

Figure 2.6b shows a histogram with ca.  $\sqrt{64800}$  bins and a normal density function fit. The data distribution seems to be slightly skewed to the right and has a larger peak than the normal distribution. This is confirmed by finding that the skewness, 0.242912, is positive, and that the kurtosis, 3.80351, is greater than 3. The lack of symmetry is indicated too by the fact that the mean is smaller than the median.

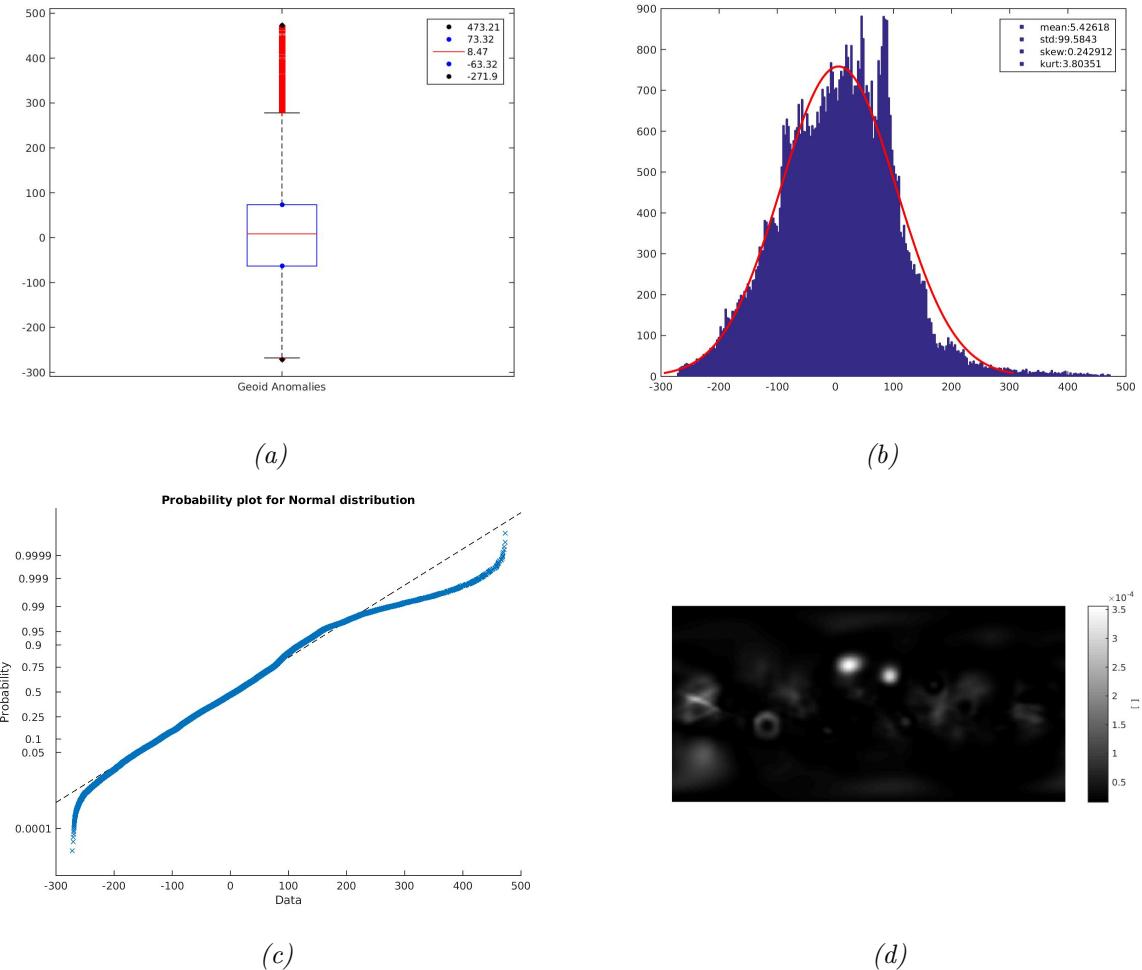


Figure 2.6: Summary for Geoid Anomalies: (a) boxplot; (b) histogram; (c) probability; (d) leverages.

Finally, Figure 2.6c shows the deviation of data (dashed line) from normality (crosses).

## 2.3 Crustal Thickness

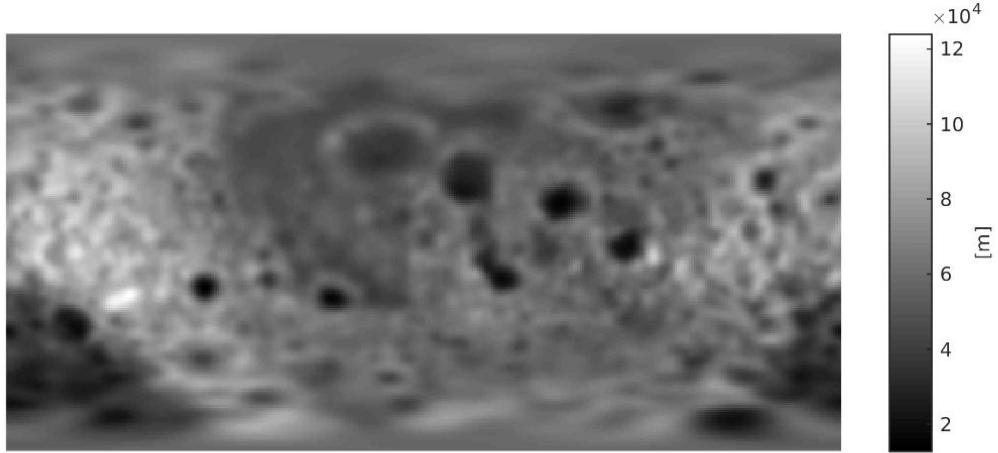
Crustal thickness is derived data that results from the difference between inferred Moho topography and the surface topography, assuming constant density crust and mantle. In the Moon, these constant values are  $2800 \text{ kg/m}^3$  and  $3300 \text{ kg/m}^3$ , respectively, and in addition it needs to be taken into account the mare basalt layer with a density of  $3200 \text{ kg/m}^3$ . Data are available from [4] under the **gravity** directory as file **thickgrd.dat** in ASCII format.

The file contains a 2D regular grid of crustal thickness of the Moon relative to a 70th degree and order spherical harmonic gravitational field model, designated Goddard Lunar Gravity Model 2 (GLGM-2), and the own-mission (Clementine) topography model that is relative to a spheroid of radius 1738 km at the equator, with a flattening of  $1/3234.93$ . The grid is a  $180 \times 360$  grid at  $1 \times 1$  degree resolution with latitudes in the range  $[-89.5, 89.5]$  and longitudes in  $[0.5, 359.5]$ . The coordinate system is selenocentric, center of mass, longitude positive east, and the location of the pole and the prime meridian are defined according to [1].

Observations are crustal thickness in kilometers in the format F8.1, and are arranged as 10 observations per line in 6480 lines. Every 360 observations read downward and rightward correspond to a whole latitude. The first observation is the crustal thickness at 89.5 degrees of latitude and 0.5 degrees of longitude. For data processing, crustal thickness values are converted

to meters.

Figure 2.7 depicts the data using a Plate Carrée projection and a gray color scale in meters.



*Figure 2.7: Crustal Thickness of the Moon.*

We observe that all maria display centers of lower thickness inscribed in rings of higher thickness. High latitude maria display centers of highest thickness, whereas maria closer to the equator have centers of lowest thickness. Notable is Mare Orientale, which shows a ring of alternating higher and lower thickness.

The magnitudes of crustal thickness for mounts seem to be close. At craters, crustal thickness are doubtful, except for craters Apollo and Hertzsprung, which display dissimilar structure. The former shows crustal thickness alike a mare, whereas the latter shows a center of higher thickness with a star-like shape.

### Statistical Summary

Variable Crustal Thickness has 64800 observations in units of meter, with a mean of 59742.9 m and a standard deviation of 14077.4 m. Figure 2.8a shows a box plot of the variable. It ranges in the interval [12756, 123887] in meters, with first, second, and third quartiles equal to 51997.5 m, 58840 m, and 66757.5 m, respectively. With a whisker length of 1.5, values seem to be symmetrically distributed, and yield more outliers of higher values.

Figure 2.8b shows a histogram with ca.  $\sqrt{64800}$  bins and a normal density function fit. The data distribution seems to be slightly skewed to the right and has a larger peak than the normal distribution. This is confirmed by finding that the skewness, 0.246853, is positive, and that the kurtosis, 3.80463, is greater than 3. The lack of symmetry is indicated too by the fact that the mean is bigger than the median.

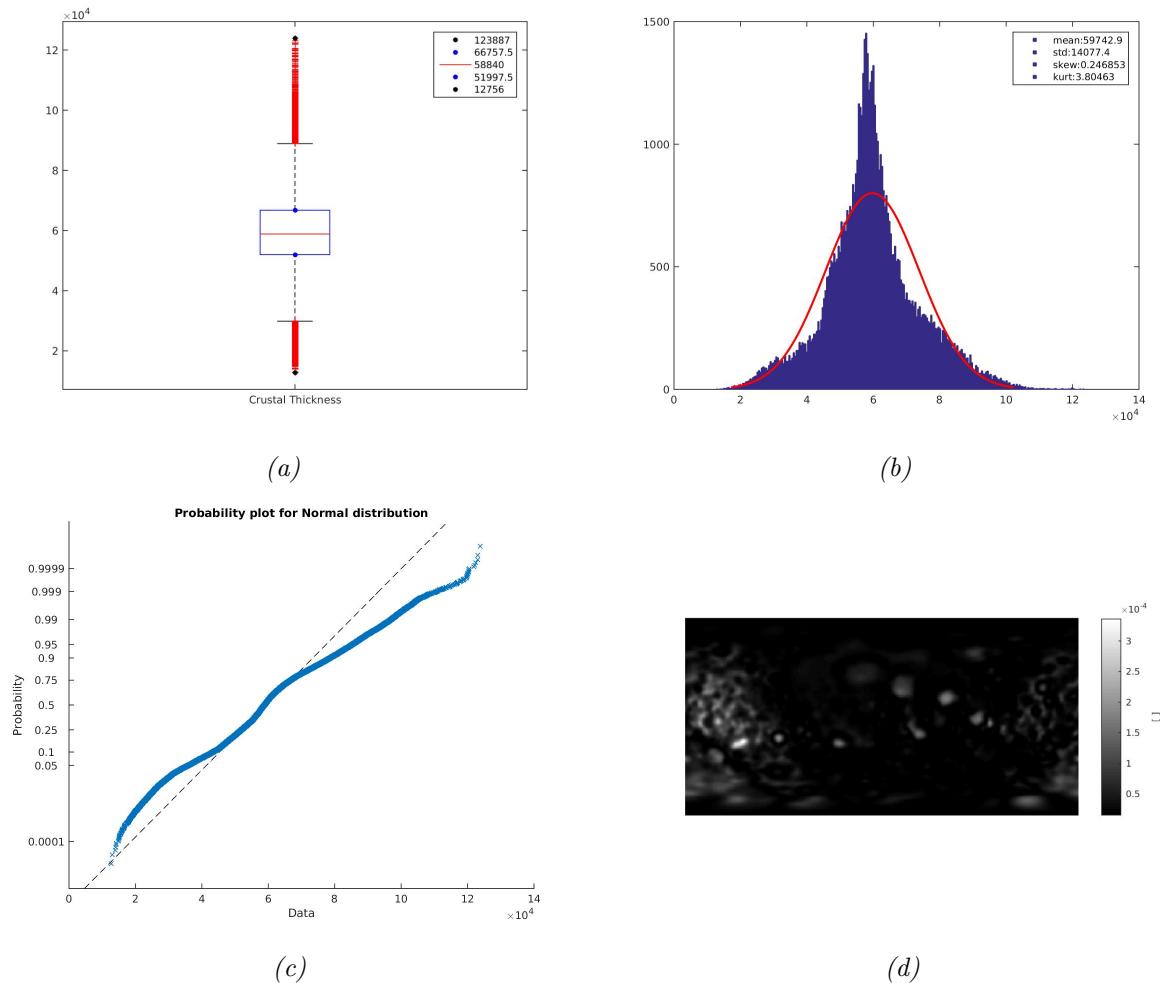


Figure 2.8: Summary for Crustal Thickness: (a) boxplot; (b) histogram; (c) probability; (d) leverages.

Finally, Figure 2.8c shows the deviation of data (dashed line) from normality (crosses).

# Chapter 3

## Surface Curvatures

A surface can be described by the geometric properties at neighborhoods of its points. From the viewpoint of Differential Geometry, the neighborhood at a given point is defined relative to the normal vector of the surface at the point. This view gives rise eventually to the notions of Gauss and mean curvatures. From the viewpoint of Geomorphometry, however, it is also important to define neighborhoods relative to the gravity vector (cf. §1) at each point. This view gives rise eventually to the notions of horizontal, vertical, and difference curvatures.

Curvatures are functions of partial derivatives of elevation. Conceptually, they are calculated by fitting a second-order polynomial to a  $3 \times 3$  square window all over the elevation *grid*, using the least-squares method. We observe that even though the fitting is performed relative to a plane grid (which is a safe assumption considering that data were acquired from a spacecraft at a mean altitude of 640 km from the Moon surface and that the fitting process takes place in surface patches of 3 arcdegrees; the assumption cannot be held though as one approaches the poles), calculations do take into account the sphericity of the surface. We refer the reader to [3] (pp. 43-45) for more details regarding this process.

Next, we provide graphical representations of the results of calculating all different types of curvatures on the surfaces for Topography, Geoid Anomalies, and Crustal Thickness. Since we are dealing with very small quantities, a logarithmic transformation is appropriate for representational purposes. In [3] (pp. 134) it is suggested to use the following transformation which preserves signs and takes into account the sphericity of the surface:

$$\Theta' = \text{sgn}(\Theta) \ln(1 + 10^{mn}|\Theta|)$$

where  $\Theta$  is the curvature variable,  $m = 2$  for Gauss curvature and  $m = 1$  for the other curvatures, and  $n \in \{2, \dots, 9\}$  takes a value depending on the local grid spacing. The minimum cell width is 264.788 m near the poles, and the maximum cell width is 30332.7 m at the equator.

