



TU WIEN  
DEPARTMENT OF GEODESY  
AND GEOINFORMATION  
RESEARCH UNIT  
ENGINEERING GEODESY

## *Diploma thesis*

# Clustering methods in the context of deformation analysis based on laser scanning point clouds

to obtain the academic degree

**Diplom-Ingenieur/in**

submitted at

Department of Geodesy and Geoinformation  
Research Unit Engineering Geodesy  
Wiedner Hauptstraße 8-10 / E120.5  
A-1040 Vienna, Austria

under the supervision of

Univ.Prof. Dr.-Ing. Hans-Berndt Neuner

and

Prof. Dr. Corinna Harmening

by

**Victoria Kostjak BSc**

Matriculation number: 11771176

Vienna, August 2, 2022

---

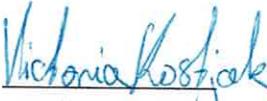
Hans-Berndt Neuner

*Affidavit*

I declare in lieu of oath, that I wrote this thesis and performed the associated research myself, using only literature cited in this volume. If text passages from sources are used literally, they are marked as such.

I confirm that this work is original and has not been submitted elsewhere for any examination, nor is it currently under consideration for a thesis elsewhere.

Vienna, August 2, 2022

  
Signature

# Acknowledgement

First and foremost, I am extremely grateful to my supervisors, Univ.Prof. Dr.-Ing. Hans-Berndt Neuner and Prof. Dr. Corinna Harmening for their invaluable advice, continuous support, and patience during my study. Thank you both for always keeping the doors open and the discussions valuable. I have learned a lot from you for this work and for the future.

I would also like to acknowledge my boyfriend as the second reader of this thesis and his unfailing support and encouragement. This accomplishment would not have been possible without him. In addition I would also like to thank my work colleagues and student colleagues who have always supported me with advice and conversations. Above all, thank you Bea for your energetic support during my studies and beyond.

And finally, I would be remiss not to mention my parents, who made this path possible in the first place. Thank you.

# Abstract

The rapid development of laser scanners in recent years has opened a new field of research. In particular, point clouds generated by terrestrial laser scanners are used in engineering geodesy. Unlike previous point-wise methods, point clouds contain spatial information about the object. Consequently, the search for area-based analysis methods to describe the scanned objects is the new challenge.

In [\[Harmening and Neuner, 2020\]](#) a deformation analysis approach is introduced. The approach bases on a least square collocation. The main objective is to model the signal and noise in the point cloud. The information about the deformation holds the signal. This thesis deals with a part of the modelling process of the signal. This consists of finding locally homogeneous areas and representing them by a suitable standard deviation.

The main task is to define a clustering process, which reliably provides the same results. The clustering method K-means used in the initial approach does not have this property. An additional task is the improvement of the calculation routine of the representative standard deviations. They are used in the stochastic process as a slowly varying factor.

The new clustering process, introduced in this work, is a combination of two clustering methods. The first one, clustering only the deformed part of the point cloud, is a spectral clustering algorithm. The clusters are constructed in relation to the eigenvectors of the Laplacian matrix. The matrix is derived from the similarity graph, which describes the connection between the data points. An additional clustering is performed using a Gaussian mixture model. This method leads to a soft clustering implementation, which means that data points can belong to multiple clusters. The posterior probability of a data point belonging to a cluster is determined considering Gaussian distributions. Moreover, soft clustering allows crossovers to be better described with the cluster result. In connection with the properties of soft clustering, the stable region was also included in the clustering.

As it turned out, finding a suitable clustering process was not the biggest challenge in this work. The feature vector for each clustering method, the choice of the number of clusters, the influence of the stable points on the clustering result and the calculation of the representative standard deviation for each point had influence on the final results. The latter, which is needed for the normalization of the residuals to the trend surface, had a greater impact on the final results than assumed.

In essence, however, an optimized clustering process has been developed that provides stable final results. Influences from other parts of the approach have not yet been considered with respect to the new clustering process.

# Kurzfassung

Die rasche Entwicklung des Laserscanners in den vergangenen Jahren eröffnete einen neuen Forschungsbereich. Vor allem werden Punktwolken, welche von terrestrischen Laserscannern erzeugt werden, in der Ingenieurgeodäsie verwendet. Im Gegensatz zu den punktweisen Methoden beinhalten Punktwolken flächenhafte Objektinformationen. Die neue Herausforderung ist, dass flächenhafte Analysemethoden zur Beschreibung des gescannten Objekts gefunden werden.

In [Harmening and Neuner, 2020](#) wird ein Ansatz zu einer Deformationsanalyse vorgestellt. Dieser basiert auf einer Kollokation nach kleinsten Quadraten. Das Hauptziel ist das Signal und Rauschen in der Punktwolke zu modellieren. Im Signal ist die Information zur Deformation enthalten. Diese Arbeit beschäftigt sich nur mit dem Modellierungsprozess des Signals. Dieser besteht aus der Findung lokal homogener Bereiche und deren Repräsentation durch eine passende Standardabweichung.

Die Hauptaufgabe ist die Definition eines Clusterablaufes, welcher wiederholt gleiche Ergebnisse liefert. Die verwendete Methode K-means im bisherigen Ansatz liefert jedoch bei einer Wiederholung der Clusterung unterschiedliche Ergebnisse. Eine zusätzliche Aufgabe ist die Verbesserung der Berechnungsroutine für die repräsentativen Standardabweichungen. Diese werden im stochastischen Modell als langsam veränderlichen Faktor eingesetzt.

Der neue Ablauf der Clusterung, welcher in dieser Arbeit vorgestellt wird, besteht aus einer Kombination von zwei Clustermethoden. Die Erste, welche nur den deformierten Teil der Punktwolke clustert, ist ein spektraler Clusteralgorithmus. Die Cluster entstehen im Zusammenhang mit den Eigenvektoren der Laplacematrix. Die Matrix wird abgeleitet vom Ähnlichkeitsgraphen, welcher die Verbindungen zwischen den Datenpunkten beschreibt. Zusätzlich wird mit einem Gaußschen Mischmodell geclustert. Diese Methode ermöglicht die Implementation einer weichen Clusterung, welche die Zuteilung von Datenpunkten zu mehreren Clustern ermöglicht. Die posteriore Wahrscheinlichkeiten der Zugehörigkeit eines Datenpunktes zu einem Cluster wird unter Berücksichtigung von gaußschen Verteilungen bestimmt. Außerdem wird eine bessere Beschreibung von Übergängen ermöglicht. Im Zusammenhang mit der weichen Clusterung, wurde auch der stabile Bereich in den Clusterablauf mit einbezogen.

Es stellte sich heraus, dass das Finden eines geeigneten Clusterablaufes nicht die größte Herausforderung in dieser Arbeit war. Der Merkmalsvektor für jede Clustermethode, die Wahl der Anzahl der Cluster, der Einfluss der stabilen Punkte auf das Clusterergebnis und die Berechnung der repräsentativen Standardabweichung für jeden Punkt hatten einen Einfluss auf das Endergebnis. Letztere, welche benötigt werden um die Abweichungen zur Trendfläche zu normieren, hatte einen größeren Einfluss auf das Endergebnis als zuvor angenommen.

Schlussendlich konnte jedoch ein optimierter Clusterablauf entwickelt werden, welcher stabile Endergebnisse liefert. Einflüsse aus anderen Teilen des Verfahrens wurden im Hinblick auf das neue Clusterverfahren noch nicht berücksichtigt.

# Contents

|   |           |
|---|-----------|
| <b>1. Introduction</b>  | <b>8</b>  |
| <b>2. Common clustering methods</b>   | <b>9</b>  |
| 2.1. Partitional methods  | 9         |
| 2.1.1. K-means  | 9         |
| 2.1.2. K-medoids  | 11        |
| 2.1.3. Fuzzy C-Means clustering   | 12        |
| 2.2. Spectral clustering  | 12        |
| 2.3. Probabilistic method - Gaussian mixture model                            | 14        |
| 2.4. Density-based method - DBSCAN  | 16        |
| 2.5. Hierarchical clustering  | 17        |
| 2.6. Clustering analysis in geodesy   | 18        |
| <b>3. Further development of the deformation analysis algorithm</b>           | <b>19</b> |
| 3.1. Overview of the deformation analysis algorithm                           | 19        |
| 3.2. Comparison of the clustering methods                                     | 20        |
| 3.2.1. Comparison based on the algorithm                                      | 21        |
| 3.2.2. Comparison based on the behaviour in the deformation analysis approach | 22        |
| 3.2.3. Discussion   | 25        |
| 3.3. Hard clustering implementation   | 25        |
| 3.4. Combination of clustering methods  | 28        |
| 3.5. Soft clustering implementation   | 30        |
| 3.6. Number of clusters   | 32        |
| 3.7. Feature vector adjustment  | 33        |
| 3.8. Edge region of the deformed area   | 35        |
| 3.9. Summary of the new approach  | 36        |
| <b>4. Data and results</b>  | <b>38</b> |
| 4.1. Data set D1  | 38        |
| 4.2. Comparison of the clustering result                                      | 39        |
| 4.3. Comparison of the calculated standard deviations                         | 40        |
| 4.4. Comparison of the analysis results                                       | 41        |
| 4.5. Further data sets  | 42        |
| <b>5. Conclusion</b>  | <b>46</b> |
| 5.1. Summary  | 46        |
| 5.2. Outlook  | 47        |
| <b>List of Figures</b>  | <b>I</b>  |

|   |              |
|---|--------------|
| <b>List of Tables</b>                   | <b>VI</b>    |
| <b>Bibliography</b>                     | <b>VII</b>   |
| <b>A. Results D1 (initial approach)</b> | <b>X</b>     |
| A.1. x-direction epoch 2 . . . . .      | X            |
| A.2. y-direction epoch 2 . . . . .      | XI           |
| A.3. z-direction epoch 2 . . . . .      | XII          |
| A.4. x-direction epoch 3 . . . . .      | XIII         |
| A.5. y-direction epoch 3 . . . . .      | XIV          |
| A.6. z-direction epoch 3 . . . . .      | XV           |
| A.7. final results . . . . .            | XVI          |
| <b>B. Results D1 (updated approach)</b> | <b>XVIII</b> |
| B.1. x-direction epoch 2 . . . . .      | XVIII        |
| B.2. y-direction epoch 2 . . . . .      | XIX          |
| B.3. z-direction epoch 2 . . . . .      | XX           |
| B.4. x-direction epoch 3 . . . . .      | XXI          |
| B.5. y-direction epoch 3 . . . . .      | XXII         |
| B.6. z-direction epoch 3 . . . . .      | XXIII        |
| B.7. final results . . . . .            | XXIV         |
| <b>C. Results D1 (new approach)</b>     | <b>XXVI</b>  |
| C.1. x-direction epoch 2 . . . . .      | XXVI         |
| C.2. y-direction epoch 2 . . . . .      | XXVIII       |
| C.3. z-direction epoch 2 . . . . .      | XXX          |
| C.4. x-direction epoch 3 . . . . .      | XXXII        |
| C.5. y-direction epoch 3 . . . . .      | XXXIV        |
| C.6. z-direction epoch 3 . . . . .      | XXXVI        |
| C.7. final results . . . . .            | XXXVIII      |

# 1. Introduction

In recent years the development of the terrestrial laser scanner opened new possibilities to describe the form of an object. Due to the areal measuring method, new demands were made for the evaluation methods of deformation analysis. The previously used point-wise evaluation methods are no longer sufficient for areal phenomena. But the implementation of areal analysis methods bring new challenges especially for deformation analysis [Wunderlich et al., 2016](#).

In [Harmening and Neuner, 2020](#) a deformation analysis approach is introduced. The basis is a least square collocation. The functional model consists of a summation of a trend, the signal and noise. The signal holds the deformation and is related to the trend, realized by the initial measurement epoch without deformation. It is assumed that the deformation can be described by a locally stationary stochastic process. This is composed of the product of a stationary correlation function and a slowly varying factor.

This thesis deals with the part of the deformation analysis approach that calculates this factor. The aim is to define homogeneous areas in the deformed region using a suitable clustering method and to calculate the representative standard deviation of the areas there. In the initial approach introduced in [Harmening and Neuner, 2020](#), a K-means clustering is performed. However, the algorithm does not allow repeatable results. Finding a suitable clustering method that reliably provides the same results is the main task of this work. The additional consideration of the calculation of the representative standard deviation was done to evaluate the new clustering process in the best possible way.

In section 2, an overview of common clustering methods and their application in geodesy is given. In section 3, the developed approach and its aspects are presented. A comparison of the results with respect to different datasets is given in section 4. Section 5 summarizes the results of this work and gives an outlook on future work.

## 2. Common clustering methods

The following chapter gives an overview of the common clustering methods and their application in geodesy. Generally speaking, each set of data has its own structure. The main task of cluster analysis is to break these down in order to be able to draw conclusions about new findings [Nagy, 1968]. However, this can be achieved through different approaches. The grouping of the methods is based on the classifications in [Aggarwal and Reddy, 2018] and [Han, 2011].

### 2.1. Partitional methods

Partitional methods are single-level approaches, which means that there is only one partition of the data set in the given number of clusters [Jain et al., 1999] [Nagy, 1968].

#### 2.1.1. K-means

The K-means algorithm is the best known partitional clustering algorithm because of the simplicity in its implication [Asgharbeygi and Maleki, 2008] [Jain et al., 1999].

The algorithm starts with selecting  $k$  initial centroids for the cluster [Aggarwal and Reddy, 2018] [The MathWorks, Inc., 2022g]. After that the assignment of each point to a cluster starts. The criteria for an assignment is a chosen distance metric between each point, for K-means Manhattan distance ( $L_1$ -Norm), Euclidean distance ( $L_2$ -Norm) and Cosine similarity are usually used. When all points are assigned to the nearest cluster centroid, the centroids will be recomputed. Then the assignment will be repeated until the number of iterations is reached. Another termination criterion that can be set is the percentage of points that change the cluster. This threshold describes the maximum number of points that are allowed to change the cluster during an iteration step. Mathematically, K-means is an optimization problem. The underlying minimization function is called the **Sum of Squared Errors** (SSE) (equation (2.1) [Aggarwal and Reddy, 2018] (sec. 4.2.1)).

$$SSE(C) = \sum_{k=1}^K \sum_{x_i \in C_k} (c_k - x_i)^2 \quad (2.1)$$

with

|       |     |   |
|-------|-----|---|
| $K$   | ... | number of data points                   |
| $C_k$ | ... | $k$ th cluster                          |
| $x_i$ | ... | point in $C_k$                          |
| $c_k$ | ... | mean of the $k$ th cluster (= centroid) |

Regarding the initial centroids one improvement is the  $k_{++}$ -algorithm at which the first centroid is picked randomly and the others are selected by a weighted probability score

[The MathWorks, Inc., 2022g]. Consequently, the next selected centroid is farthest from the current selected [Aggarwal and Reddy, 2018].

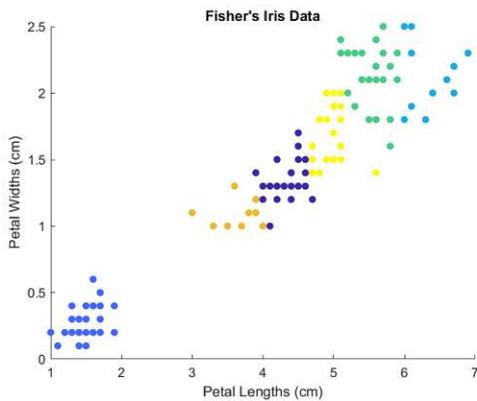


Figure 2.1.: K-means (number of clusters: 6)

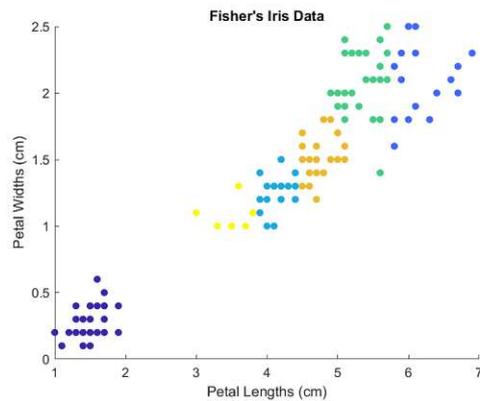


Figure 2.2.: K-means (number of clusters: 6) retry

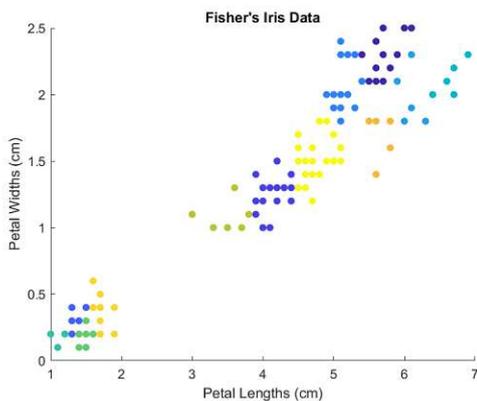


Figure 2.3.: K-means (number of clusters: 12)

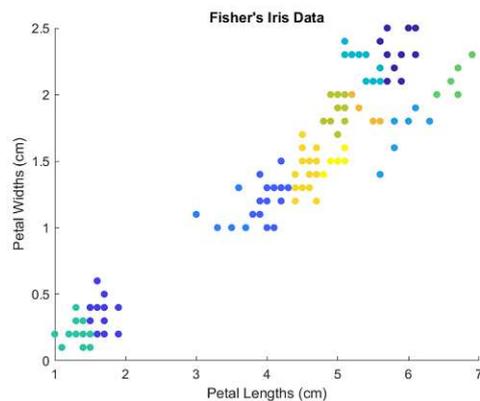


Figure 2.4.: K-means (number of clusters: 12) retry

Obviously, the result of clustering is still influenced by the centroids chosen at the beginning, as the figures 2.1 - 2.4 show. The used data is the Fisher's Iris data set [FISHER, 1936]. The strong variation of the clusters is the reason why the K-means clustering method should be replaced in the deformation analysis approach. In addition, the algorithm is strongly influenced by outliers [Han, 2011]. However, it is included in the tests for comparison with the other clustering methods.

Furthermore, finding the appropriate number of clusters for the dataset is a challenge for all clustering methods. Several researchers are confronted with this problem, therefore many approaches exist [Xu and WunschII, 2005] [Aggarwal and Reddy, 2018] [Han, 2011]. The tested methods are described in section 3.6.