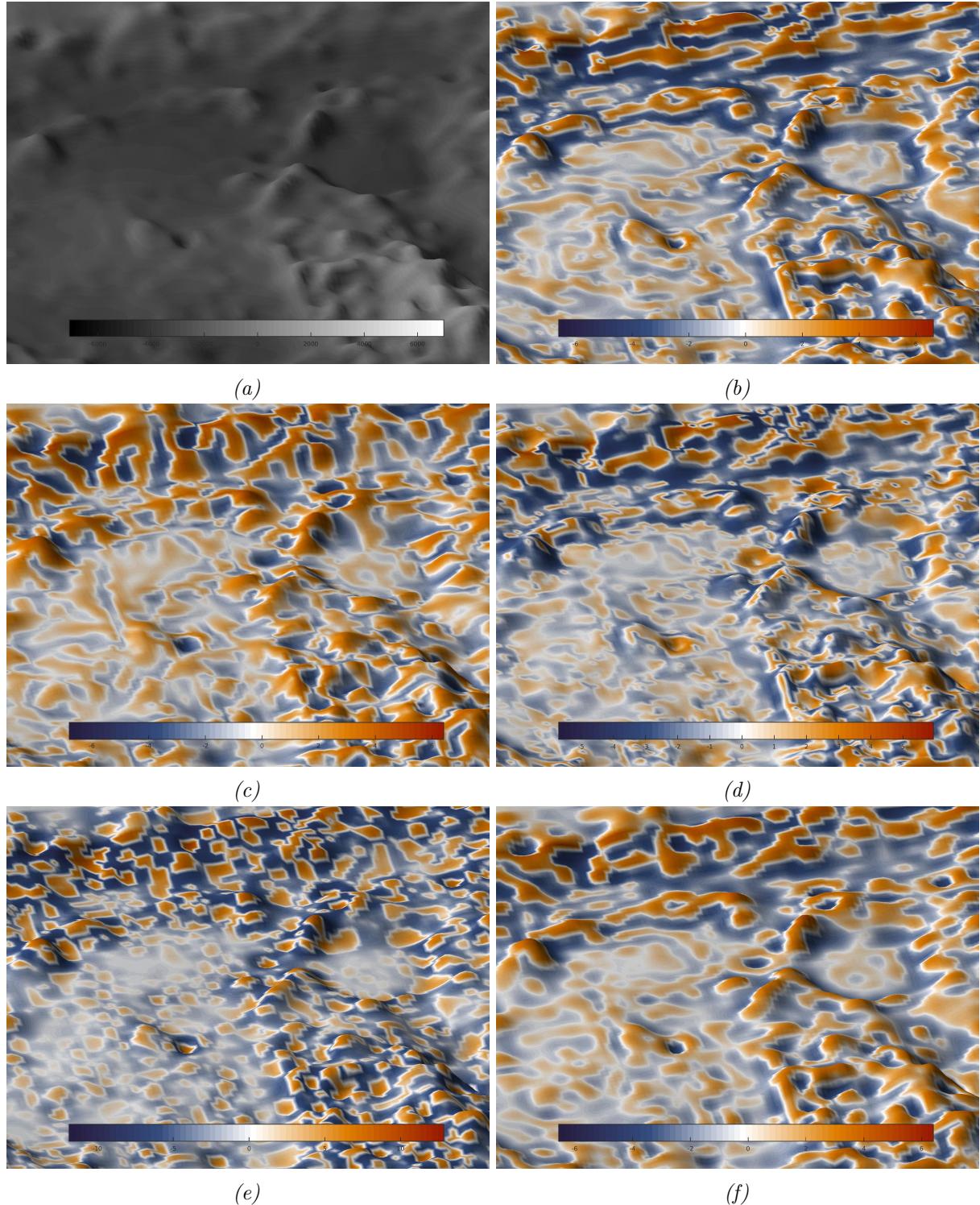
We will use the different types of curvatures only to describe the shapes of surfaces. We will discuss briefly the results in relation to some few features of interest on the Moon. For Topography, we chose a detail of its map at maria Imbrium and Serenitatis; for Geoid Anomalies, we chose a detail in the vicinity of Mare Orientale; and for Crustal Thickness the detail chosen is in the vicinity of maria Imbrium, Serenitatis, Crisium, and Smythii

In Appendix A, there can be found statistical summaries for each curvature on each of the surfaces.

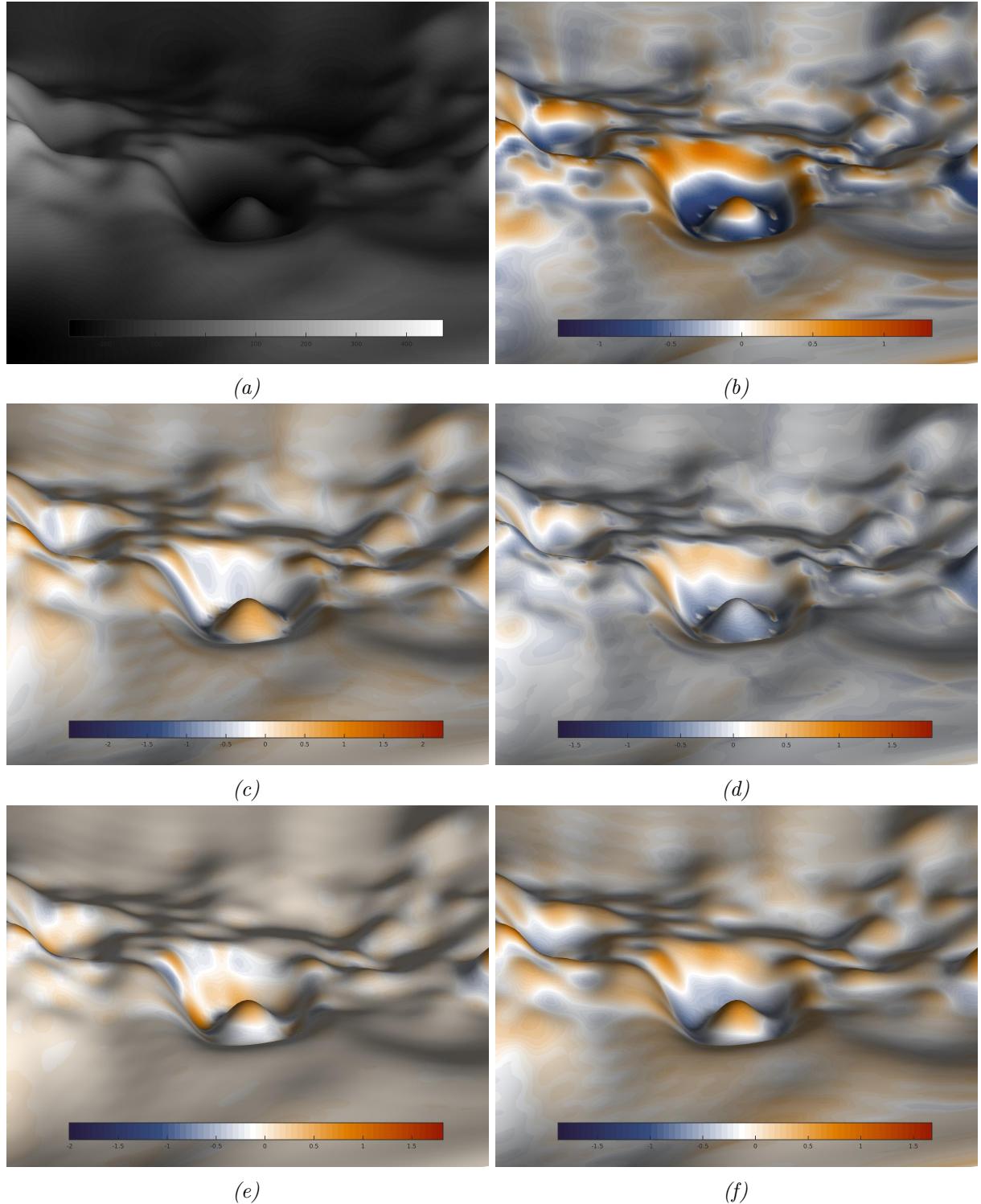
### 3.1 Vertical Curvatures

Figures 3.1b, 3.2b, and 3.3b, depict the values of vertical curvature on the Topography, Geoid Anomalies, and Crustal Thickness surfaces respectively, using a 3D Plate Carrée projection and a diverging color log scale.

In all cases, we observe positive values of vertical curvature on along vertical slopes that are convex. Negative values are observed along vertical slopes that are concave. Boundaries between positive and negative values are loci of zero curvature and they correspond to straight lines relative to vertical sections.



*Figure 3.1: Curvatures on Moon Topography. Alphabetically: Topography,  $kv$ ,  $kh$ ,  $E$ ,  $K$ , and  $H$ . Detail at maria Imbrium and Serenitatis. Log scales with diverging blue-white-orange color maps for negative-zero-positive values. Data aspect ratio [1 1 500].*

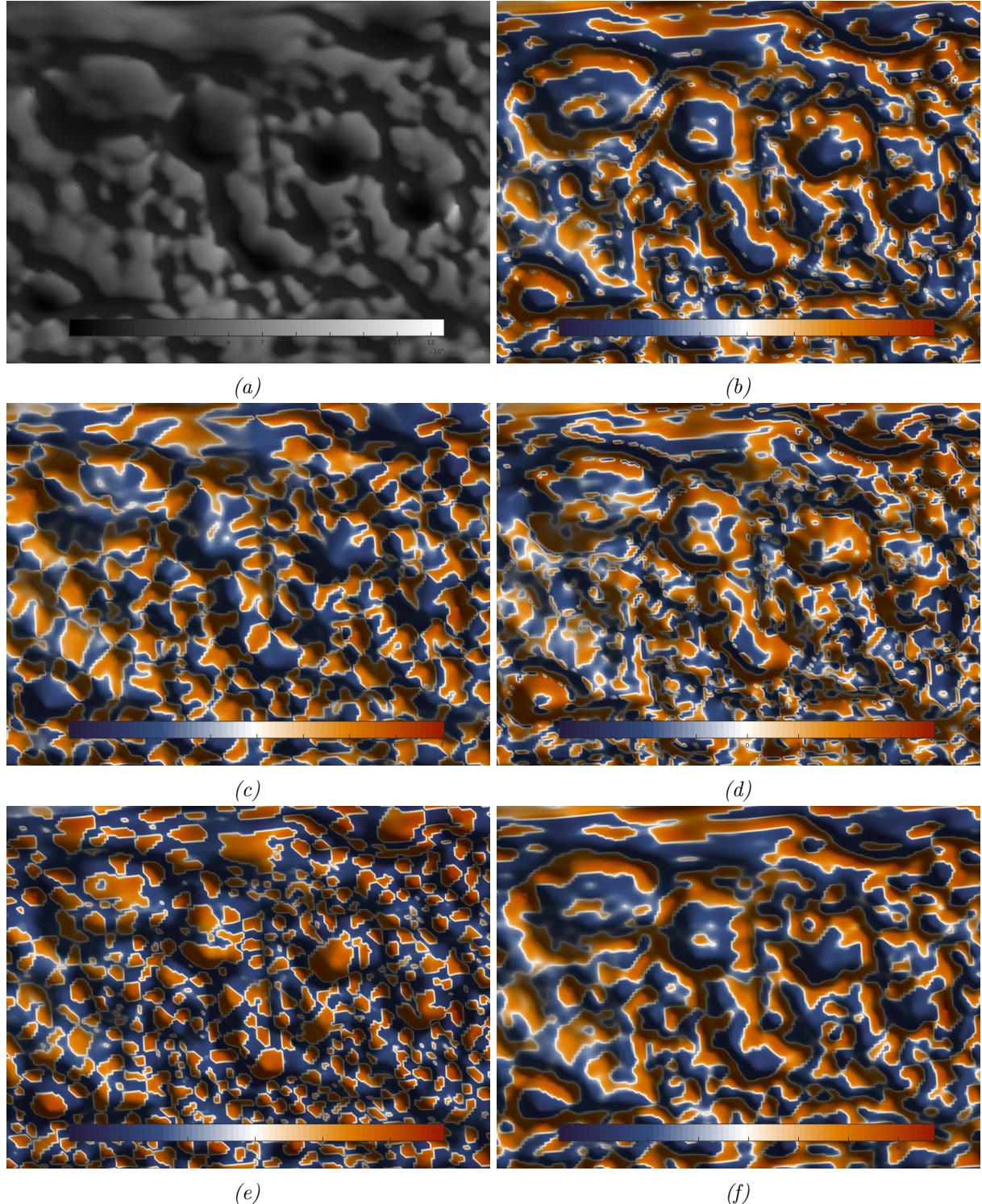


*Figure 3.2: Curvatures on Moon Geoid Anomalies. Alphabetically: Geoid Anomalies,  $kv$ ,  $kh$ ,  $E$ ,  $K$ , and  $H$ . Detail corresponding to vicinity of Mare Orientale. Log scales with diverging blue-white-orange color maps for negative-zero-positive values. Data aspect ratio [1 1 20].*

## 3.2 Horizontal Curvatures

Figures 3.1c, 3.2c, and 3.3c, depict the values of horizontal curvature on the Topography, Geoid Anomalies, and Crustal Thickness surfaces respectively, using a 3D Plate Carrée projection and a diverging color log scale.

In all cases, we observe positive values of horizontal curvature along horizontal sections that are convex. Negative values are observed along horizontal sections that are concave. Boundaries between positive and negative values are loci of zero curvature and they correspond to straight lines relative to horizontal sections.



*Figure 3.3: Curvatures on Moon Crustal Thickness. Alphabetically: Crustal Thickness,  $kv$ ,  $kh$ ,  $E$ ,  $K$ , and  $H$ . Detail corresponding to vicinity of maria Imbrium, Serenitatis, Crisium, and Smythii. Log scales with diverging blue-white-orange color maps for negative-zero-positive values. Data aspect ratio [1 1 1250].*

### 3.3 Difference Curvature

Figures 3.1d, 3.2d, and 3.3d, depict the values of difference curvature on the Topography, Geoid Anomalies, and Crustal Thickness surfaces respectively, using a 3D Plate Carrée projection and a diverging color log scale.

In all cases, we observe positive values of difference curvature on areas where there is more vertical bending than horizontal bending, regardless of convexity or concavity, whereas the opposite holds for values of negative difference curvature. Boundaries between these values are loci of zero curvature and they correspond to points of equal vertical and horizontal curvature.

### 3.4 Gauss Curvature

Figures 3.1e, 3.2e, and 3.3e, depict the values of Gauss curvature on the Topography, Geoid Anomalies, and Crustal Thickness surfaces respectively, using a 3D Plate Carrée projection and a diverging color log scale.

In all cases, we observe positive values of Gauss curvature on ellipsoidal surface patches, and negative values are observed on saddles, whether convex or concave. Boundaries between these values are loci of zero curvature and they correspond to straight lines relative to normal sections.

### 3.5 Mean Curvature

Figures 3.1f, 3.2f, and 3.3f, depict the values of mean curvature on the Topography, Geoid Anomalies, and Crustal Thickness surfaces respectively, using a 3D Plate Carrée projection and a diverging color log scale.

In all cases, we observe positive values of mean curvature on antiforms, whereas negative values are observed on synforms. Boundaries between these values are loci of zero curvature and they correspond to normal sections of extreme curvature, equal magnitude, but opposite sign.

# Chapter 4

## Moonform Classification

The five types of curvatures that we are concerned with in this work are used to define classification schemes of surface forms (or landforms, when referring to topographic surfaces). The classifications distinguish local attributes of a surface regarding geometric forms (e.g. domes, basins or saddles), as well as geometric properties of a surface that are physically interpretable (e.g. acceleration or convergence of gravity-driven flows). We recall that although we are also applying these techniques to non-topographic surfaces, we are limiting ourselves to treat only mathematical and statistical aspects, and that interpretations are out of the scope of this work.

The classification schemes used in this chapter are defined in [6] and [3] and we refer the reader to those works for details. Classes are associated *exclusively* to the signum of each curvature, i.e., whether it is negative, zero, or positive. The schemes can be described as follows:

**Gauss Classification:** It is based on the signa of Gauss (K) and mean (H) curvatures. Only four out of eight possible classes are considered main types:

1. Hills:  $K>0, H>0$ .
2. Depressions:  $K>0, H<0$ .
3. Convex Saddles:  $K<0, H>0$ .
4. Concave Saddles:  $K<0, H<0$ .

**Efremov-Krcho Classification:** It is based on the signa of vertical ( $kv$ ) and horizontal ( $kh$ ) curvatures. Only four out of nine possible classes are considered main types:

1. Dissipation:  $kv>0, kh>0$ .
2. Convex-concave Transit:  $kv>0, kh<0$ .
3. Concave-convex Transit:  $kv<0, kh>0$ .
4. Accumulation:  $kv<0, kh<0$ .

**Shary Classification:** It is based on the signa of Gauss (K), mean (H), vertical ( $kv$ ), horizontal ( $kh$ ), and difference (E) curvatures. Only 12 out of 96 possible classes are considered main types:

1. Hills:  $K>0, H>0, E>0$ .
2. Hills:  $K>0, H>0, E<0$ .
3. Depressions:  $K>0, H<0, E>0$ .
4. Depressions:  $K>0, H<0, E<0$ .
5. Convex Saddles:  $K<0, H>0, kh>0, kv>0, E>0$ .
6. Convex Saddles:  $K<0, H>0, kh>0, kv>0, E<0$ .

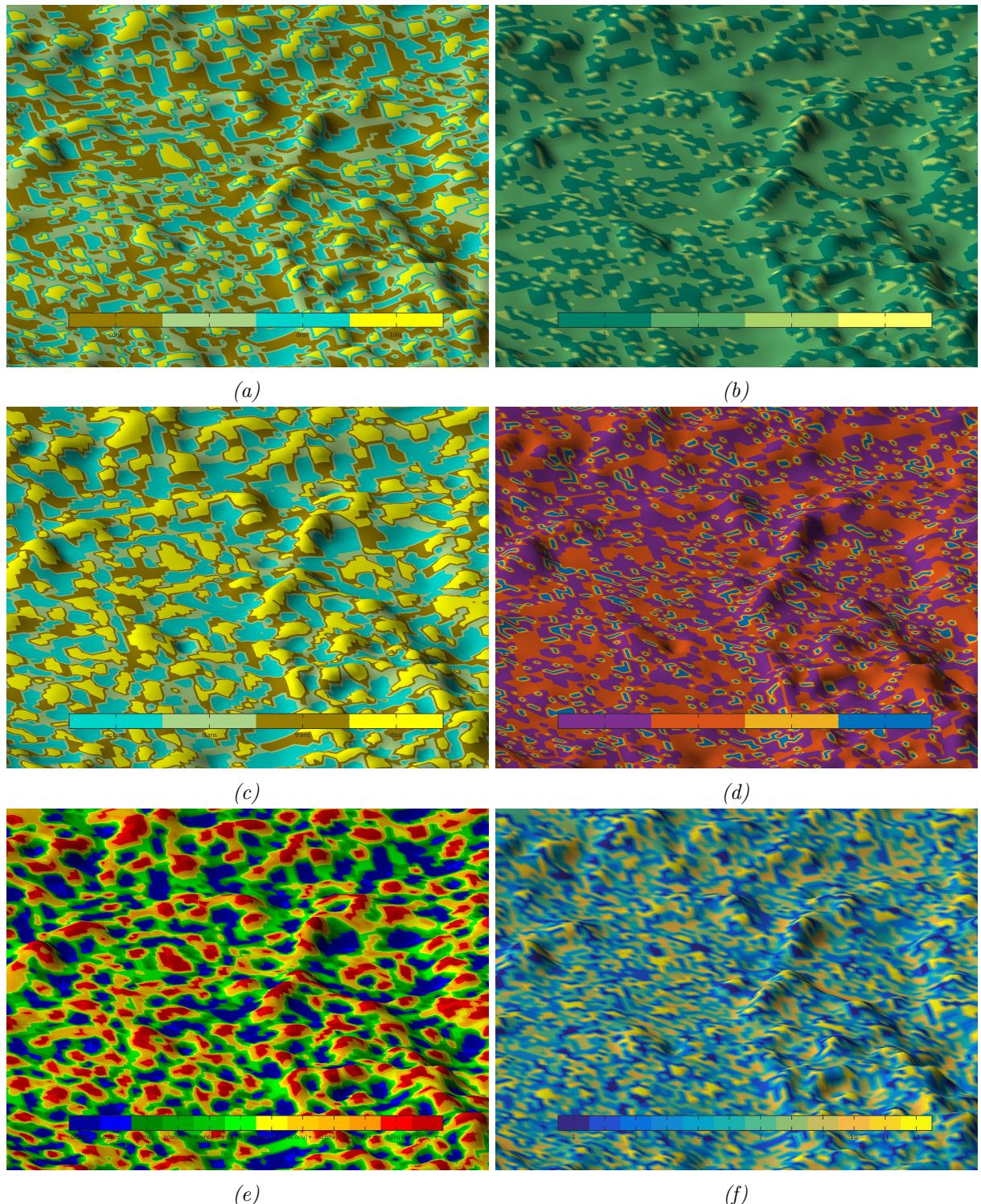


Figure 4.1: Classifications on Topography: (a) Gauss; (b) 4-means with  $K$  and  $H$ ; (c) Efremov-Krcho; (d) 4-means with  $kh$  and  $kv$ ; (e) Shary; (f) 12-means with all curvatures. Detail at maria Imbrium and Serenitatis. Data aspect ratio [1 1 500].

7. Convex Saddles:  $K < 0, H > 0, kh > 0, kv < 0$ .
8. Convex Saddles:  $K < 0, H > 0, kh < 0$ .
9. Concave Saddles:  $K < 0, H < 0, kh > 0$ .

10. Concave Saddles:  $K < 0$ ,  $H < 0$ ,  $kh < 0$ ,  $kv > 0$ .
11. Concave Saddles:  $K < 0$ ,  $H < 0$ ,  $kh < 0$ ,  $kv < 0$ ,  $E > 0$ .
12. Concave Saddles:  $K < 0$ ,  $H < 0$ ,  $kh < 0$ ,  $kv < 0$ ,  $E < 0$ .

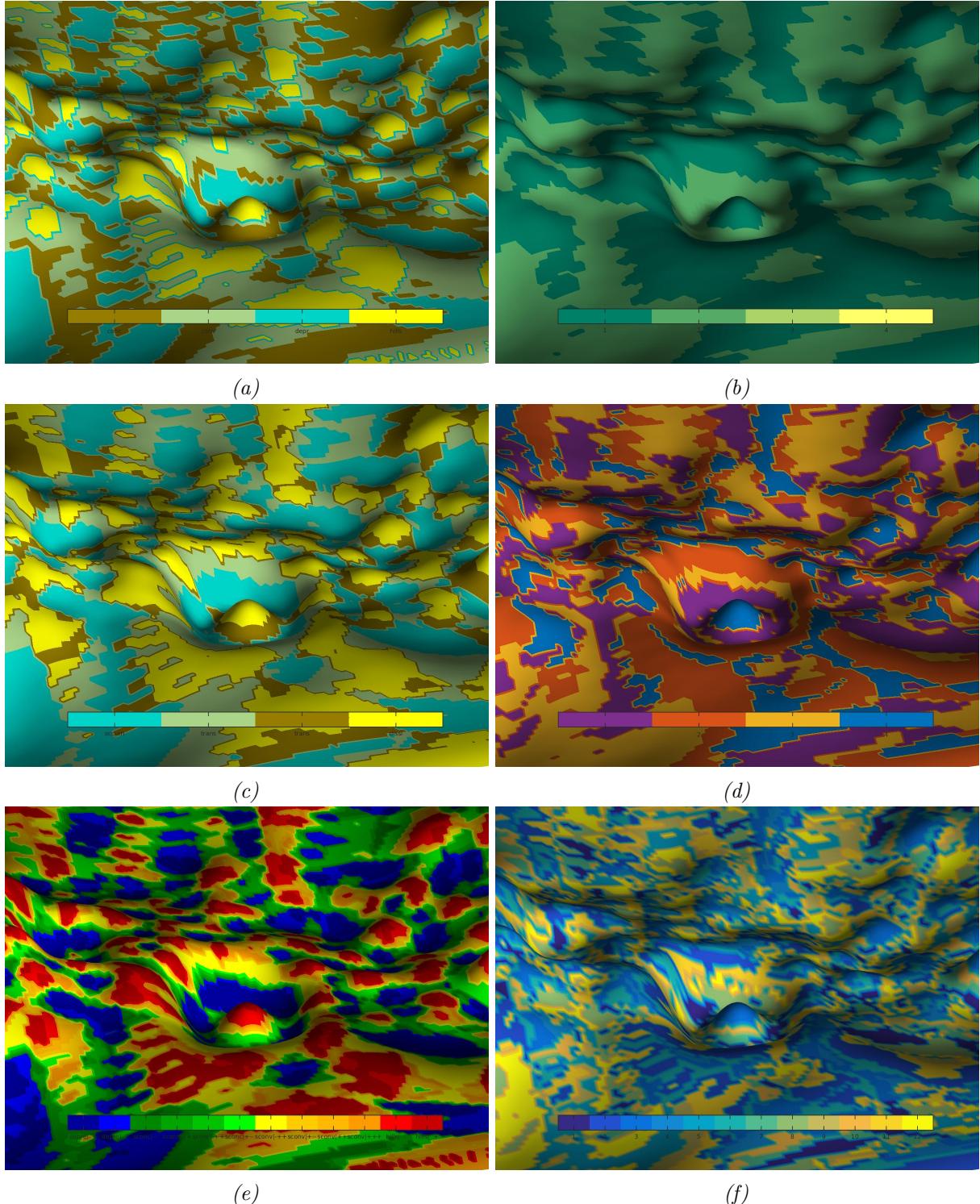


Figure 4.2: Classifications on Geoid Anomalies: (a) Gauss; (b) 4-means with  $K$  and  $H$ ; (c) Efremov-Krcho; (d) 4-means with  $kh$  and  $kv$ ; (e) Shary; (f) 12-means with all curvatures. Detail at maria Imbrium and Serenitatis. Data aspect ratio [1 1 20].

A couple of remarks are in place regarding computer implementations of the schemes described

above. First, there is an inherent difficulty in deciding the signum of a quantity when its magnitude is too small, so that even if the quantity is positive, for example, it could be considered practically a zero. For a given application, it is always possible to define a zero-centered interval, in such a way that any number included in the interval is assumed to be a zero value, but this is done with some degree of arbitrariness. In our work, we decided to rely on the MATLAB built-in function `sign`.

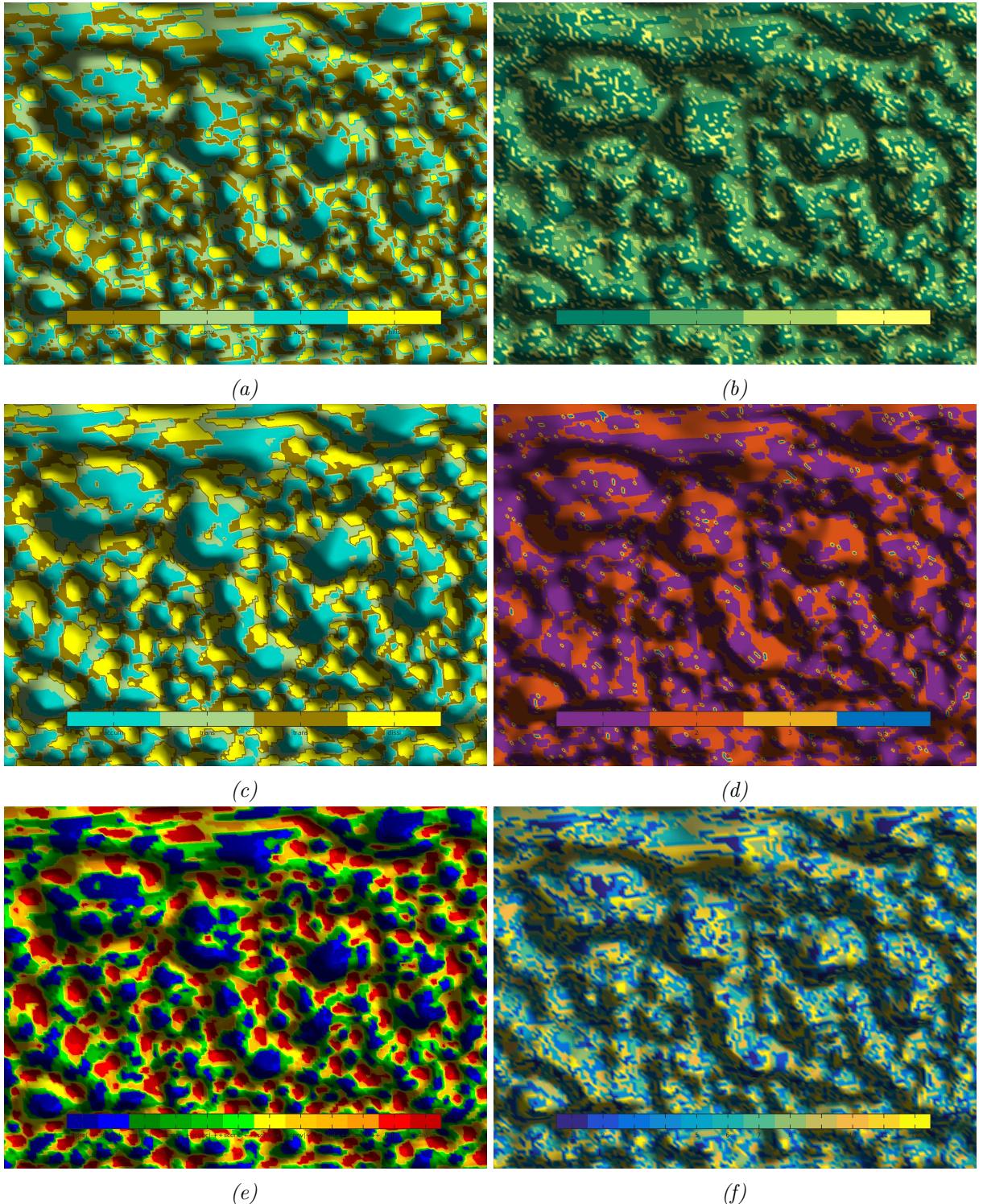


Figure 4.3: Classifications on Crustal Thickness: (a) Gauss; (b) 4-means with  $K$  and  $H$ ; (c) Efremov-Krcho; (d) 4-means with  $kh$  and  $kv$ ; (e) Shary; (f) 12-means with all curvatures. Detail at maria Imbrium and Serenitatis. Data aspect ratio [1 1 1250].

Second, the arithmetic ranges of curvature variables can be very large. For visualization purposes it can be convenient to modify the data aspect ratio, but this must be made in an instructed manner in order to avoid misinformation. In our work, we did not follow criteria other than obtaining a pleasant representation of data, and this may have had the side-effect of flattening some surface areas that may be otherwise uneven.

The remaining of this chapter presents our results of classification of all three lunar surfaces in our study.

## 4.1 Gauss Classification

Figures 4.1a, 4.2a, and 4.3a depict the results of Gauss classification for the Topography, Geoid Anomalies, and Crustal Thickness surfaces respectively, and figures 4.1b, 4.2b, and 4.3b, depict corresponding 4-means clustering with K and H. All visualizations use a 3D Plate Carrée projection.

For each surface we observe that Gauss classes correspond to surface features: yellow classes are hills, blue classes are depressions, and green and brown classes are saddles, convex and concave, respectively. There are, however, classified surface patches that do not seem to correspond to their predicted features, e.g. yellow patches that do not look like hills, but this is likely due to our choices of data aspect ratios and lack of zero-centered interval.

Compared to Gauss classification, k-means over-represents hills and depressions and under-represents saddles. Such clustering seems to discriminate well prominences and depressions, and could be an approximation toward discriminating ridges and valleys.

## 4.2 Efremov-Krcho Classification

Figures 4.1c, 4.2c, and 4.3c depict the results of Efremov-Krcho classification for the Topography, Geoid Anomalies, and Crustal Thickness surfaces respectively, and figures 4.1d, 4.2d, and 4.3d, depict corresponding 4-means clustering with kv and kh. All visualizations use a 3D Plate Carrée projection.

For each surface we observe that Efremov-Krcho classes correspond to surface flow dynamics: yellow classes are dissipation areas, blue classes are accumulation areas, and green and brown classes are transition areas, convex and concave, respectively. There are, however, classified surface patches that do not seem to correspond to their predicted features, e.g. yellow patches that look flat, but this is likely due to our choices of data aspect ratios and lack of zero-centered interval.

Compared to Efremov-Krcho classification, k-means seems to show a tendency to represent dissipation and accumulation, but does not perform consistently. The representation in the case of Geoid Anomalies seems better than that in the other two cases, where dissipation seems under-represented and accumulation over-represented.

## 4.3 Shary Classification

Figures 4.1e, 4.2e, and 4.3e depict the results of Shary classification for the Topography, Geoid Anomalies, and Crustal Thickness surfaces respectively, and figures 4.1f, 4.2f, and 4.3f, depict corresponding 12-means clustering with kv and kh. All visualizations use a 3D Plate Carrée projection.

For each surface we observe that Shary classes correspond to combinations of form and flow attributes. For instance, from the point of view of forms: red classes are hills, blue classes are depressions, and green and yellow classes are saddles, concave and convex, respectively. There are, however, classified surface patches that do not seem to correspond to their predicted attributes,

e.g. red patches that look flat, but this is likely due to our choices of data aspect ratios and lack of zero-centered interval.

Compared to Shary classification, k-means seems to show a tendency to represent convexity and concavity, but does not perform consistently. The representation in the case of Geoid Anomalies seems better than that in the other two cases, where it is difficult to match clusters with corresponding Shary classes.