### 2.1.2. K-medoids

K-medoids is basically a variation of the K-means clustering method [Aggarwal and Reddy, 2018] [Han, 2011]. In contrast selecting the centroid of a cluster by the mean value of the objects in one cluster as in K-means, an actual data point is representing the cluster center. This is where the name medoids came from. The K-medoids algorithm is less sensitive to outliers since an absolute-error criterion is used as minimization function. Equation (2.2) shows the criterion [Han, 2011] (p. 455).

$$E = \sum_{i=1}^k \sum_{p \in C_i} \text{dist}(p, o_i) \quad (2.2)$$

with

|               |     |   |
|---------------|-----|---|
| $k$           | ... | number of clusters                        |
| $C_i$         | ... | $i$ th cluster                            |
| $p$           | ... | points of the data set                    |
| $o_i$         | ... | representative object of $C_i$ (= medoid) |
| $\text{dist}$ | ... | distance function                         |

In this work the **P**artitioning **A**round **M**edoids (PAM) algorithm is used for iteratively minimizing the absolute-error function [The MathWorks, Inc., 2022h]. The figures 2.5 and 2.6 show exemplary clustering results with the K-medoids clustering method. It is applied to the Fisher's Iris data set [FISHER, 1936].

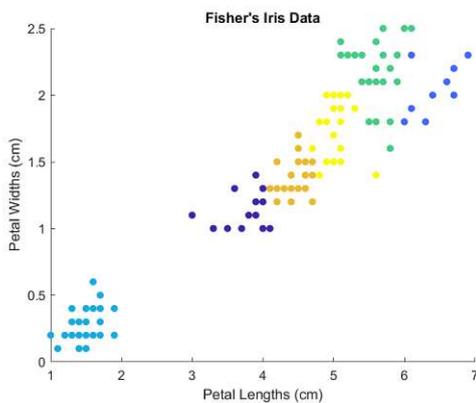


Figure 2.5.: K-medoids (number of clusters: 6)

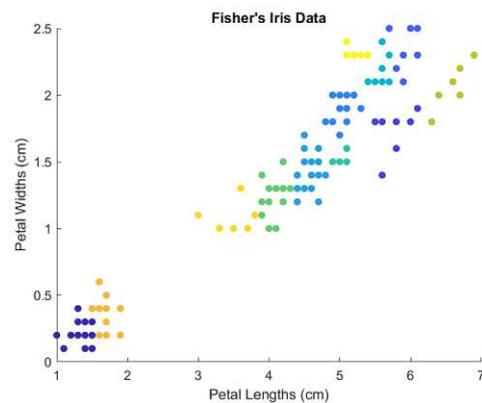


Figure 2.6.: K-medoids (number of clusters: 12)

### 2.1.3. Fuzzy C-Means clustering

Fuzzy clustering also known as soft clustering is another variation of the K-means clustering method [Aggarwal and Reddy, 2018] [The MathWorks, Inc., 2022e]. However, in this method points have the possibility to be a member of more than one cluster. It follows the possibility for clusters to have overlapping areas. Therefore, the membership is indicated with additional values between 0 and 1. Equation (2.3) shows the SSE function [Aggarwal and Reddy, 2018] (sec. 4.2.4.4). The **Fuzzy C-Means** clustering algorithm (FCM) has the same steps as K-means. In addition to the cluster centroid, the membership weight in the minimization of the SSE is also updated. Because of the similarity to K-means, FCM is also sensitive to outliers.

$$SSE(C) = \sum_{k=1}^K \sum_{x_i \in C_k} w_{xik}^\beta \|x_i - c_k\|^2 \quad (2.3)$$

with

|             |     |   |
|-------------|-----|---|
| $K$         | ... | number of clusters                                  |
| $C_k$       | ... | $k$ th cluster                                      |
| $x_i$       | ... | point in $C_k$                                      |
| $c_k$       | ... | mean of the $k$ th cluster (= centroid)             |
| $w_{xik}$   | ... | membership weight of point $x_i$ belonging to $C_k$ |
| $\beta > 1$ | ... | degree of fuzzy overlap                             |

## 2.2. Spectral clustering

In contrast to section 2.1, spectral clustering uses the given data set to derive a similarity graph [von Luxburg, 2007] [Azran and Ghahramani,] [The MathWorks, Inc., 2022j]. There are three types of them:  $\epsilon$ -neighbourhood graph,  $k$ -nearest neighbour graph and the fully connected graph. In general, the similarity graph contains the pairwise distances to the chosen neighbour points. The appropriate calculation of the distances should also be considered. After setting up the graph the distances are transformed into similarity measures with weights between the data points. Usually, the Gaussian similarity function is used.

Now two matrices can be derived from the graph: the similarity matrix  $S$  and the degree matrix  $D$ . The similarity matrix  $S$  represents the connected nodes of the similarity graph. The degree matrix  $D$  is a diagonal matrix. The entries are the sum of each row of the similarity matrix  $S$ . The degree matrix  $D$  represents the degree of a node, which is the number of the connected edges to it.

The matrix connecting the two matrices  $D$  and  $S$  is called Laplacian matrix  $L$ . It is the basis for clustering, since its eigenvalues tell a lot about the structure of the similarity graph [von Luxburg, 2007](#). The eigenvalues of a matrix is also called spectrum. This is were the name spectral clustering comes from. Usually, three Laplacian matrix types are listed in the literature: the unnormalized Laplacian matrix  $L_u$ , the normalized random-walk Laplacian matrix  $L_n$  introduced in [Shi and Malik, 2000](#) and the normalized symmetric Laplacian matrix  $L_{ns}$  introduced in [Ng et al., 2001](#). The calculations of the different matrices are given in the equations [\(2.4\)](#) - [\(2.6\)](#).

$$L_u = D - S \tag{2.4}$$

$$L_n = D^{-1}L_u \tag{2.5}$$

$$L_{ns} = D^{-\frac{1}{2}}L_uD^{-\frac{1}{2}} \tag{2.6}$$

For the clustering the smallest eigenvalue of the Laplacian matrix  $L$  is found [The MathWorks, Inc., 2022](#). Then the matrix consisting of the eigenvectors according to the chosen eigenvalue is calculated. Finally, the matrix, where the rows are treated as points, is clustered by a suitable clustering method. Usually, K-means or K-medoids is used. In figure [2.7](#) and [2.8](#) two examples of an spectral clustering with the Fisher's Iris data set [FISHER, 1936](#) are shown.

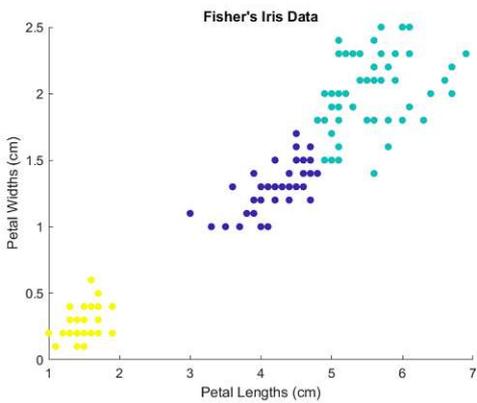


Figure 2.7.: Spectral clustering  
 number of clusters: 3  
 distance metric: Euclidean  
 similarity graph: k-nearest neighbour  
 Laplacian matrix: normalized  
 random-walk

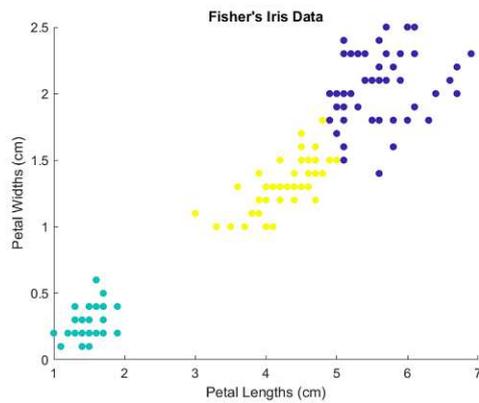


Figure 2.8.: Spectral clustering  
 number of clusters: 3  
 distance metric: Euclidean  
 similarity graph:  $\epsilon$ -neighbourhood  
 (radius = 1)  
 Laplacian matrix: normalized  
 symmetric

### 2.3. Probabilistic method - Gaussian mixture model

The **Gaussian mixture model** (GMM) is one representative clustering method in the group of probabilistic methods [Aggarwal and Reddy, 2018] [Xu and WunschII, 2005]. In general, it is assumed that the data set can be represented by a mix of Gaussian distributions. The amount of these is represented by the variable  $k$  (see equation (2.7) [Aggarwal and Reddy, 2018] (sec. 3.2.1)).

$$k \in \{1, \dots, K\} \quad (2.7)$$

with

|     |     |                              |
|-----|-----|------------------------------|
| $k$ | ... | Gaussian distribution number |
| $K$ | ... | number of clusters           |

The main objective of this clustering method is to find the three describing parameters of the Gaussian distributions. These are the mean  $\mu$ , the covariance  $\Sigma$  and the mixing probability  $\pi$ . Based on these parameters, the data points can finally be assigned to clusters. The mean  $\mu$  is the value with the highest probability and fixes the center position of the distribution.  $\Sigma$  defines the width of the distribution and  $\pi$  the height. However, the mixing probability of each cluster are, as the name says, probabilities. Therefore, all mixing probabilities have to meet equation (2.8) [Aggarwal and Reddy, 2018] (sec. 3.2.1).

$$\sum_{k=1}^K \pi_k = 1 \quad (2.8)$$

with

|         |     |  |
|---------|-----|--|
| $\pi_k$ | ... | mixing probability of $k$ th Gaussian distribution |
|---------|-----|--|