# Chapter 5

## Considerations on Discriminant Analysis

In the previous chapter we applied three different types of classifications over moon surfaces based on the signs of five different types of curvatures, and obtained consistent results with theory. In addition, we contrasted the results with clustering of points over the surfaces applying k-means discriminant analysis. In this chapter we discuss the quality of these analyses, as well as alternatives.

### 5.1 Quality of Discrimination with k-means

As an example of the quality of discrimination of the k-means method, we computed silhouette plots for each of the three classification schemes in the case of Geoid Anomalies.

Silhouette is a measure of intra-cluster similarity relative to inter-cluster difference. The silhouette value of an observation  $x_i$  is the quantity  $(b_i - a_i) / \max(b_i, a_i)$ , where  $a_i$  is the average distance from  $x_i$  to all other observations in its cluster, and  $b_i$  is the minimum of the average distances of  $x_i$  to observations in each other cluster. The metric used is the same employed in the classification algorithm.

Values of silhouettes range in the interval  $[-1, 1]$ . A silhouette close to one indicates that the observation is well matched to observations of its own cluster, and well separated from observations of other clusters. We deem of 0.6 as an acceptable value of a silhouette to indicate well matching. When many observations have high silhouettes then data are considered well discriminated, whereas if many observations have low or negative silhouettes then it is an indication that the discrimination is inappropriate. The latter may be due to the use of too few or too many clusters in the classification, or the use of an inappropriate metric.

Figure 5.1b shows the silhouettes for Geoid Anomalies after performing a 4-means discriminant analysis with Gauss and mean curvatures and the cosine metric.

We notice that almost all observations were grouped as clusters either 1 or 2 and that these clusters are perfectly discriminated. The mean of clustering is 0.99983. When comparing this clustering to the Gauss classification in Figure 4.2a, it can be seen that cluster 1 seems to correspond to the union of classes *hill* and *convex saddles* and cluster 2 seems to correspond to the union of classes *depressions* and *concave saddles*. The result suggests that maybe the cosine metric is not appropriate for this discrimination.

Figure 5.1d shows the silhouettes for Geoid Anomalies after performing a 4-means discriminant analysis with horizontal and vertical curvatures and the cosine metric.

Relative to a silhouette value of 0.6, we observe that the clustering is acceptable, with clusters 1 and 2 better discriminated. The mean of clustering is 0.712723. The majority of observations (more than half) within each cluster appear to be well matched, but each cluster has too some observations poorly matched. When comparing this clustering to the Efremov-Krcho classification

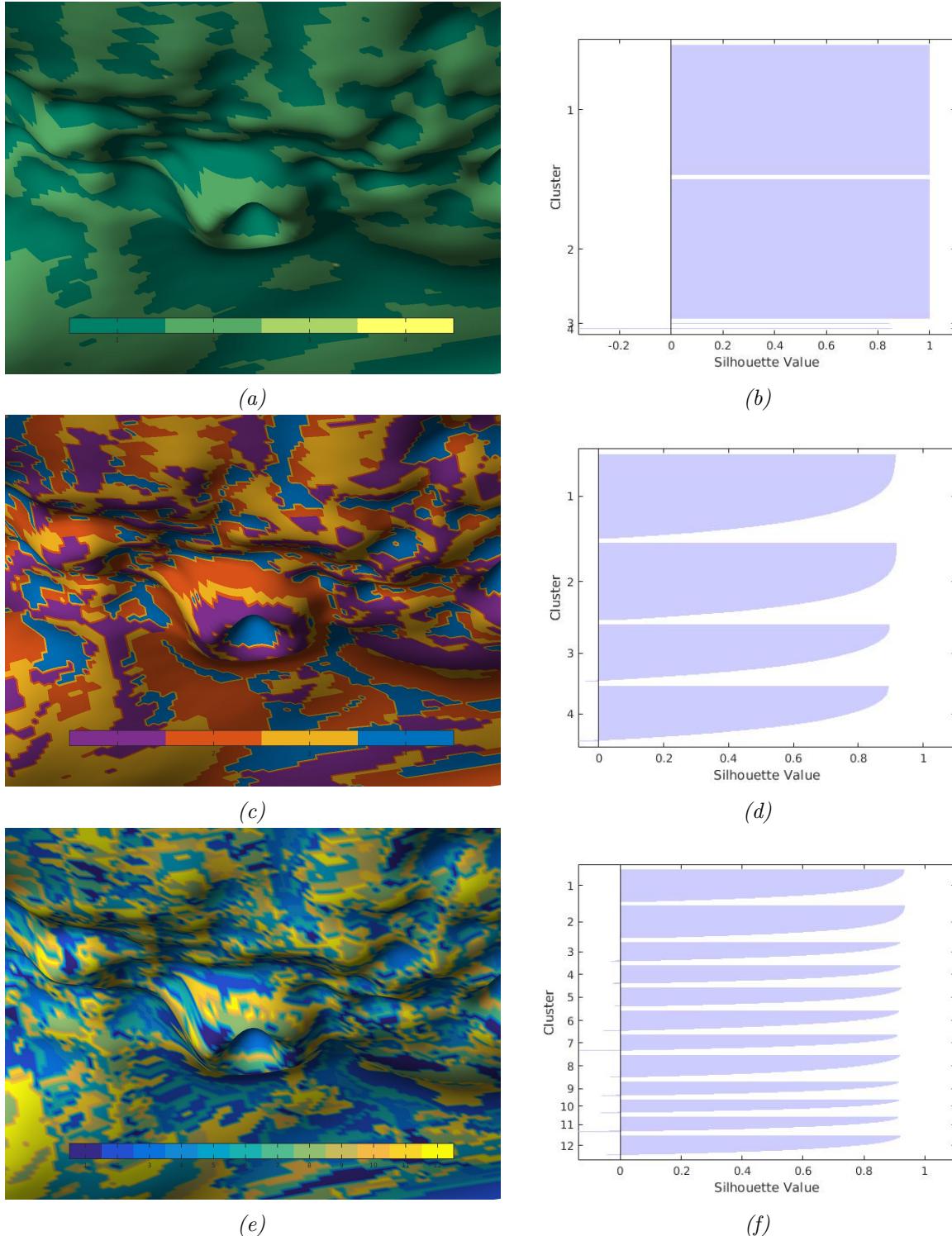


Figure 5.1: Silhouettes for Geoid Anomalies. (a) & (b) 4-means with  $K$  and  $H$ ; (c) & (d) 4-means with  $kh$  and  $kv$ ; (e) & (f) 12-means with  $K$ ,  $H$ ,  $kh$ ,  $kv$ , and  $E$ . In all cases, the cosine metric was used.

in Figure 4.2c, it can be seen that cluster 1 seems to be a subset of class *accumulation*, cluster 2 is related to class *dissipation*, cluster 3 seems related to both *accumulation* and *transition* (green), and cluster 4 seems related to both *dissipation* and *transition* (brown). In spite of acceptable discrimination, the result is not sound.

Figure 5.1f shows the silhouettes for Geoid Anomalies after performing a 12-means discriminant

analysis with Gauss, mean, horizontal, vertical, and difference curvatures and the cosine metric. Relative to a silhouette value of 0.6, we observe that most of clusters seem to be well discriminated except for clusters 7, 9, 10, and 11, since in these cases the majority of observations (more than half) appear to be poorly matched. The mean of clustering is 0.692905. When comparing this clustering to the Shary classification in Figure 4.2e, it can be seen that cluster 1 seems related to *lower concave*, cluster 2 seems related to *hills*, and cluster 8 seems related to *lower depressions*. The analysis suggests that the discrimination may be acceptable, but not sound.

## 5.2 Data Standardization and other Metrics

The analyses we have presented were made with non standardized data. Although we did experiment with k-means discrimination on standardized curvatures, there were no sensible results that we could exhibit. In fact, we met two kinds of problems: On the one hand, curvature values were scaled to a point that we could not find diverging color maps that would allow us to replicate visualizations of the original data. On the other hand, we could not execute the k-means algorithm with the cosine metric, and the only metric that yielded no execution errors was the squared euclidean distance, that grouped almost all observations into a single cluster.

# Chapter 6

# Principal Component Analysis

It is known that there are dependencies between the five type of curvatures that we are dealing with ([3] pp. 14,15), and also that three of them are sufficient to describe a surface ([6], and see also [3] pp. 331). It is therefore natural to perform a Principal Component Analysis to obtain a new set of variables that are independent but maximize variance, and then make a discriminant analysis. We have done so, once more using Geoid Anomalies as an example.

## 6.1 Analysis

Figure 6.1 shows the result of Principal Component Analysis on [kv kh E K H], where curvatures were previously logarithmically scaled and then standardized.

The box-plot in Figure 6.1a shows that Gauss curvature has the smallest variability, and that mean and difference curvature have similar variability. The latter can be explained by the fact that both curvatures can be expressed as linear combinations of vertical and horizontal curvature. In Figure 6.1b we observe that the first two principal components explain ca. 80% of the variance among all curvature information, and the first three components over 95% of the variance. There does not seem to be a clear difference in change of variance between the first and second components, and between the second and third components. To be sure, these are the components variances in decreasing order: 2.38669576929787, 1.68276260657926, 0.920571274261764, 0.00710532928495729, and 0.00286502057612739. We therefore decided to retain all first three components.

The retained components are then visualized in Figure 6.1c. We observe that indeed there is more variance along the first component than the second, along the second component than the third (although less), and along the first component than the third. In this last case, we can also see a non-linear correlation.

## 6.2 Discrimination

We performed subsequently a discriminant analysis upon the first three principal components. Figures 6.2a and 6.2b show the result of a 4-mean discrimination on the retained components, and Figures 6.2c and 6.2d show the result of a 12-mean discrimination.

In the case of 4-mean discrimination, we observe that relative to a silhouette value of 0.6 most of clusters seem to be acceptably discriminated, except for cluster 3. Cluster 1 appears to have the highest number of observations, following clusters 4, 2, and 3. The mean of clustering is 0.654478.

The result of this clustering seems closer to Efremov-Krcho classification (Figure 4.2c) than to Gauss classification (Figure 4.2a). It can be seen that cluster 4 seems related to *accumulation*, and cluster 3 seems related to *dissipation*. Cluster 1, being the most abundant seems over-representing class *green transition*. The analysis suggests that the discrimination is barely acceptable.

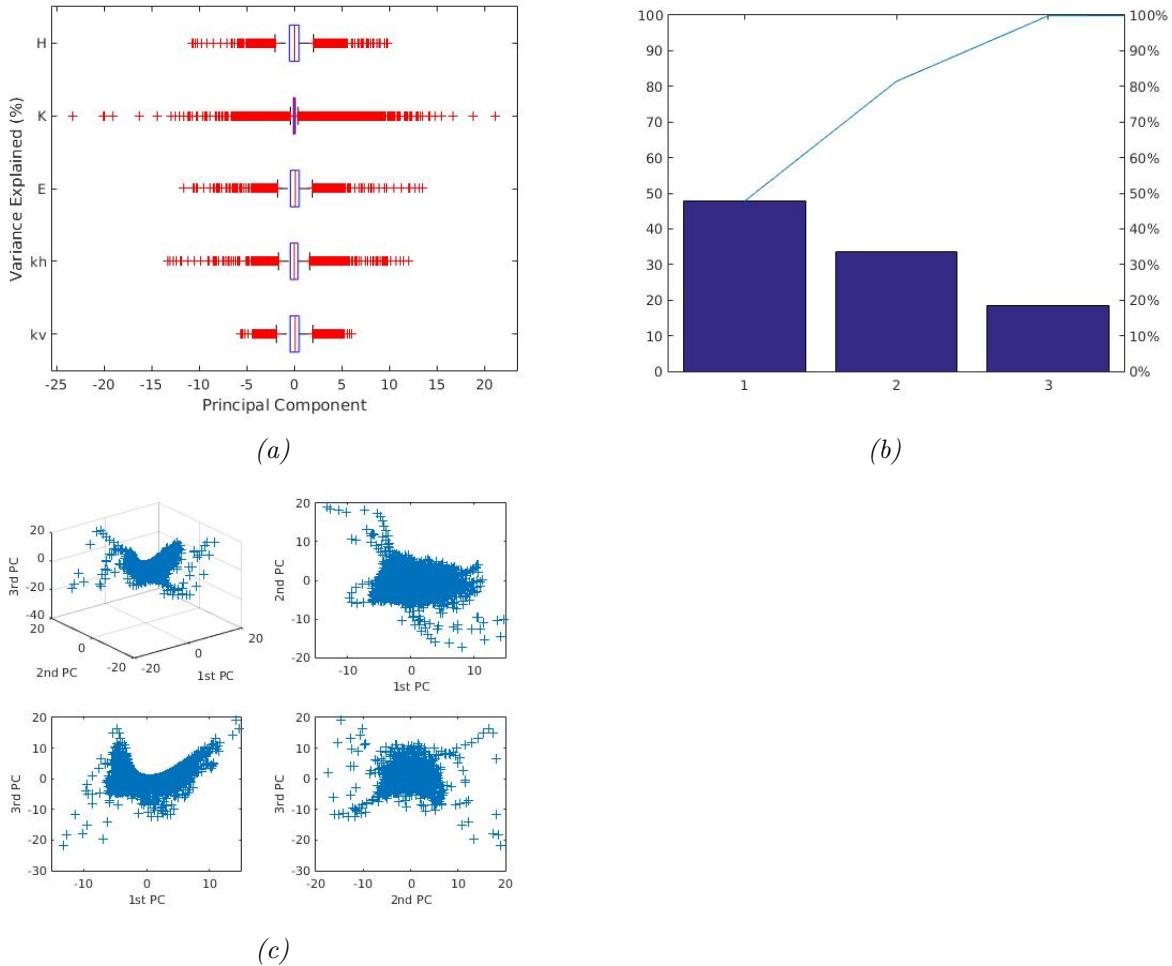


Figure 6.1: PCA on  $[kv\ kh\ E\ K\ H]$  and subsequent 4-means clustering for Geoid Anomalies. (a) Multivariate box-plot; (b) Screeplot with first three principal components; (c) Scores.

In the case of 12-mean discrimination, we observe that relative to a silhouette value of 0.6 all clusters seem poorly discriminated, except for cluster 11 and perhaps cluster 4. The mean of clustering is 0.546731.

When comparing this clustering to the Shary classification in Figure 4.2e, it can be seen that cluster 11 seems somewhat related to *lower convex*, and cluster 4 seems somewhat related to *depressions* and *lower concave*. The analysis suggests that the discrimination is poor.

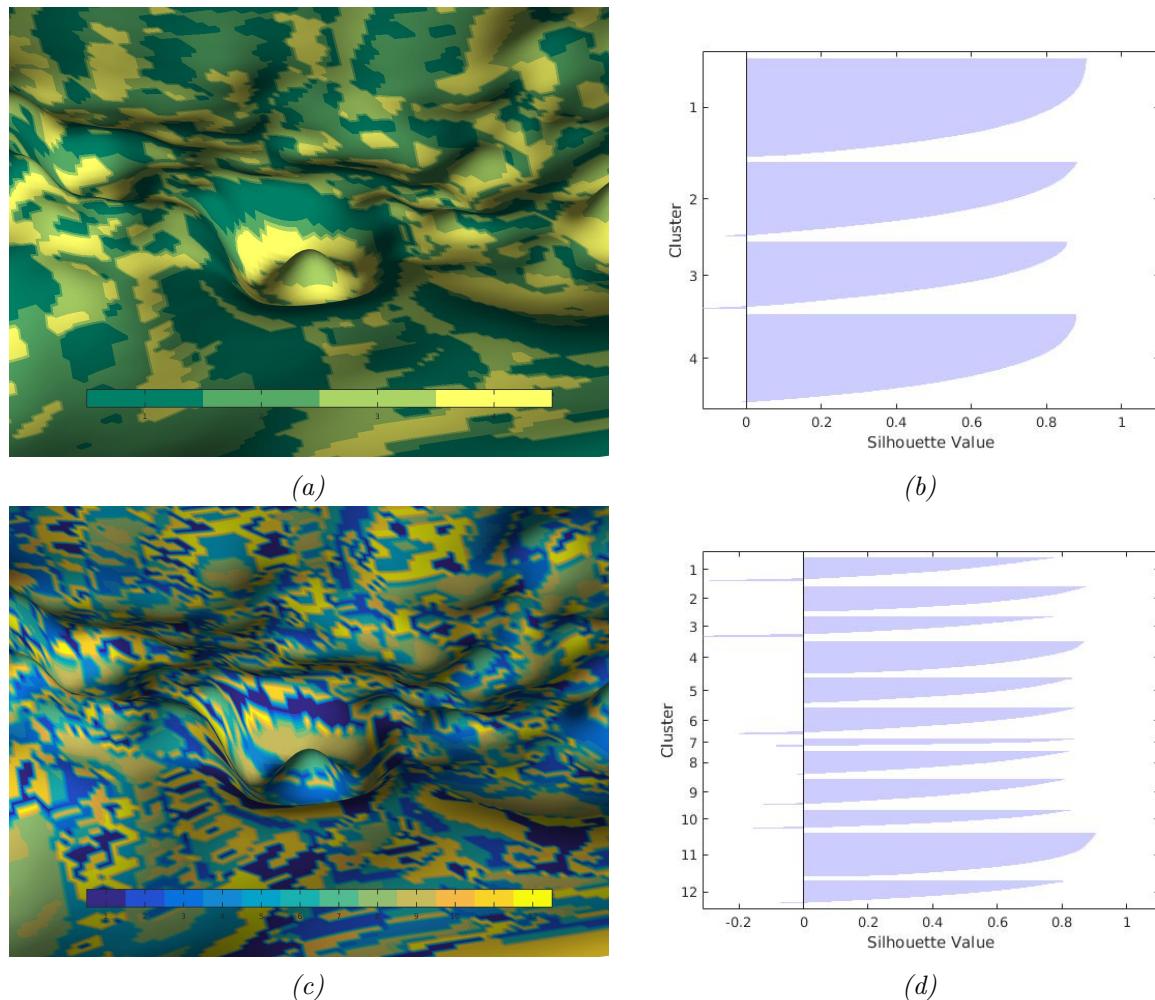


Figure 6.2: Discriminant analysis with retained components for Geoid Anomalies. (a)&(b) 4-means; (c)&(d) 12-means.

# Chapter 7

## Conclusions

In this work we applied techniques of Differential Geometry and Geomorphometry to study three surfaces that are defined on selenographic coordinates: lunar Topography, Geoid Anomalies, and Crustal Thickness. We performed three different types of classification of surface forms [3] [6]: Gauss Classification, Efremov-Krcho Classification, and Shary Classification. The focus of this work, however, was the application of multivariate statistical tools in the context of surface form classification, and in this respect we were concerned with the use of Discriminant Analysis and Principal Component Analysis. Below, we mention several issues of interest that were not discussed in the the main body of the report.

### On Geomorphometric classifications

In section 4 we presented geomorphic classifications on the three lunar surfaces under study, and contrasted the results to corresponding clusterings using the k-means method. We did not compare the results of the three geomorphic classifications themselves, since that was our focus. It is worth, nevertheless, to summarize some of our observations.

For each surface we noticed that: (a) Gauss hills are subsets of Efremov-Krcho dissipations; (b) Gauss hills correspond tightly to Shary hills; (c) Efrem dissipations correspond tightly to combined Shary hills and upper Shary convex saddles.

On another hand, and perhaps dually: (a) Gauss depressions are subsets of Efrem accumulations; (b) Gauss depressions correspond tightly to Shary depressions; (c) Efrem accumulations correspond tightly to combined Shary depressions and lower Shary concave saddles.

The above observations are qualitative and they should be verified numerically.

### On Statistical Tools

Several statistical tools were employed in this work and we would like to emphasize the value of some of them.

Gaussian smoothing helps to reduce image noise. We applied it to variable Topography and were able to see features on the lunar topographic surface that were otherwise difficult to recognize. Least Squares adjustment helps to fit a surface. We applied it to approximate partial derivatives and thereby calculate curvature values.

Histograms of variables help to design diverging color maps that enhance data representation, without obscuring information nor introducing artifacts. Although we were guided by the use of histograms we fell short of doing it programmatically.

Finally, similarities between observations and dependencies between variables can be examined with Discriminant Analysis and Principal Component Analysis. We discuss them next.

## On Discriminant Analysis

Discriminant analysis in the form of k-means is an important method to separate observations in groups according to certain measure of similarity. The results we obtained showed that relevant information can be revealed with this type of analysis, however it should be complemented with other methods. The problem of surface form classification is very particular in that only the signs of variables are meaningful for segmentation. The cosine metric, which was the metric used for measuring data similarity, does consider signs, but it also carries more information that hampers the classification process. It is likely that a k-means algorithm provided with a metric that takes into account only the signs of variables performs better.

## On Principal Component Analysis

In order to examine interdependency of variables and redundancy of information Principal Component Analysis is one among other statistical methods of importance. It can be used to reduce the dimensionality of data by creating a few new variables that are linear combinations of the original variables, that are not linearly correlated, and that maximize the variance of data. The advantages of the product of this analysis are however limited by the quality of a subsequent discriminant analysis upon the created new variables. In our work we were unable to overcome such limitation.

# Bibliography

- [1] Davies, M. E., Abalakin, V. K., Brahic, A., Bursa, M., Chovitz, B. H., Lieske, J. H., Seidelmann, P. K., Sinclair, A. T., and Tjuflin, Y. S., *Report of the IAU/IAG/COSPAR working group on Cartographic Coordinates and Rotational Elements of the Planets and Satellites: 1991*, Celestial Mechanics and Dynamical Astronomy, 53, 377-397, 1992.
- [2] Dickey, J., Faller, J.E., Newhall, X.X., Ricklefs, R.L., Ries, J.G., Shellus, P.J., Veillet, C., Whipple, A.L., Wiant, J.R., Williams, J.G., and Yoder, C.F., *Lunar Laser Ranging: A Continuing Legacy of the Apollo Program*, Science, 265, 482-490, 1994.
- [3] Florinsky, I. V., *Digital Terrain Analysis in Soil Science and Geology*, First Edition, Elsevier, 2012.
- [4] Lemoine, Frank G., NASA/Goddard Space Flight Center, Greenbelt, MD, *CLEM1 LUNAR GRAVITY V1.0*, release date 1996-01-01, PDS Geosciences Node, Washington University in St. Louis, <http://pds-geosciences.wustl.edu/missions/clementine/gravtopo.html>.
- [5] PDS Geosciences Node, Washington University in St. Louis, Clementine, <http://pds-geosciences.wustl.edu/missions/clementine/default.htm>.
- [6] Shary, P. A., *Land surface in gravity points classification by complete system of curvatures*, Mathematical Geology, 27, 1995. pp. 373-390.
- [7] Williams, J. G., Newhall, X.X., Standish, E.M., *Users of Ephemerides of the Earth-Moon System*, JPL Internal Memorandum, December 18, 1993.
- [8] Zuber, Maria T., Massachusetts Institute of Technology, Cambridge, MA, Smith, David E., NASA/Goddard Space Flight Center, Greenbelt, MD, and Neumann, Gregory A., John Hopkins University, Baltimore, MD, *CLEM1 LUNAR TOPOGRAPHY V1.0*, release date 1996-01-01, PDS Geosciences Node, Washington University in St. Louis, <http://pds-geosciences.wustl.edu/missions/clementine/gravtopo.html>.

# Appendix A

## Statistical Summaries of Curvatures

### Topography: Horizontal Curvature

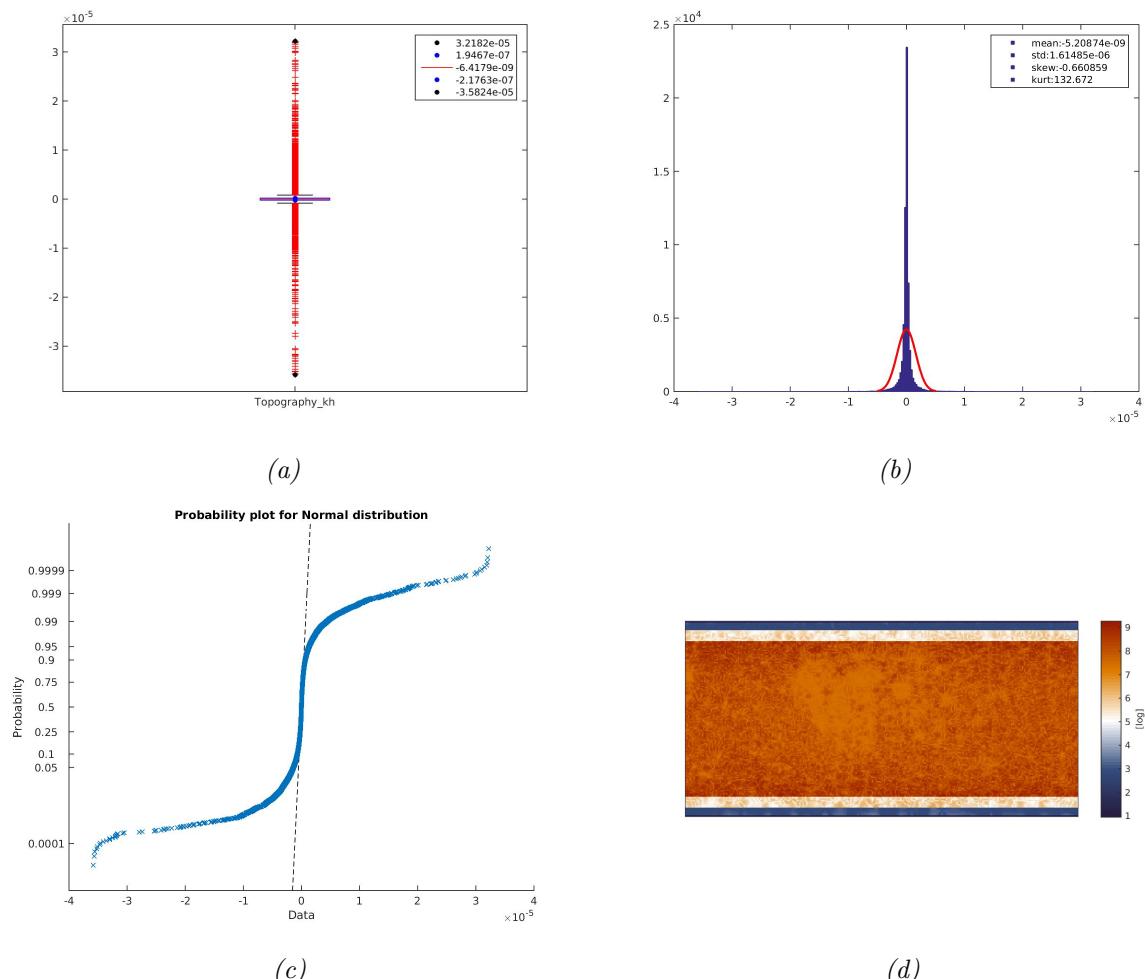


Figure A.1: Summary for Topography\_kh: (a) boxplot; (b) histogram; (c) probability; (d) leverages.

## Topography: Vertical Curvature

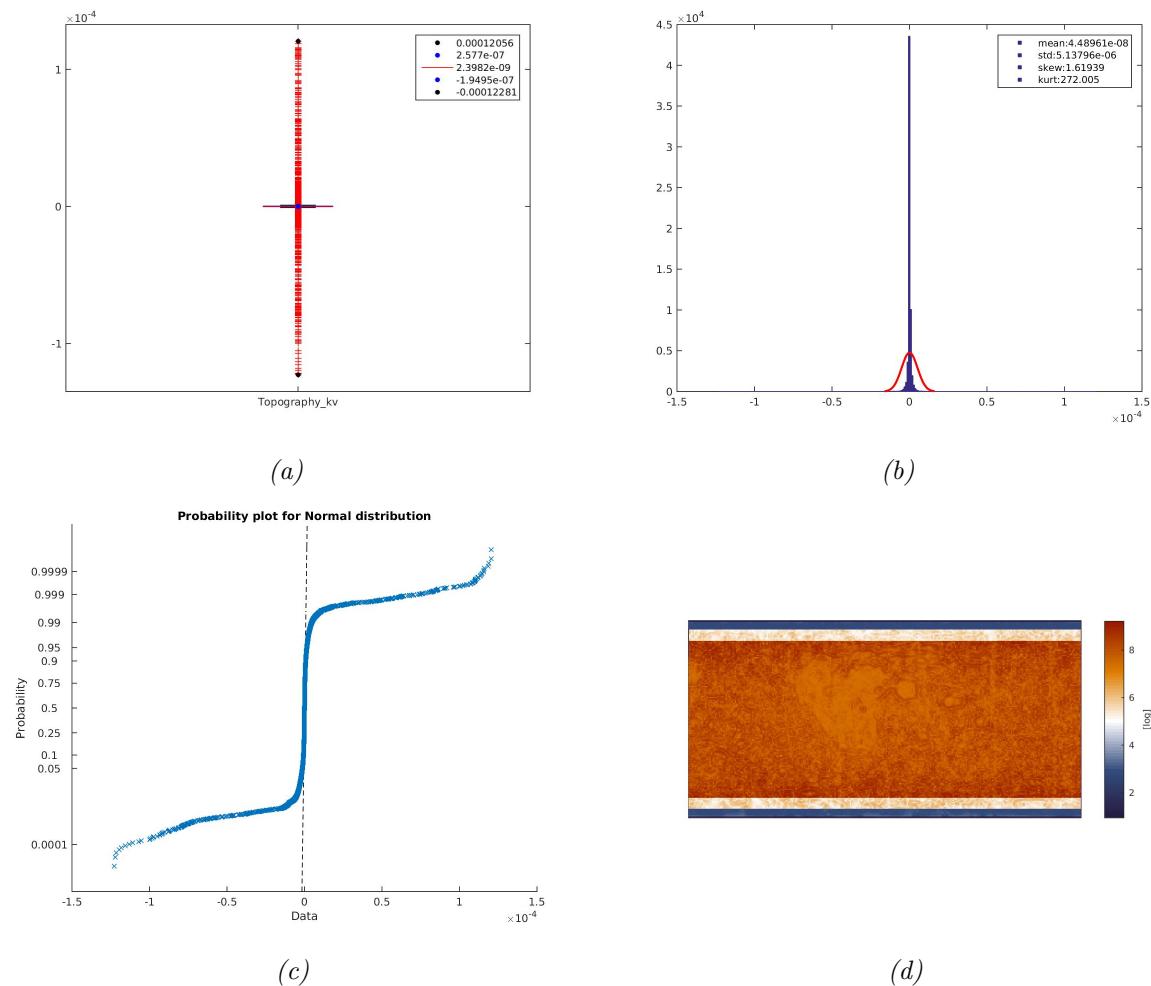


Figure A.2: Summary for Topography\_kv: (a) boxplot; (b) histogram; (c) probability; (d) leverages.