To find the optimal parameters so that each Gaussian distribution per cluster fits the data points it contains, the maximum likelihood is sought. However, the assumption for describing the data points is that they can be described with a mixture of Gaussian distributions. Therefore, finding the maximum likelihood for the data points is challenging. Equation (2.9) and (2.10) show the mathematical description of a Gaussian mixture model [Aggarwal and Reddy, 2018] (sec. 3.2.2).

$$p(x_n|\pi, \underline{\mu}, \underline{\Sigma}) = \sum_{k=1}^K \pi_k \mathcal{N}(x_n|\underline{\mu}_k, \underline{\Sigma}_k) \quad (2.9)$$

$$\mathcal{N}(\underline{x}|\underline{\mu}, \underline{\Sigma}) = \frac{1}{(2\pi)^{\frac{D}{2}}} \frac{1}{|\underline{\Sigma}|^{\frac{1}{2}}} \exp\left\{-\frac{1}{2}(\underline{x} - \underline{\mu})^T \underline{\Sigma}^{-1}(\underline{x} - \underline{\mu})\right\} \quad (2.10)$$

with

|                      |     |   |
|----------------------|-----|---|
| $\underline{x}_n$    | ... | vector of data points from $k$ th Gaussian distribution |
| $\mathcal{N}$        | ... | multivariate Gaussian distribution                      |
| $\underline{\mu}$    | ... | D-dimensional mean vector                               |
| $\underline{\Sigma}$ | ... | DxD covariance matrix                                   |
| $\underline{x}$      | ... | D-dimensional data point vector                         |
| $D$                  | ... | dimension   |
| $ \dots $            | ... | determinant   |

Due to the complexity of equation (2.9), an iterative method is used to determine the optimal parameters of each Gaussian distribution. The name of this method is **Expectation - Maximation (EM) algorithm** [The MathWorks, Inc., 2022f] [Aggarwal and Reddy, 2018]. As the name suggests, it consists of two main steps. These are executed until convergence. First, initial values for the parameters ( $\mu$ ,  $\Sigma$  and  $\pi$ ) are needed. These could be found by pre-clustering the data set with another method. Then, posterior probabilities for each point are computed in respect to the membership of a cluster. This is the so called E-step. Afterwards, during the M-step, the parameters are updated and the log-likelihood is computed. Finally, the data points can be assigned to the cluster related to the posterior probabilities [The MathWorks, Inc., 2022a].

To give an example, the Fisher's Iris data set [FISHER, 1936] is clustered with a Gaussian mixture model with 3 clusters. The result of clustering is shown in figure 2.9. Figure 2.10 shows the underlying structure of the multivariate Gaussian distribution.

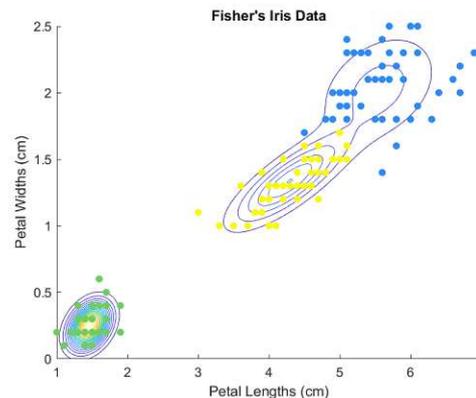
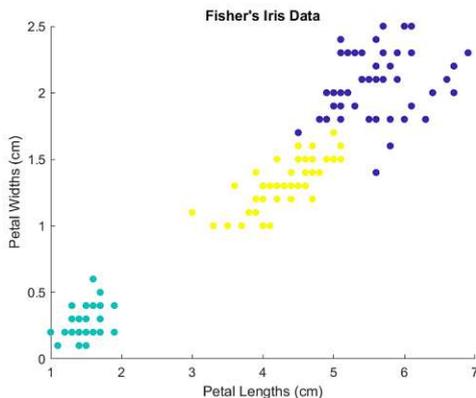


Figure 2.9.: Gaussian mixture model (number of clusters: 3)

Figure 2.10.: Gaussian mixture model with contours (number of clusters: 3)

## 2.4. Density-based method - DBSCAN

DBSCAN (Density-Based Spatial Clustering of Applications with Noise) is one representative method of density-based clustering [Han, 2011] [Schubert et al., 2017]. As input parameters the  $\epsilon$ -neighbourhood and the *MinPts* (Minimum number of neighbours) are needed [The MathWorks, Inc., 2022c]. When clustering a data set with DBSCAN the algorithm starts to sign each point as "unvisited". Then a random point is selected and marked as "visited". With the  $\epsilon$ -neighbourhood, which defines the radius for the neighbourhood search, the neighbours of the current point are found. If the number of them is equal or greater than the number of *MinPts*, the current point is assigned as a new cluster. Next, the algorithm goes through the neighbours to mark them as "visited", if they are not yet marked. In addition, the  $\epsilon$ -neighbourhood of these points are computed. The new found points are also added to the amount of neighbours. After all neighbouring points have been marked as "visited", the cluster is output. The process is repeated until no more points are marked as "unvisited". If there are not enough neighbours in an area, the point is marked as "noise". Furthermore, if a point has already been assigned to a cluster, it cannot be assigned to a cluster again. It follows that the clusters depend on how the "unvisited" points are selected one after the other. The two parameters  $\epsilon$ -neighbourhood and *MinPts* also have a large influence. However, the choice of suitable values is not trivial.

The figures 2.11 and 2.12 show the DBSCAN clustering method. The data set is the Fisher's Iris data set [FISHER, 1936]. The  $\epsilon$ -neighbourhood has the value 1.2. *MinPts* were assumed to be 50.

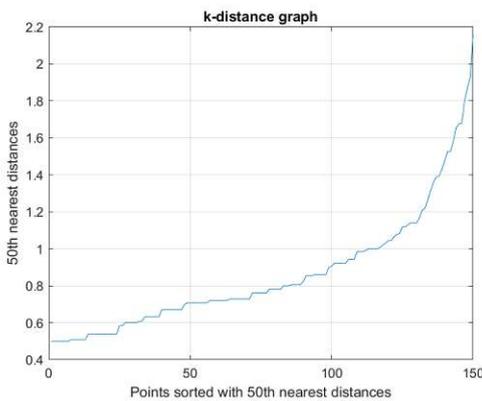


Figure 2.11.: k-distance graph (*MinPts* = 50)

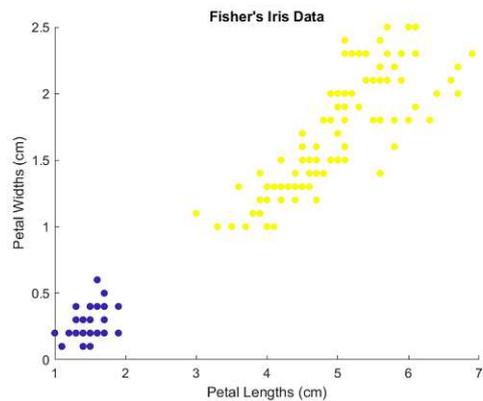


Figure 2.12.: DBSCAN clustering ( $\epsilon = 1.2$ )

## 2.5. Hierarchical clustering

In general, there are two possibilities to do a hierarchical clustering [Aggarwal and Reddy, 2018]. If all points are considered as one cluster at the beginning, which is divided afterwards, it is the so-called divisive clustering method. Consequently, a top-down hierarchy of clusters is generated. In contrast, in the agglomerative method, each point is a separate cluster which are subsequently merged. This generates a bottom-up hierarchy. Hence, hierarchical methods perform nested series of partitions, which is in contrast to paritional methods [Jain et al., 1999] [Nagy, 1968].

In this work the agglomerative clustering method is used in the context of the deformation analysis method. As mentioned in the beginning each point is a separate cluster [Jain et al., 1999] [The MathWorks, Inc., 2022b]. Due to that the first step in the algorithm is to compute pairwise distances between all data points. An agglomerative hierarchical cluster tree (dendrogram, figure 2.13) is then derived from the calculated distances. It shows the individual links between the points which leads to the similarity level. In this work the single link approach, which is similar to the nearest neighbour search, is used [Aggarwal and Reddy, 2018] [The MathWorks, Inc., 2022b]. Hence, a link between two points shows the nearest neighbour. The same applies to groups of data points. Finally, the dendrogram is cut at a certain level, which leads to the generation of a cluster. However, different levels of cutting produce different clustering results. Furthermore, there are various values for defining the cutting level. In this work the level is defined by setting the maximum number of clusters.

The figures below (2.13, 2.14) shows the dendrogramm and the hierarchical clustering with 6 cluster of the Fisher's Iris data set [FISHER, 1936].