## Topography: Difference Curvature

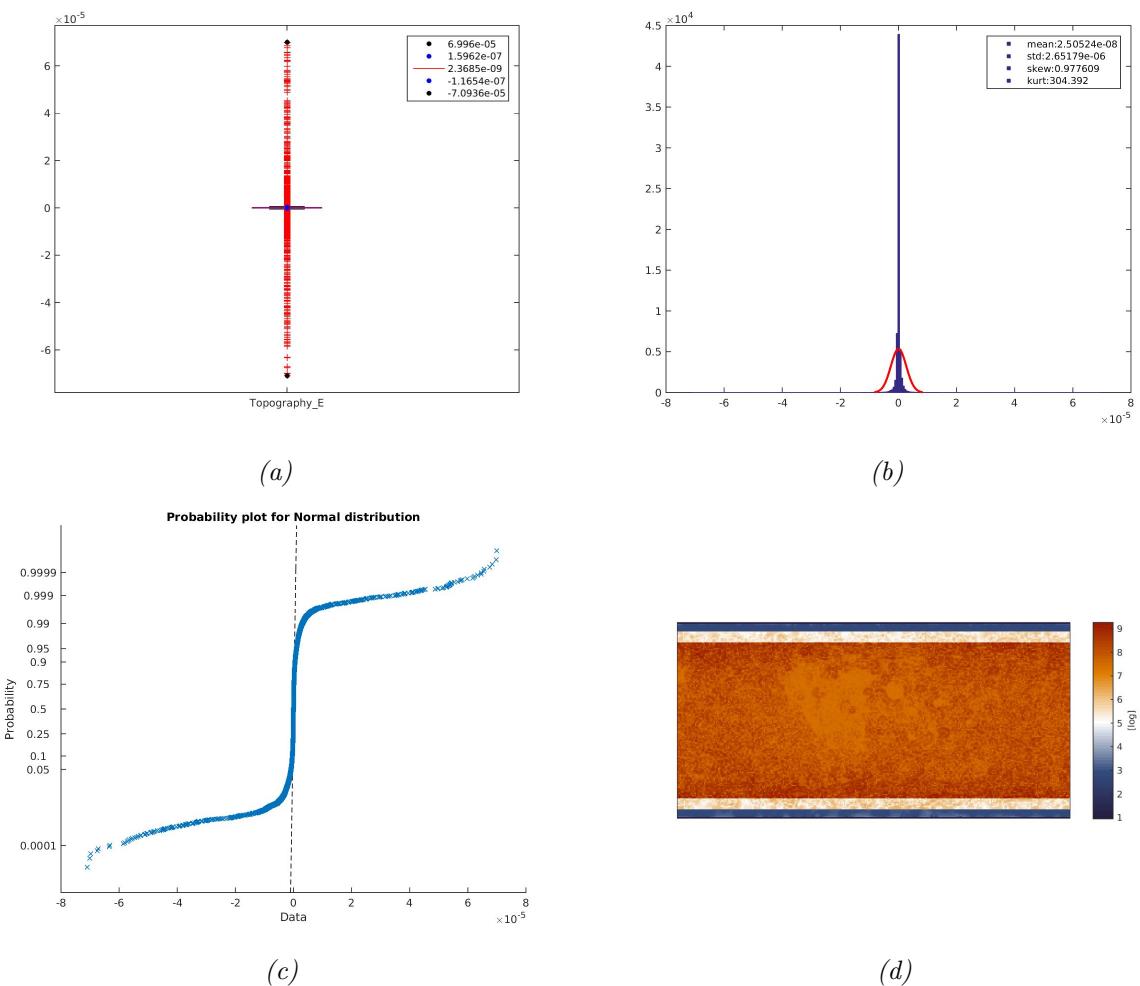


Figure A.3: Summary for `Topography_E`: (a) boxplot; (b) histogram; (c) probability; (d) leverages.

## Topography: Gauss Curvature

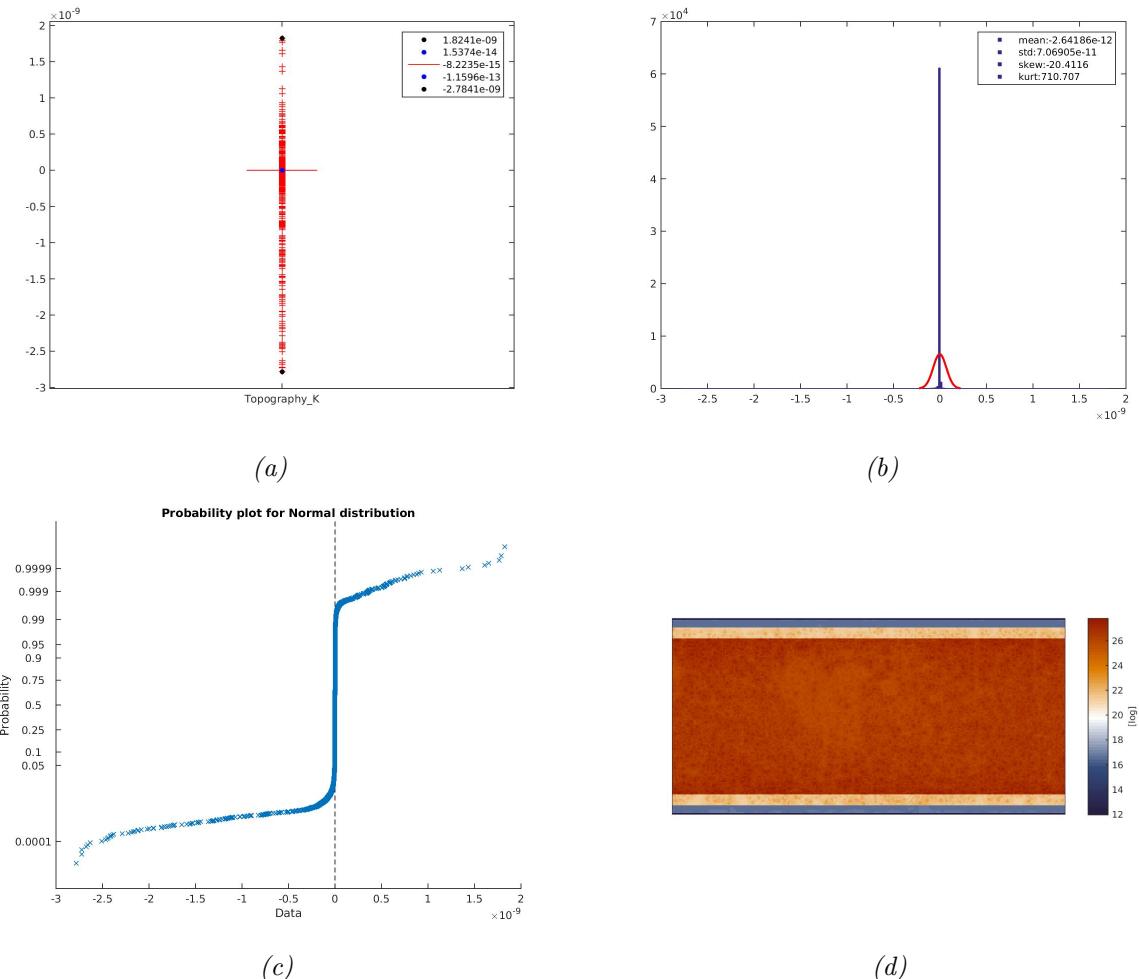


Figure A.4: Summary for Topography\_K: (a) boxplot; (b) histogram; (c) probability; (d) leverages.

## Topography: Mean Curvature

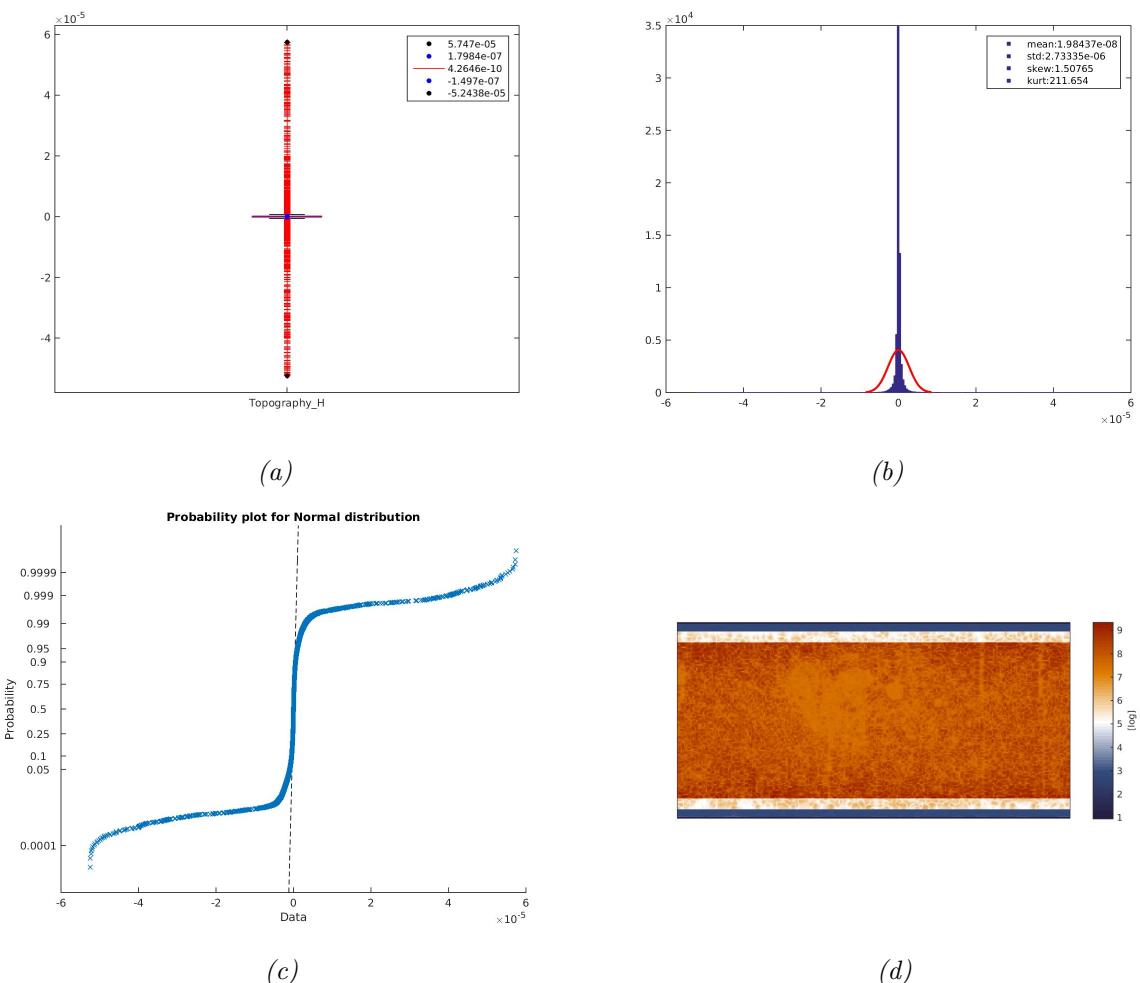


Figure A.5: Summary for Topography\_H: (a) boxplot; (b) histogram; (c) probability; (d) leverages.

## Geoid Anomalies: Horizontal Curvature

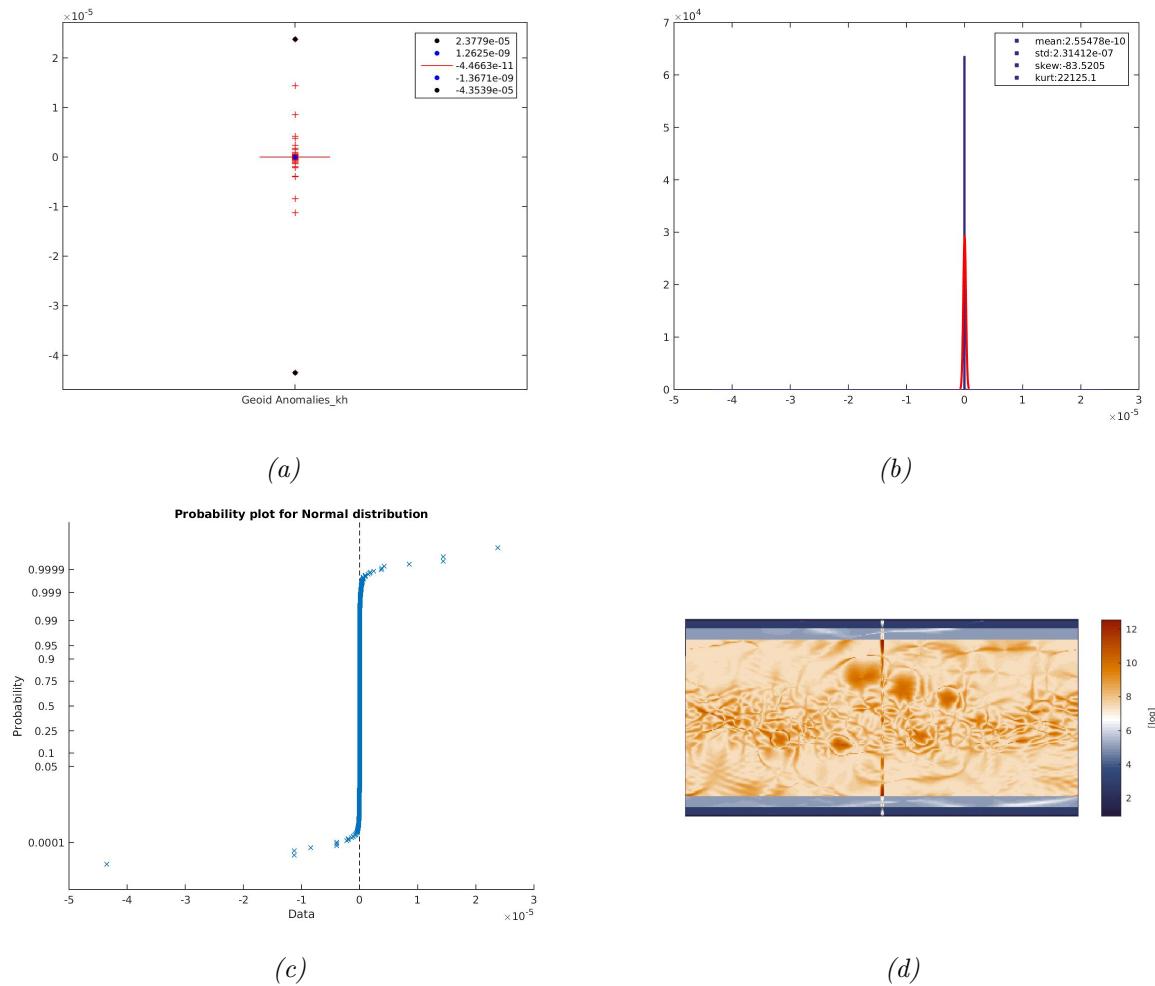


Figure A.6: Summary for GeoidAnomalies\_kh: (a) boxplot; (b) histogram; (c) probability; (d) leverages.

## Geoid Anomalies: Vertical Curvature

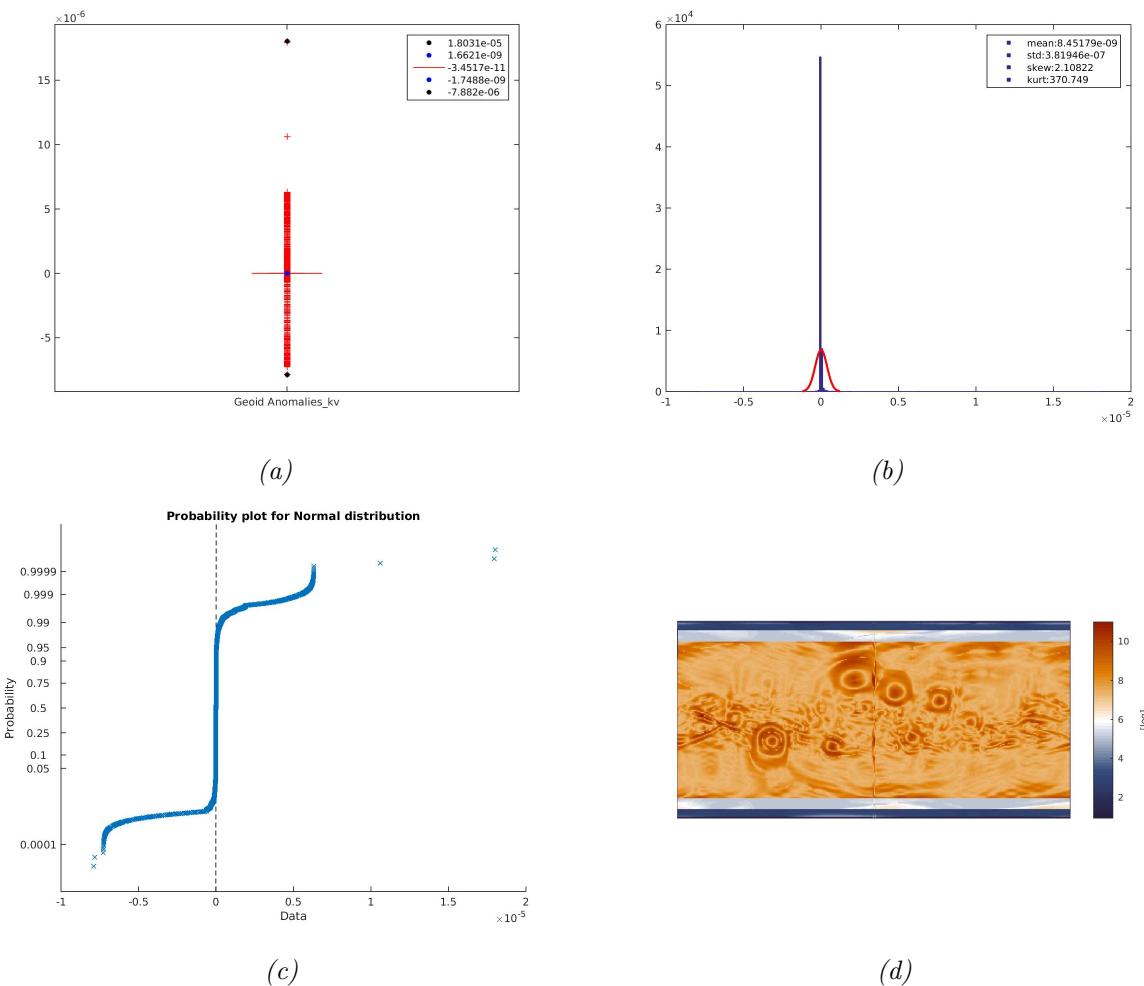


Figure A.7: Summary for GeoidAnomalies\_kv: (a) boxplot; (b) histogram; (c) probability; (d) leverages.

## Geoid Anomalies: Difference Curvature

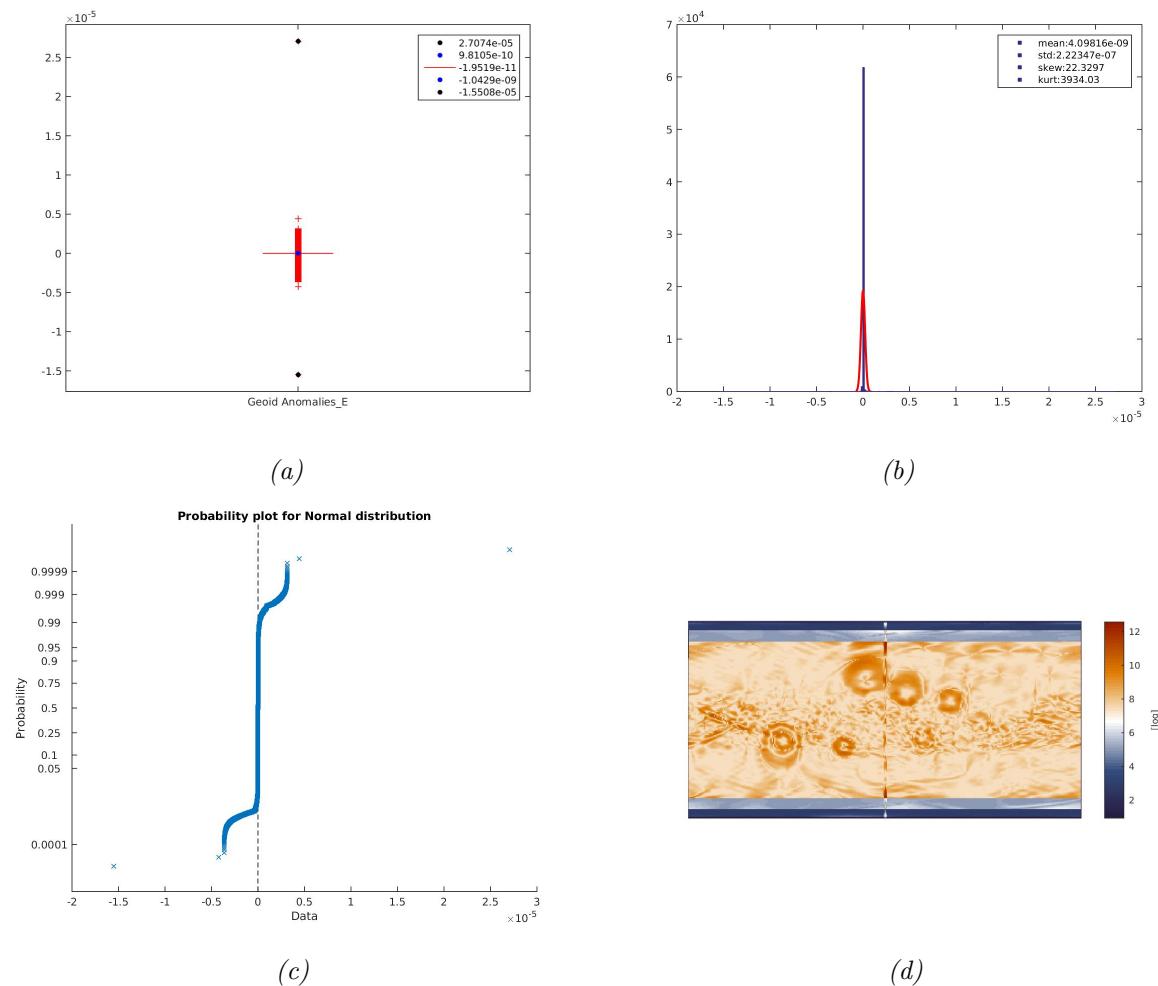


Figure A.8: Summary for *GeoidAnomalies\_E*: (a) boxplot; (b) histogram; (c) probability; (d) leverages.

## Geoid Anomalies: Gauss Curvature

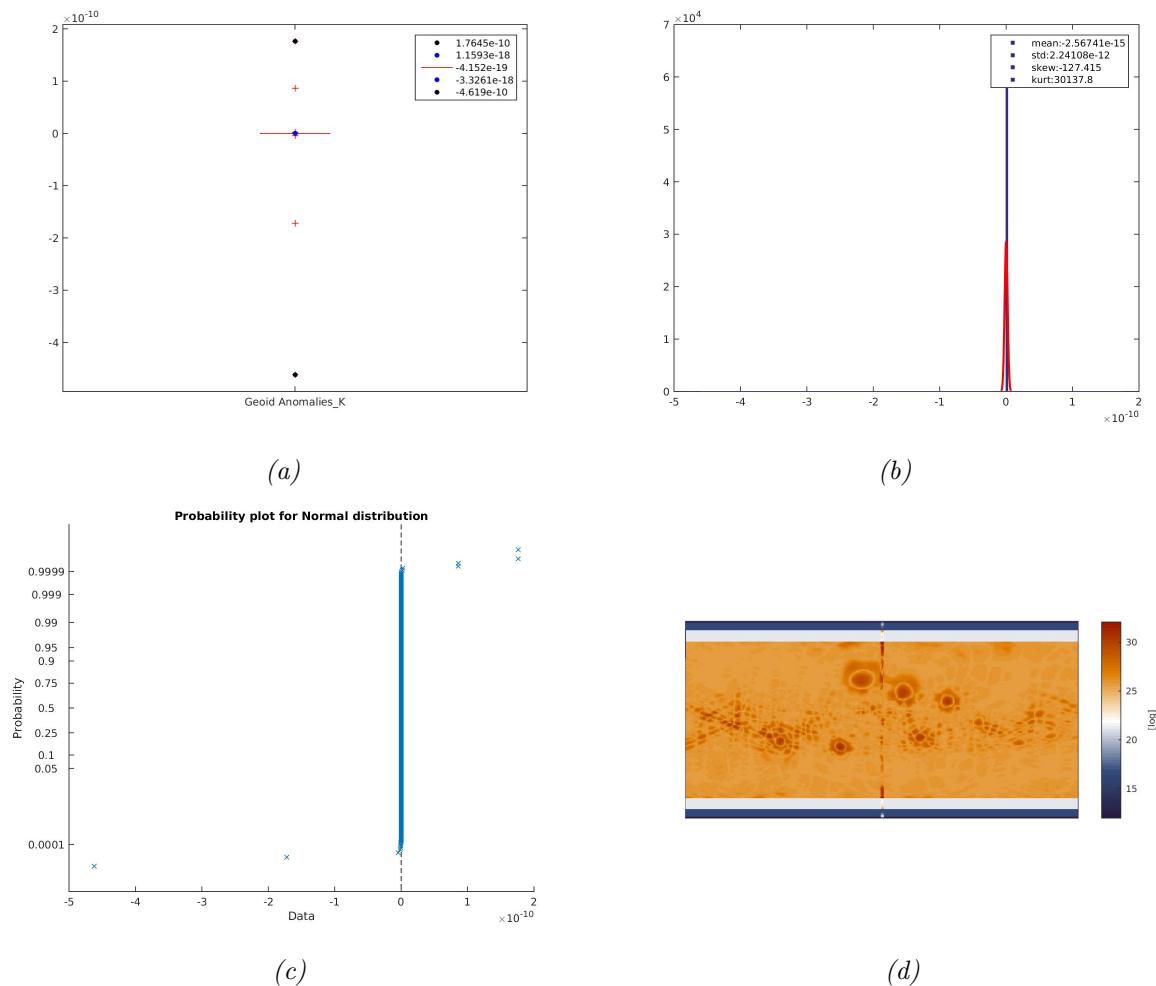


Figure A.9: Summary for GeoidAnomalies\_K: (a) boxplot; (b) histogram; (c) probability; (d) leverages.

## Geoid Anomalies: Mean Curvature

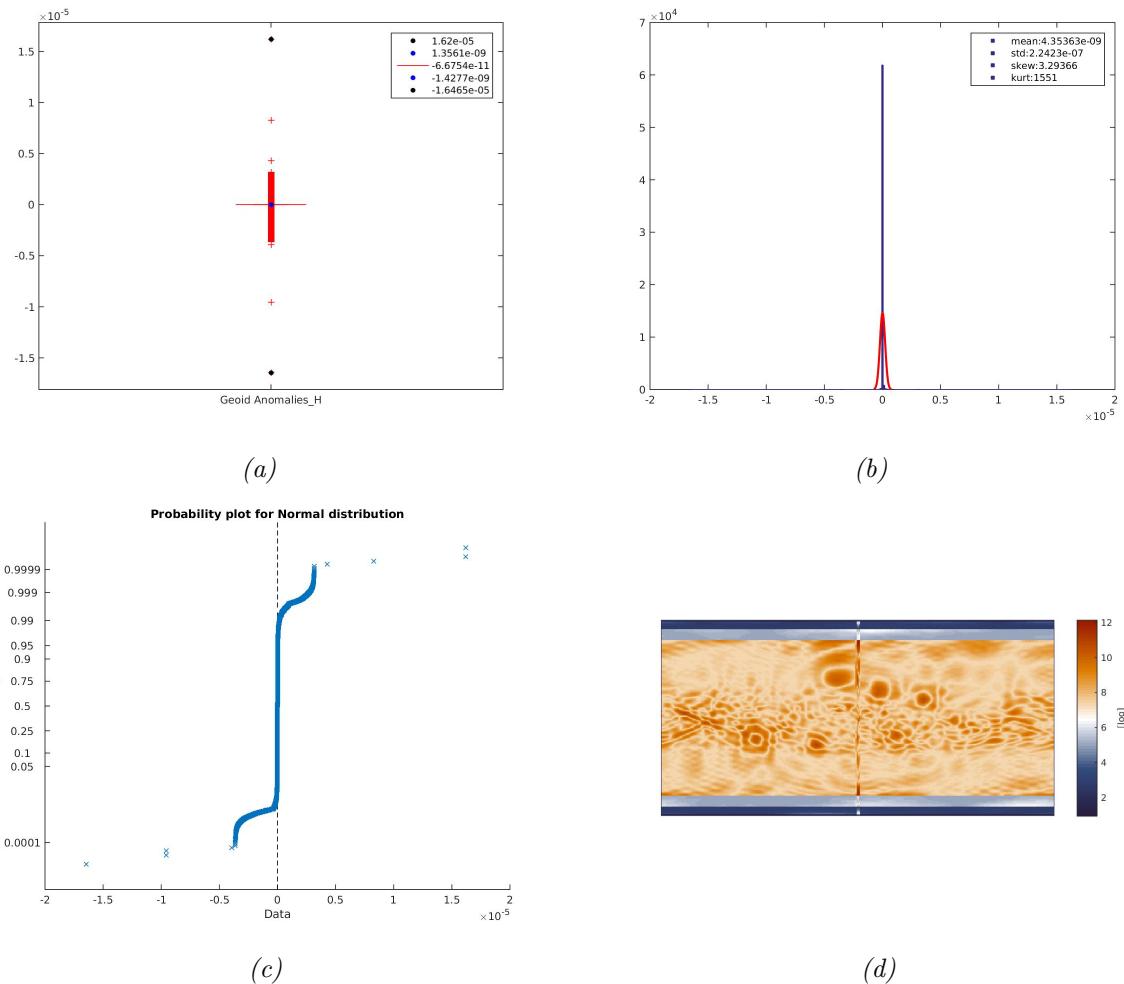


Figure A.10: Summary for GeoidAnomalies\_H: (a) boxplot; (b) histogram; (c) probability; (d) leverages.

## Crustal Thickness: Horizontal Curvature

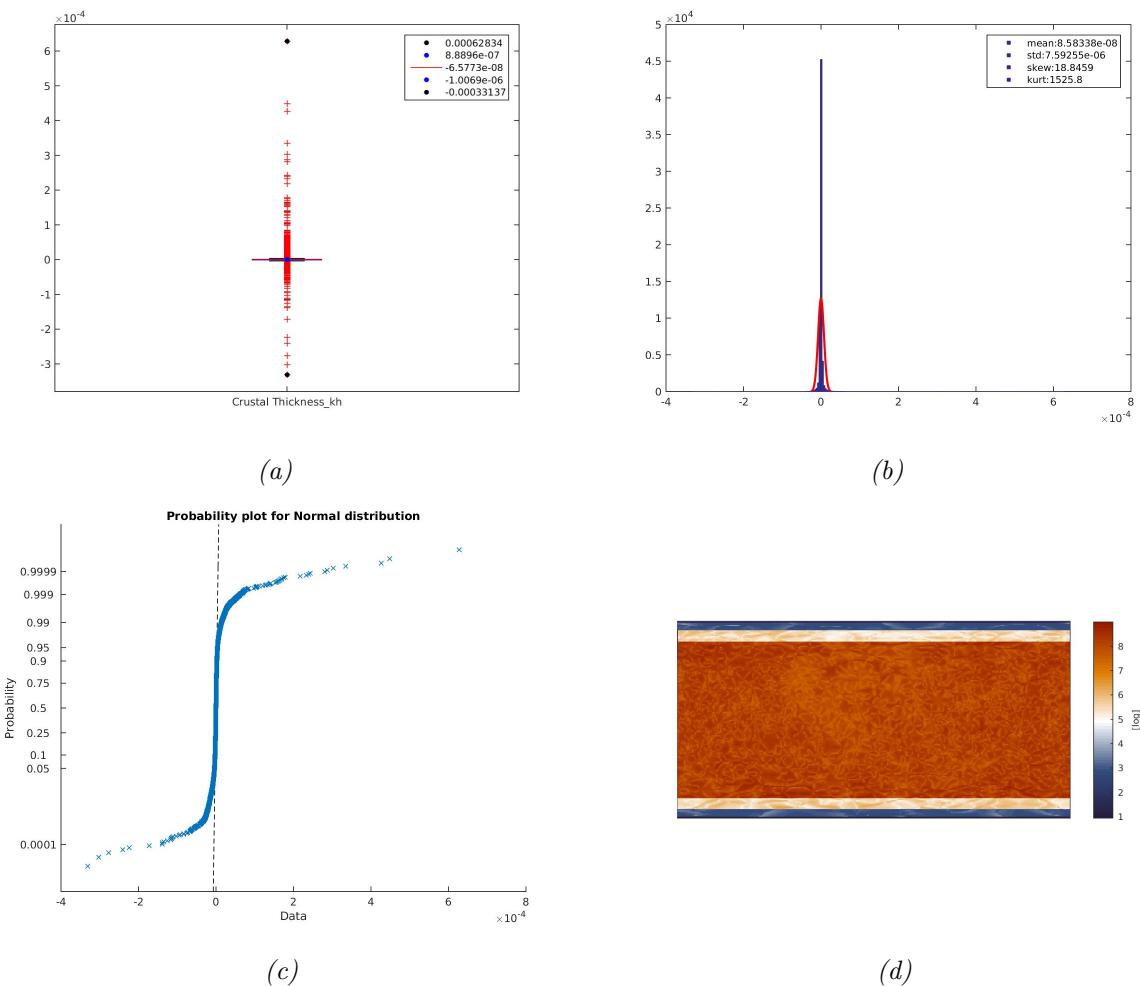


Figure A.11: Summary for CrustalThickness\_kh: (a) boxplot; (b) histogram; (c) probability; (d) leverages.