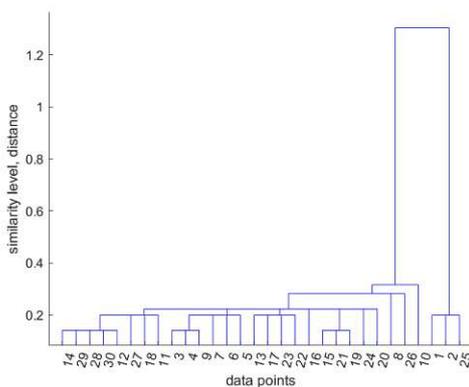


Figure 2.13.: Dendrogram with Euclidean distance

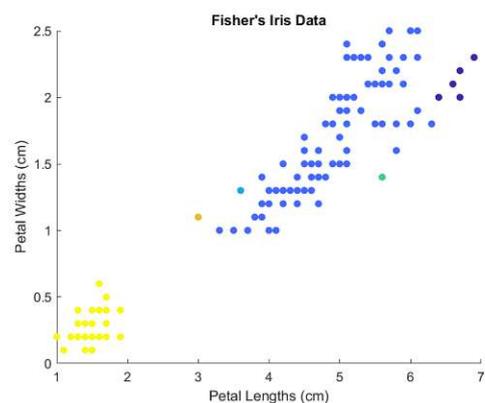


Figure 2.14.: Hierarchical clustering (number of clusters: 6)

## 2.6. Clustering analysis in geodesy

Historically, the first clustering methods were used in biology [Everitt, 2011]. Nowadays, cluster analysis is used in many research areas, including geodesy. There are many applications where cluster analysis is needed. The methods used in the course of cluster analyses are often those that were also compared in this work.

As expected, K-means is the most commonly used method when performing a cluster analysis. For instance, it is used to cluster GNSS data [Higginson et al., 2015], airborne laser scanning raw data [Morsdorf et al., 2003], satellite imagery [Lv et al., 2010] and in the study on the distribution of geodetic control points [Pokonieczny et al., 2017].

But also the other methods presented in section 2 can be found in publications. For example, the hierarchical method is also used for satellite imagery [Häme et al., 2020]. A density-based clustering method is used in [Amjad and Shah, 2020] for finding sites for solar farms. An example of spectral clustering is given in [Xu et al., 2016], where it is used to cluster point clouds derived from pictures. [Li et al., 2022] uses a fuzzy clustering method to analyse the signal for GNSS measurements. The publication [Singh et al., 2021] uses the K-medoids method in connection with a terrestrial laser scanning point cloud.

A publication, where a Gaussian mixture model is applied, is [Granat et al., 2021]. It is used for the clustering of GNSS data to identify California's Major Faults. It should also be said that several cluster methods are also used in approaches. For example, in the study of [Praene et al., 2019] about climate zone mapping a hierarchical method and K-means is used.

This chapter is intended to give a small insight into the different areas of application of a cluster analysis in geodesy. It is by no means a complete description, but the result of a literature research in the course of this work. In summary, it can additionally be said that in the research directions of geodesy, cluster analysis is a frequently used tool.

## 3. Further development of the deformation analysis algorithm

In the following section an overview of the deformation analysis approach is given. Afterwards, the changes are described. Each aspect, which has significant relevance to the final result, is discussed in detail.

### 3.1. Overview of the deformation analysis algorithm

In the publication "A spatio-temporal deformation model for laser scanning point clouds" by C. Harmening and H. Neuner an approach based on a least square collocation is described [Harmening and Neuner, 2020](#). This subsection should give a short introduction to the method. Formula (3.1) shows the functional model of the deformation analysis approach.

$$\underline{l} = \underline{A}\vartheta + \underline{s} + \underline{\epsilon} \quad (3.1)$$

The first term  $\underline{A}\vartheta$  describes the deterministic trend where  $\vartheta$  is the vector of unknowns. The trend is subtracted from the vector of observations  $\underline{l}$ . Consequently the phenomena is described by the summation of the signal and noise (equation (3.2)).

$$\underline{l} - \underline{A}\vartheta = \underline{s} + \underline{\epsilon} \quad (3.2)$$

The deformation is contained in the stochastic signal  $\underline{s}$ . The signal is modelled also similar to a least squares collocation. The last term represents the stochastic measurement noise  $\underline{\epsilon}$ . The signal and the noise are both assumed as normally distributed shown in equation (3.3) and (3.4). The expected value is 0.  $\underline{\Sigma}$  is the covariance matrix. Moreover correlations between the two covariance matrices are excluded (equation (3.5)).

$$\underline{s} \sim \mathcal{N}(0, \underline{\Sigma}_{ss}) \quad \text{with} \quad \underline{\Sigma}_{ss} = \sigma_0^2 \cdot \underline{Q}_{ss} \quad (3.3)$$

$$\underline{\epsilon} \sim \mathcal{N}(0, \underline{\Sigma}_{\epsilon\epsilon}) \quad \text{with} \quad \underline{\Sigma}_{\epsilon\epsilon} = \sigma_0^2 \cdot \underline{Q}_{\epsilon\epsilon} \quad (3.4)$$

$$\underline{\Sigma}_{s\epsilon} = \underline{\Sigma}_{\epsilon s} = 0 \quad (3.5)$$

The trend  $\underline{A}$  is realised by a B-Spline surface, which describes the initial geometry of the object. It is related to the first measurement epoch. Since the deformed area is linked with the signal, a distinction between deformed and stable regions is required. The distinction is realised by a threshold. The criterion for distorted points is related to the standard deviation of the noise in

the first measuring epoch. It is given in equation (3.6). It states that the absolute value of the residual to the trend surface  $r_x$  must be greater than 1.5 times of the standard deviation of the noise in the first measuring epoch  $\sigma_x^1$ .

$$|r_x| > 1.5 \sigma_x^1 \quad (3.6)$$

with

|              |     |  |
|--------------|-----|--|
| $r_x$        | ... | pointwise residual to trend surface                          |
| $\sigma_x^1$ | ... | standard deviation of the noise in the first measuring epoch |
| $x$          | ... | coordinate (x,y,z)   |
| $ \cdot $    | ... | absolute value   |

The estimation of the signal is done in a multi-step way. In short, the localisation of locally homogeneous areas is significant for modelling the corresponding standard deviation. For this, a K-means clustering method is used. The corresponding standard deviations are the basis for describing the deformation. Thereafter, the estimation of empirical autocorrelations and cross-correlations can be done. The last steps are modelling the noise, filtering the results and doing a prediction between the measured points.

Furthermore, it should be mentioned that the entire processes in the deformation analysis algorithm are carried out separately in terms of coordinates. In addition, cross-coordinate and cross-epoch processes are considered.

The used data set in this approach is simulated. It consist of three epochs. The epoch 1 is describing the trend and has no deformed areas. In the two others one control point of the B-Spline surface of the trend is shifted in the z-direction. The details of the dataset is described below in section 4.

Finally, this work deals with only one part of the approach, the search for homogeneous areas and the calculation of the representative standard deviations to normalize the residuals to the trend surface. The aim is to close the gap between the computed residuals to the trend and the normalized ones. The following sections discuss each aspect in detail.

### 3.2. Comparison of the clustering methods

In the above described approach the detection of locally homogeneous areas is realised by a K-means clustering. As mentioned in section 2.1.1 the K-means algorithm selects the cluster centroids randomly. Consequently, the clusters are different for each run, which partially leads to different final results. First of all, finding an alternative to the K-means algorithm is discussed.

### 3.2.1. Comparison based on the algorithm

As a first step the single clustering methods will be discussed in more detail. The clustering methods introduced in section 2 have different criteria to select the points for clusters. In addition, the aim is to have the same output after clustering the same data.

Compared to the described methods, the DBSCAN method needs an additional threshold to compute the clusters ( $\epsilon$  neighbourhood, section 2.4). The threshold is dependent on the distance between the points and the number of points in the neighbourhood. This is a handicap for an automated algorithm, because the value of the  $\epsilon$  neighbourhood has to be specified by hand.

As the Fuzzy C-Means clustering algorithm (FCM) is just a soft clustering variant of K-means and the Gaussian mixture model algorithm chooses the initial parameters with the  $k_{++}$ -algorithm, it is not surprising that both clustering methods have different results according to the composition of the clusters after each run of the deformation analysis algorithm. Therefore, both methods are not suitable for the aim of this work. Nevertheless, it should be mentioned that only these two methods support the implementation of soft clustering. According to that, combinations will be tried to achieve soft clustering implementation in the deformation analysis approach (see section 3.4).

Although the K-medoids method is also a variant of K-means, it could lead to more similar cluster results. Since the medoids are actual data points and the minimization process is performed with an absolute-error criterion. The cluster assignment is less influenced by outliers in the data points. However, the randomly chosen initial medoids lead to significant variations in the clusters.

In contrast, the spectral clustering and the hierarchical clustering partition the data set according to a computed data graph. Thus, using distances between the data points should always lead to the same clustering result. Even though the influence of the K-means algorithm in the spectral clustering is not assessable, the method will still be considered.

As a result the two clustering methods: spectral clustering and hierarchical clustering look promising in terms of the computational algorithm. Table 3.1 gives an overview of the first exclusion. Finally, the need to specify the number of clusters has been deliberately omitted. This aspect is described in more detail in section 3.6.

Table 3.1.: Overview of the first exclusion

| method       | usability     |
|--------------|---------------|
| K-means      | $\mathcal{X}$ |
| FCM          | $\mathcal{X}$ |
| GMM          | $\mathcal{X}$ |
| DBSCAN       | $\mathcal{X}$ |
| K-medoids    | $\mathcal{X}$ |
| spectral     | ✓             |
| hierarchical | ✓             |

### 3.2.2. Comparison based on the behaviour in the deformation analysis approach

After a look at the algorithm of the individual clustering methods, they were tested in the deformation analysis approach. Although a first exclusion was made in section 3.2.1 all clustering methods are considered. The reason was to see the analysis results in comparison to the use of the K-means method in the clustering process. This also avoids exclusion of methods that are based only on theoretical assessment. It also allows the potential of each clustering method to be evaluated in terms of final results.