## Crustal Thickness: Vertical Curvature

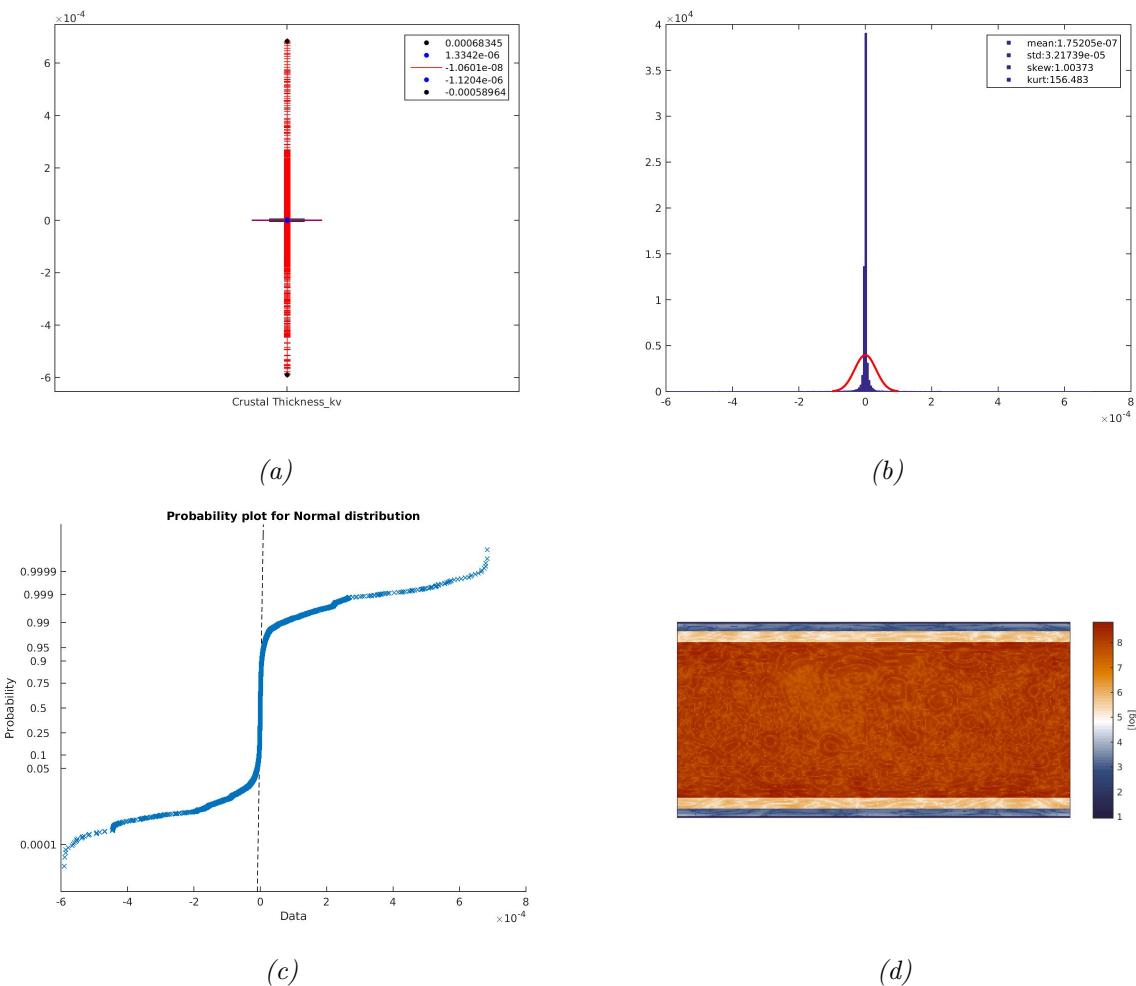


Figure A.12: Summary for CrustalThickness\_kv: (a) boxplot; (b) histogram; (c) probability; (d) leverages.

## Crustal Thickness: Difference Curvature

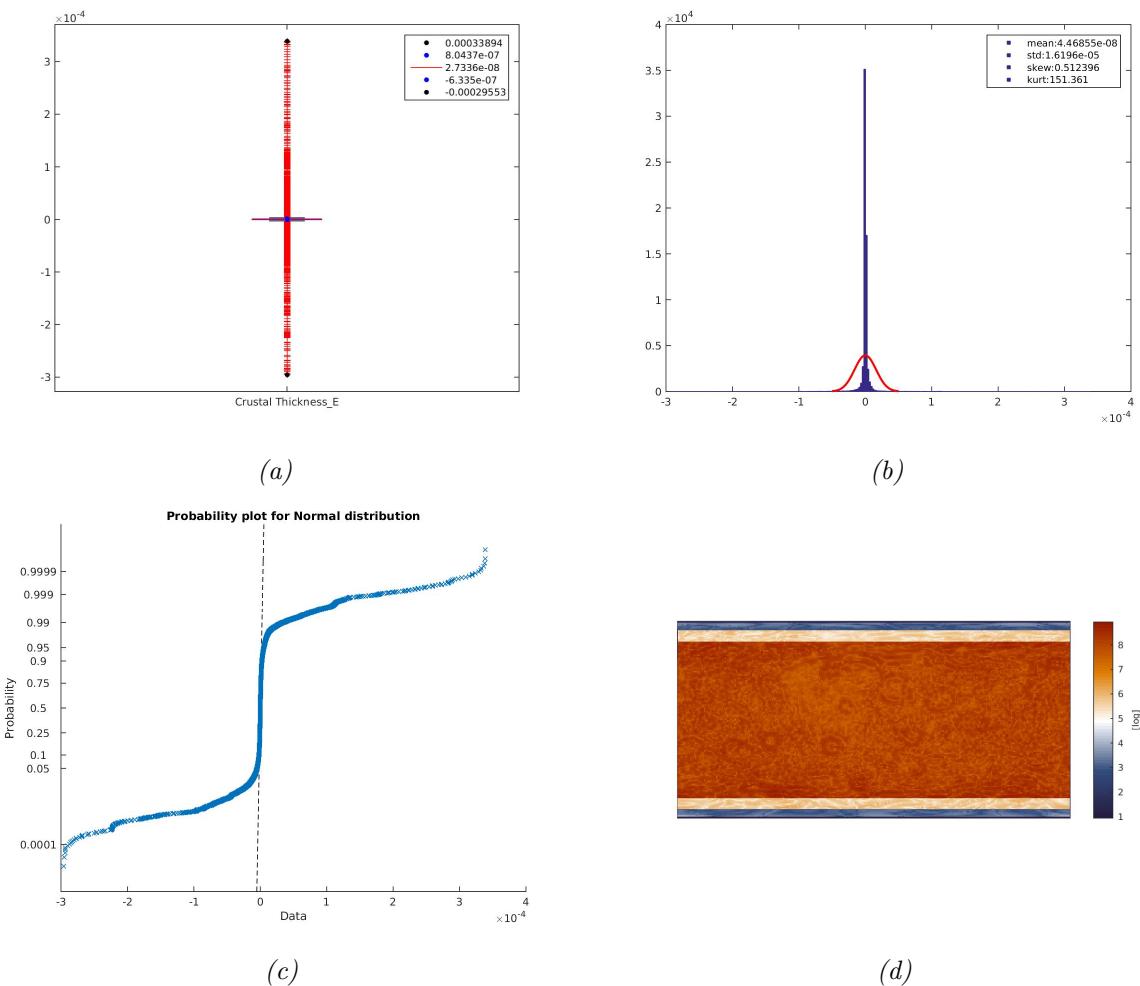


Figure A.13: Summary for `CrustalThickness_e`: (a) boxplot; (b) histogram; (c) probability; (d) leverages.

## Crustal Thickness: Gauss Curvature

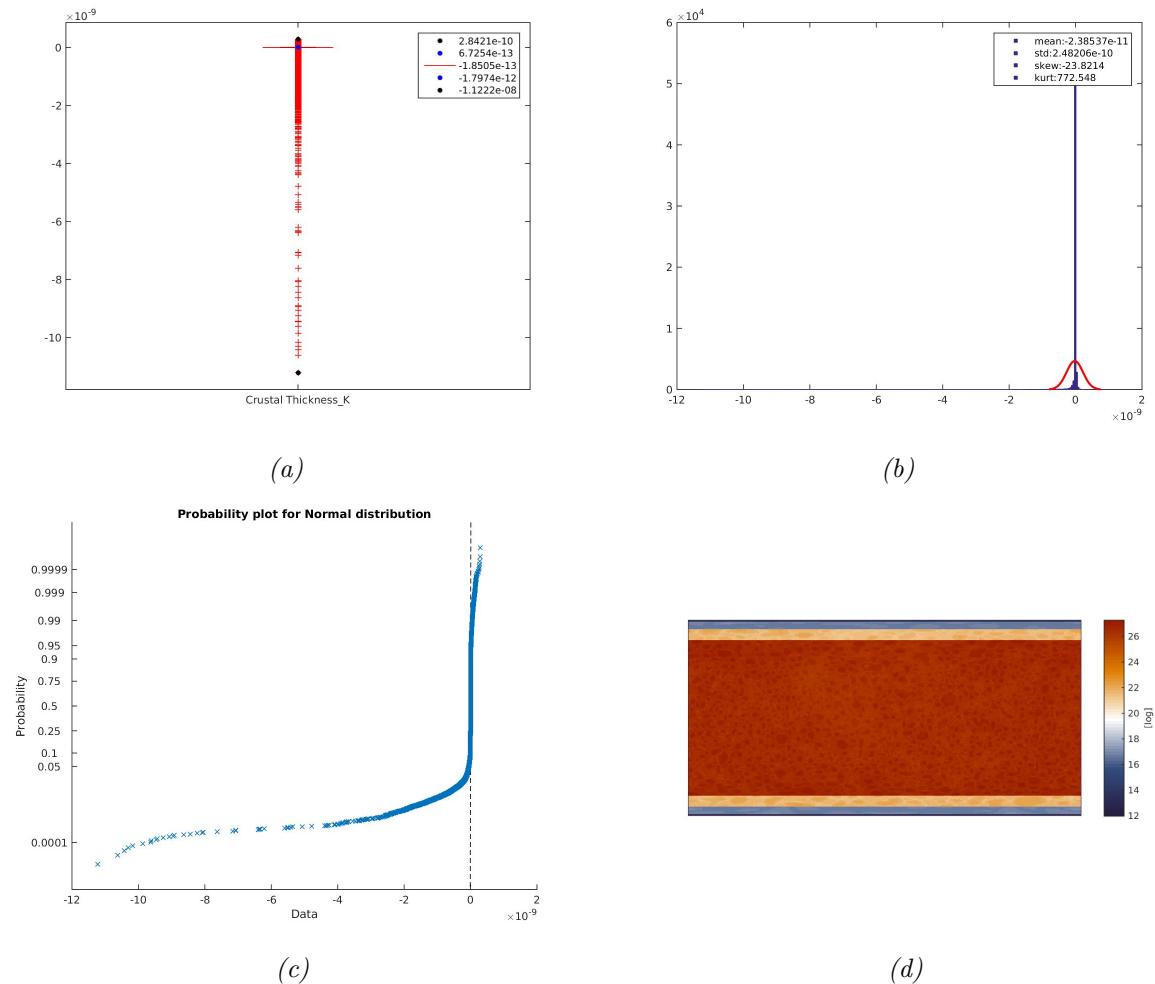


Figure A.14: Summary for CrustalThickness\_k: (a) boxplot; (b) histogram; (c) probability; (d) leverages.

## Crustal Thickness: Mean Curvature

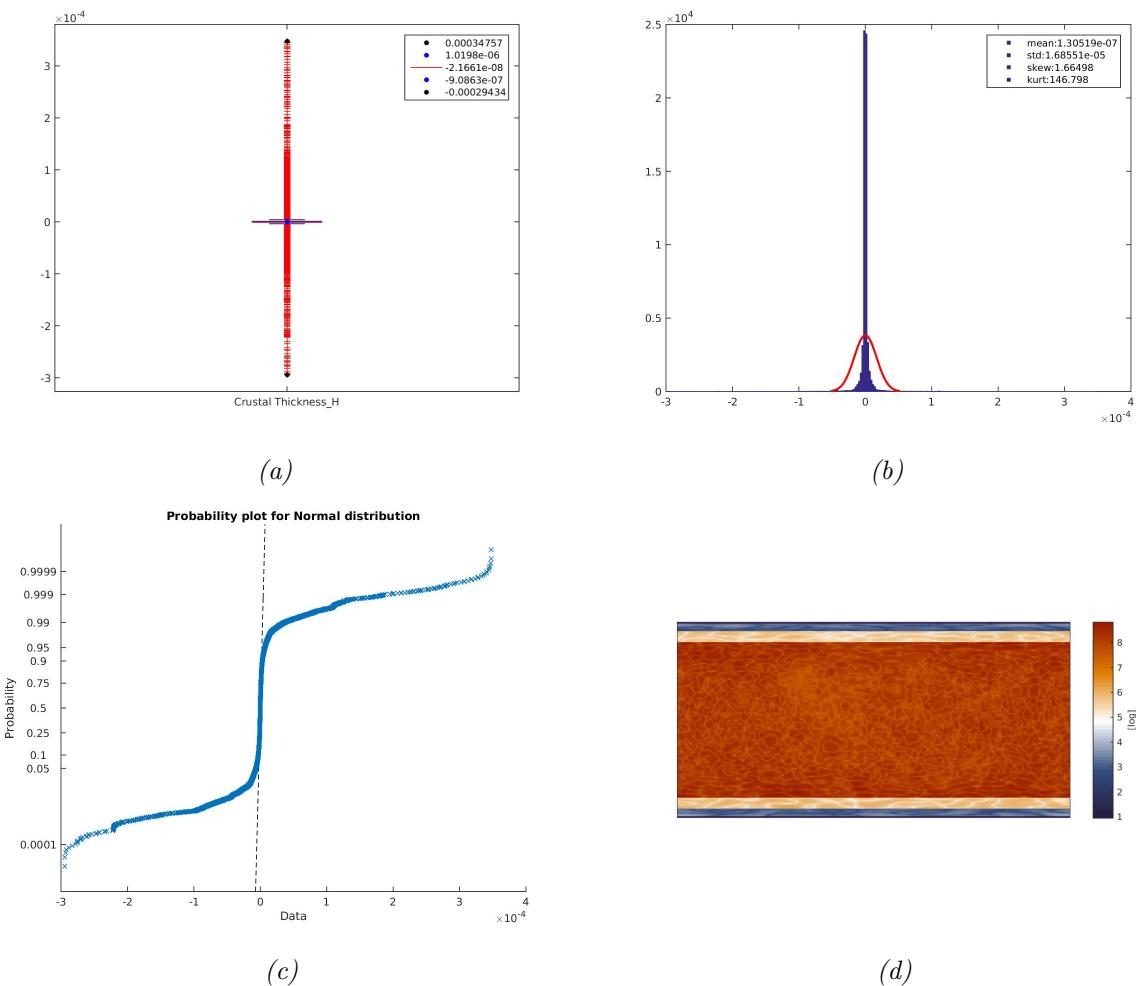


Figure A.15: Summary for `CrustalThickness_h`: (a) boxplot; (b) histogram; (c) probability; (d) leverages.