The following tables 3.2 - 3.4 show the results of the tests. The methods were tested with the default parameters of the MATLAB<sup>®</sup> functions [The MathWorks, Inc., 2022i]. The analysis consisted of the variation of the number of clusters and the weight in the used feature vector. Equation (3.7) shows the feature vector with the changeable variables.

$$\underline{f}_{x,y,z}^e = \left[ \underline{r}_{x,y,z}^e \cdot a \quad \underline{x}^e \cdot b \quad \underline{y}^e \cdot b \quad \underline{z}^e \cdot b \right] \quad (3.7)$$

with

$\underline{f}_{x,y,z}$  ... feature vector in x,y,z-direction

$\underline{r}_{x,y,z}$  ... residuals to trend surface in x,y,z-direction

$\underline{x}, \underline{y}, \underline{z}$  ... vector of x,y,z-coordinates

$a$  ... weight for residuals

$b$  ... weight for coordinates

$e$  ... epoch

Table 3.2.: Overview of results part 1

| number<br>of<br>clusters | method                      |                       |                               |                       |
|--------------------------|-----------------------------|-----------------------|-------------------------------|-----------------------|
|                          | K-means [mm]<br>with weight |                       | K-medoids [mm]<br>with weight |                       |
|                          | $a = 1000$<br>$b = 0.01$    | $a = 1$<br>$b = 0.05$ | $a = 1000$<br>$b = 0.01$      | $a = 1$<br>$b = 0.05$ |
| 5                        | -5 - 5                      | -5 - 5                | <b>-4 - 2</b>                 | -5 - 5                |
| 6                        | -5 - 5                      | <b>-6 - 4</b>         | ---                           | -5 - 5                |
| 7                        | -5 - 5                      | -500 - 500            | ---                           | -10 - 10              |
| 8                        | ---                         | -40 - 20              | ---                           | -5 - 10               |
| 9                        | ---                         | -5 - 5                | ---                           | -20 - 80              |
| 10                       | -5 - 5                      | -20 - 20              | ---                           | -5 - 5                |
| 11                       | ---                         | -100 - 200            | ---                           | -8 - 4                |
| 12                       | ---                         | -20 - 20              | ---                           | -5 - 5                |

The shown results in millimetres are the range of the discrepancies to the nominal surface in the z-coordinate of epoch  $e = 3$ . This is the epoch with the biggest deformation in the z-direction. The best results are in bold letters. These are determined with the lowest range in millimetres and the lowest values in the two considered weights of the feature vector.

Table 3.3.: Overview of results part 2

| number<br>of<br>clusters | method                   |                       |                            |                       |
|--------------------------|--------------------------|-----------------------|----------------------------|-----------------------|
|                          | FCM [mm]<br>with weight  |                       | DBSCAN [mm]<br>with weight |                       |
|                          | $a = 1000$<br>$b = 0.01$ | $a = 1$<br>$b = 0.05$ | $a = 1000$<br>$b = 0.01$   | $a = 1$<br>$b = 0.05$ |
| 5                        | ---                      | ---                   | ---                        | ---                   |
| 6                        | ---                      | -5 - 5                | ---                        | ---                   |
| 7                        | ---                      | -5 - 5                | ---                        | ---                   |
| 8                        | ---                      | ---                   | ---                        | ---                   |
| 9                        | ---                      | ---                   | ---                        | ---                   |
| 10                       | ---                      | <b>-2 - 2</b>         | ---                        | ---                   |
| 11                       | ---                      | ---                   | ---                        | ---                   |
| 12                       | ---                      | -4 - 2                | ---                        | ---                   |

Table 3.4.: Overview of results part 3

| number<br>of<br>clusters | method                   |                       |                              |                       |                                  |                       |
|--------------------------|--------------------------|-----------------------|------------------------------|-----------------------|----------------------------------|-----------------------|
|                          | GMM [mm]<br>with weight  |                       | spectral [mm]<br>with weight |                       | hierarchical [mm]<br>with weight |                       |
|                          | $a = 1000$<br>$b = 0.01$ | $a = 1$<br>$b = 0.05$ | $a = 1000$<br>$b = 0.01$     | $a = 1$<br>$b = 0.05$ | $a = 1000$<br>$b = 0.01$         | $a = 1$<br>$b = 0.05$ |
| 5                        | -4 - 4                   | -4 - 2                | ---                          | -5 - 10               | <b>-2 - 4</b>                    | ---                   |
| 6                        | -5 - 5                   | <b>-2 - 2</b>         | ---                          | -20 - 40              | -5 - 5                           | ---                   |
| 7                        | -5 - 5                   | -5 - 5                | ---                          | ---                   | -10 - 10                         | -4 - 4                |
| 8                        | <b>-2 - 2</b>            | -5 - 5                | ---                          | -10 - 10              | -5 - 10                          | -5 - 5                |
| 9                        | -10 - 5                  | -4 - 4                | ---                          | -5 - 10               | -5 - 5                           | -5 - 5                |
| 10                       | <b>-2 - 2</b>            | -5 - 5                | ---                          | -6 - 4                | -4 - 4                           | -5 - 5                |
| 11                       | -4 - 2                   | <b>-2 - 2</b>         | ---                          | <b>-6 - 2</b>         | -100 - 50                        | -4 - 4                |
| 12                       | -2 - 2                   | <b>-2 - 2</b>         | ---                          | -8 - 4                | ---                              | -4 - 4                |

Obviously, the DBSCAN method was not able to pass the test. Although different parameters were tested, the deformation analysis approach was not able to compute final results. The problem was the approximation of autocorrelation and cross-correlation functions, which are linked to the clusters and their distribution on the object.

Figure 3.1 shows the clusters with DBSCAN with 10 clusters in z-direction of epoch  $e = 3$ . The distribution of the darkblue cluster is the reason for no final results. It is related to the assumption of the underlying process which describes the signal (see section 3.3). To fit this description, the clusters must have a compact form. This is obviously not the case here.

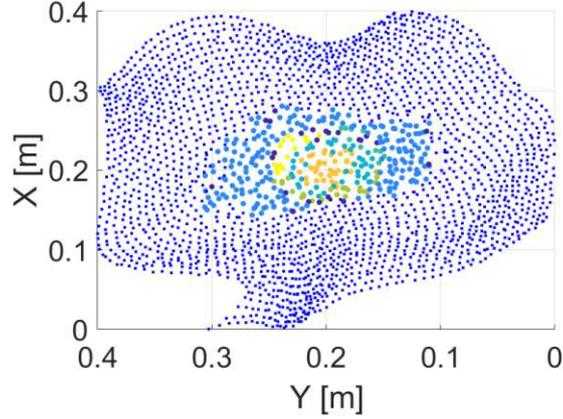


Figure 3.1.: Clusters DBSCAN (number of clusters = 10)

All other methods were able to produce analysis results. But the range of the remaining discrepancies is very different. As mentioned above also the potential of the clustering methods according to the analysis results was an aspect of the tests. It should be emphasized that the two methods GMM and FCM, which allow soft clustering, force the smallest discrepancies to the nominal surface. Just as with the DBSCAN method the distribution of the clusters is crucial for computing final results. Adding to this is the strikingly strong dependence of the weights in the feature vector. Also the selection of a proper number of clusters is so far not clearly evident.

Table 3.5 shows another overview of the second exclusion in context with the performance in the deformation analysis approach. The K-means is excluded not only because of the aim to replace it in the analysis approach. It also provides the biggest results in comparison with the other methods. Due to the good results of GMM and FCM, they will be considered again. The K-medoids method is also considered again because of the better results than K-means.

Table 3.5.: Overview of the second exclusion

| method       | usability     |
|--------------|---------------|
| K-means      | $\mathcal{X}$ |
| FCM          | ✓             |
| GMM          | ✓             |
| DBSCAN       | $\mathcal{X}$ |
| K-medoids    | ✓             |
| spectral     | ✓             |
| hierarchical | ✓             |

### 3.2.3. Discussion

As a last check the best constellation from the clustering methods were executed 5 times repeatedly. The result reflected the statement done in section [3.2.1](#). In summary, the two clustering methods, spectral clustering and hierarchical clustering, are suitable methods for the deformation analysis method. Because their clustering result and their final results were exactly the same every time. The following table [3.6](#) shows the final exclusion.

Additionally, as mentioned in [3.2.2](#), the remaining discrepancies to the nominal surface with the Fuzzy C-Means clustering algorithm and the Gaussian mixture model are the smallest. This intensified the idea to combine cluster methods to gain the advantages of the soft clustering.

Table 3.6.: Overview of the final exclusion

| method       | usability     |
|--------------|---------------|
| K-means      | $\mathcal{X}$ |
| FCM          | $\mathcal{X}$ |
| GMM          | $\mathcal{X}$ |
| DBSCAN       | $\mathcal{X}$ |
| K-medoids    | $\mathcal{X}$ |
| spectral     | ✓             |
| hierarchical | ✓             |

The cluster result is not the only factor, which has a high influence on the final results. Two examples of the others are the calculation of the standard deviations to normalize the residuals to the trend surface and the approximation of the autocorrelation and cross-correlation functions. In sections [3.3](#) and [3.5](#) the calculation of the standard deviations is described in detail. The influence of the correlation functions is briefly discussed in section [4](#) in respect to the chosen data sets. The detailed discussion is part of future work.

### 3.3. Hard clustering implementation

In this section the modelling of the corresponding standard deviations of each cluster with respect to the hard clustering implementation will be discussed. In [\[Harmening and Neuner, 2020\]](#), it was found that the change of a locally stationary standard deviations can be used to produce deformations on the surface. For that only the residuals to the trend are needed. The process is assumed to be a locally stationary stochastic process. This consists of two parts: a stationary correlation function and a slowly varying scale. Both aspects are assumed to be normally distributed. In this work the stationary correlation function used in the deformation analysis approach are taken over. However, the slowly varying scale represented by the locally stationary standard deviation of each cluster is influenced by the clustering result. Therefore, the calculation of the locally stationary standard deviation is an important component for evaluating the modelling of the signal.

Figure 3.2 shows the clustering result of the z-coordinate of epoch  $e = 3$ . Each cluster represents an area with similar residuals. Therefore, a representative standard deviation for each cluster  $\sigma_{x,c}^e$  can be determined. For this purpose, the absolute value of the maximum residual of the cluster is divided according to the 3-sigma rule (equation (3.8)).

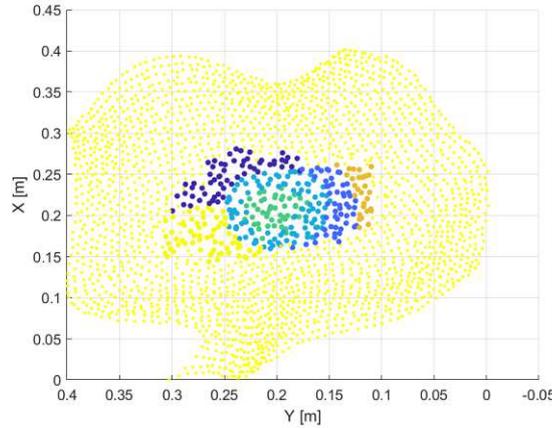


Figure 3.2.: Hard clustering: clustering result

$$\sigma_{x,c}^e = \frac{\max(|r_{x,c}^e|)}{3} \quad (3.8)$$

with

|                  |     |   |
|------------------|-----|---|
| $\sigma_{x,c}^e$ | ... | representative standard deviation per cluster |
| $r_{x,c}^e$      | ... | residuals per cluster                         |
| $x$              | ... | coordinate (x,y,z)                            |
| $c$              | ... | cluster number                                |
| $e$              | ... | epoch   |
| $ \cdot $        | ... | absolute value                                |

Due to the fact that the signal exists only in the distorted area a representative standard deviation of the stable area is not needed. Finally, the residuals of each point were normalized by the related representative standard deviation. Figure 3.3 shows the calculated standard deviations in the z-coordinate of epoch  $e = 3$ . The residuals to the trend surface in the z-coordinate of epoch  $e = 3$  are shown in figure 3.4.

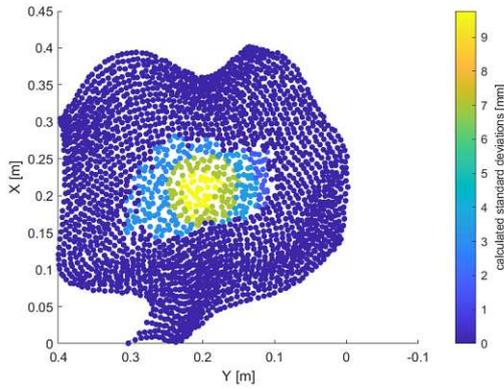


Figure 3.3.: Hard clustering: calculated standard deviations

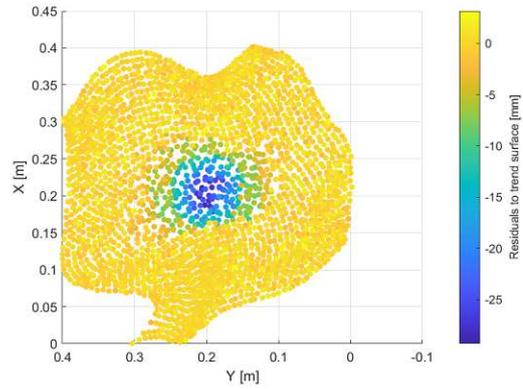


Figure 3.4.: Hard clustering: residuals to trend surface

The main objective is to reproduce the progression of the residuals to the trend surface with the calculated standard deviations. The result looks promising. However, the process is described with a slowly varying scale. Now, the representative standard deviation are calculated for each cluster. Therefore, the value changes abruptly. This causes sharp edges around each cluster. Furthermore, the edge region also has a gap between the calculated standard deviations. For improving this crossover the soft clustering method was investigated (see section 3.5).

Figure 3.5 shows the schematic structure of the initial and updated clustering process. The initial one is the approach introduced in Harmening and Neuner, 2020. The updated approach has its only difference in the applied clustering method. The figure states that in the updated approach the combination of the two clustering methods, spectral clustering and Gaussian mixture model, is used. This is the outcome of the investigation, where the hierarchical method and spectral method were combined with the Gaussian mixture model. It is described in the following section 3.4.

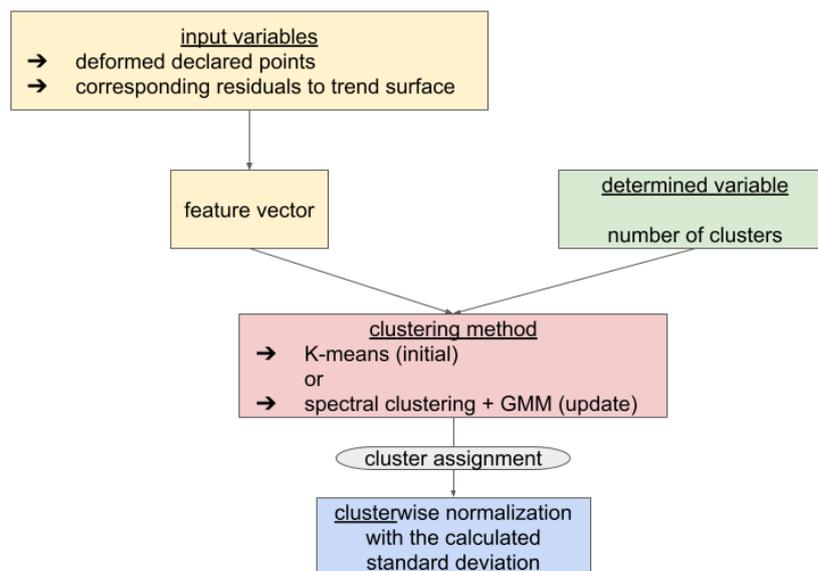


Figure 3.5.: Initial and updated approach

### 3.4. Combination of clustering methods

During the test of the clustering methods within the deformation analysis approach, the additional parameters for each methods described in [The MathWorks, Inc., 2022i] where also briefly investigated. Due to the good results of the Gaussian mixture model (GMM) and the Fuzzy C-Means clustering algorithm (FCM) section 3.2.2 these two were closely looked at regarding the implementation in the analysis approach. Especially, the function for the GMM has the advantage to set a previous clustering as a starting parameter. For the FCM this is not possible. As mentioned above, implementing the soft clustering was an additional aspect which should be tested in the deformation analysis approach. Also the transition from a deformed to a stable region could then be considered in the clustering result. The reason why soft clustering would be an advantage lies in the description of the underlying process. It is discussed in section 3.5. But as a first test only the deformed area with hard clustering implementation was investigated.

Since the spectral clustering method and the hierarchical clustering method are the most suitable (see section 3.2.3), these will be tested in combination with the Gaussian mixture model. The same aspects and feature vectors used in the test in section 3.2.2 were used for the combinations of clustering methods. Also the table 3.7 has the same structure.

Table 3.7.: Overview of results with combined clustering methods

| number<br>of<br>clusters | method                               |                       |  |                       |
|--------------------------|--------------------------------------|-----------------------|--|-----------------------|
|                          | spectral and GMM [mm]<br>with weight |                       | hierarchical and GMM [mm]<br>with weight |                       |
|                          | $a = 1000$<br>$b = 0.01$             | $a = 1$<br>$b = 0.05$ | $a = 1000$<br>$b = 0.01$                 | $a = 1$<br>$b = 0.05$ |
| 5                        | -5 - 5                               | -2 - 4                | -4 - 2                                   | - -                   |
| 6                        | - - -                                | <b>-2 - 2</b>         | -4 - 2                                   | - - -                 |
| 7                        | -5 - 10                              | -2 - 4                | -5 - 5                                   | - -                   |
| 8                        | - - -                                | -100 - 100            | -5 - 5                                   | - - -                 |
| 9                        | -5 - 5                               | -10 - 10              | -4 - 2                                   | - - -                 |
| 10                       | -20 - 20                             | -5 - 10               | <b>-2 - 2</b>                            | - - -                 |
| 11                       | -40 - 60                             | -6 - 4                | -5 - 5                                   | - - -                 |
| 12                       | -5 - 10                              | -5 - 5                | -20 - 20                                 | - - -                 |

In contrast to the results from the test in section 3.2.2, the spectral clustering produces final results with one weighting of the feature vector as well as with the other. However, the hierarchical methods only works with one weighting. This shows, that the feature vector has different impact on each clustering method and combination. Which opened another question regarding the feature vector. The further investigations are described in section 3.7.

However, both combinations could stabilize the analysis results also after repeated executions. The following two figures (3.6 and 3.7) show the clustering result with the best constellation of the combination in the z-direction of epoch  $e = 3$ . For comparison, the clustering result of the K-means method is shown in figure 3.8.

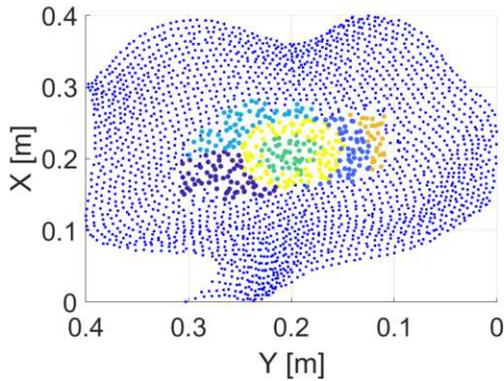


Figure 3.6.: Updated clustering process: spectral + GMM method (number of clusters: 6)

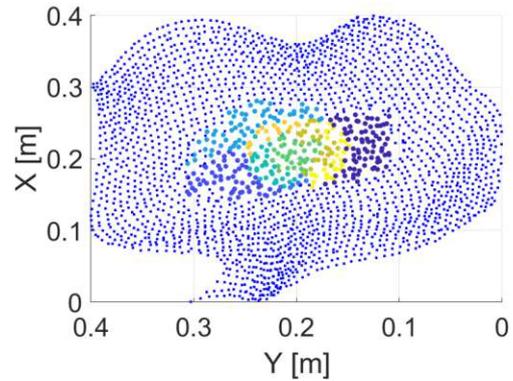


Figure 3.7.: Updated clustering process: hierarchical + GMM method (number of clusters: 10)

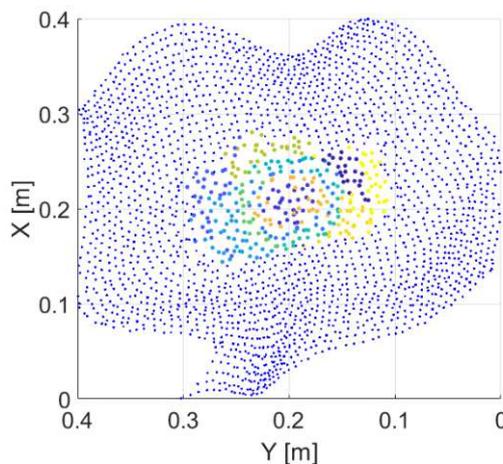


Figure 3.8.: Initial clustering process: K-means method

Finally, the clustering result of the combination of spectral clustering with the Gaussian mixture model reflects the deformation slightly better than with the hierarchical clustering. Especially, the yellow and green clusters in figure 3.6 show the conical deformation caused by the shift of the control point of the B-Spline surface of the trend. Therefore the combination of spectral clustering and Gaussian mixture model is chosen as the most suitable alternative to K-means.

### 3.5. Soft clustering implementation

Since the combination of clustering methods works, it is now feasible to benefit from the properties of the Gaussian mixture model. As described in section 2, the posterior probability of points in a cluster are distributed according to the Gaussian distribution. Therefore, it is now possible for points to be assigned to multiple clusters. Especially the edge regions of a cluster gain smoothness. Consequently, the clustering result can reflect the shape of the locally stationary stochastic process more accurately.

The results of testing the Gaussian mixture model in combination with the spectral clustering method on the deformed part of the point cloud are described in section 3.4. There it was used as a hard clustering implementation. Since it makes no difference on the clustering result when using soft clustering, no additional test is needed. Although, the posterior probabilities of the data points regarding to the clusters influences the calculation of the representative standard deviations.

However, also the edge region from the stable to the deformed region should be included in the clustering process. For this two feature vectors, each for one clustering method is defined. Therefore, the clustering process is split up in two steps. The feature vector for the spectral clustering method consists of the values belonging to the deformed declared points. The second feature vector, the one for the Gaussian mixture model, is basically an extension of the first feature vector. The values added are associated with the stable declared points in the vicinity of the deformed area. Additionally, the clustering result of the spectral clustering method is set as starting parameter for the Gaussian mixture model. The added points in the second feature vector are introduced as one additional cluster. Therefore, the number of cluster has to be increased by one.

This approach is the result of testing various combinations of properties in the clustering methods and different uses of the feature vector. When pre-clustering the deformed area, the focus in the clustering result remains on this region. During the tests the importance of choosing a suitable number of clusters became evident. Also the specification of a feature vector proved to be challenging. The results are presented in sections 3.6 and 3.7. The influence of adding the surrounding points is discussed in section 3.8.

Figure 3.9 shows the clustering result. It is noticeable that the edges of the clusters are blurrier than with the hard clustering implementation in the updated approach. This provides additional support in modelling the locally homogeneous standard deviations.

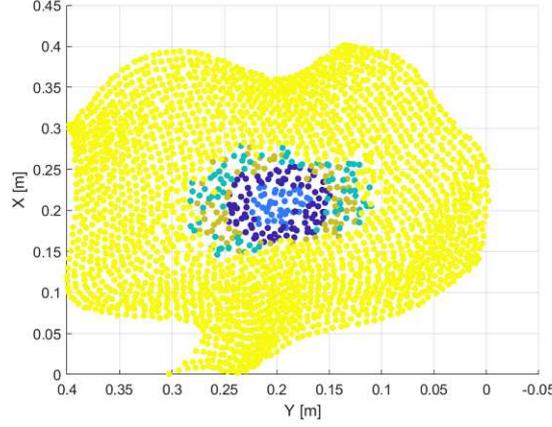


Figure 3.9.: Soft clustering: clustering result

In the modification the posterior probability for each point is included. After finding  $\sigma_{x,c}^e$  (equation (3.8)), the representative standard deviation per point  $\sigma_{x,point}^e$  is computed. For this purpose, the sum of the posterior probabilities of the representative variances  $v_{x,c}^e$  is calculated. Equations (3.9) and (3.10) show the calculation depending on the membership of the point to the deformed and stable regions. The final step is to normalize the residuals with the representative standard deviation for each point.

$$\sigma_{1,x,point}^e = \frac{\sqrt{\sum_{k=1}^{c_{max}-1} p_k^e \cdot v_{x,c,k}^e}}{\sqrt{\sum_{j=1}^{c_{max}-1} p_j^e}} = \frac{\sqrt{\sum_{k=1}^{c_{max}-1} p_k^e \cdot \sigma_{x,c,k}^{e2}}}{\sqrt{\sum_{j=1}^{c_{max}-1} p_j^e}} \quad (3.9)$$

$$\sigma_{2,x,point}^e = \frac{\sqrt{\sum_{k=1}^{c_{max}-1} p_k^e \cdot v_{x,c,k}^e}}{\sqrt{\sum_{j=1}^{c_{max}} p_j^e}} = \frac{\sqrt{\sum_{k=1}^{c_{max}-1} p_k^e \cdot \sigma_{x,c,k}^{e2}}}{\sqrt{\sum_{j=1}^{c_{max}} p_j^e}} \quad (3.10)$$

with

|                        |     |   |
|------------------------|-----|---|
| $\sigma_{1,x,point}^i$ | ... | representative standard deviation per point<br>point is included in the deformed area               |
| $\sigma_{2,x,point}^i$ | ... | representative standard deviation per point<br>point is included in both areas, stable and deformed |
| $x$                    | ... | coordinate (x,y,z)  |
| $c$                    | ... | cluster number  |
| $c_{max}$              | ... | maximal number of clusters  |
| $p$                    | ... | posterior probability   |
| $e$                    | ... | epoch   |

By adding points from the stable region, these must also be taken into account when normalizing the residuals to the trend surface. With the implementation of the soft clustering the points can change their affiliation to the deformed or stable area after the clustering. In this approach, the assignment before the clustering process is used to normalize the residuals to the trend surface. However, the assignment after the clustering process is used for further calculations.

The figures 3.10 and 3.11 shows the calculated standard deviations and the residuals to the trend surface. Both show the results in the z-coordinate of epoch  $e = 3$ . In contrast to the results with the updated approach the edges of the clusters are smoother and blend into each other. The effects in the final results will be discussed in section 4.

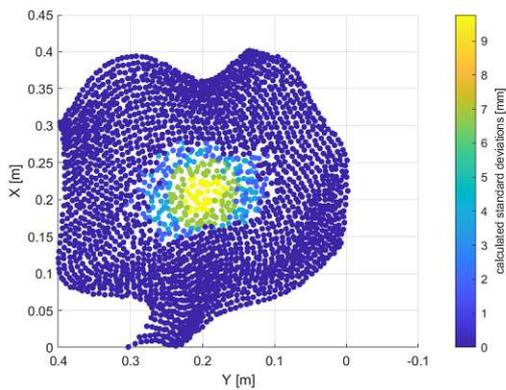


Figure 3.10.: Soft clustering: calculated standard deviations

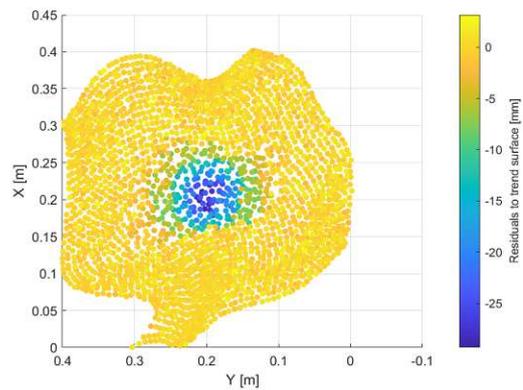


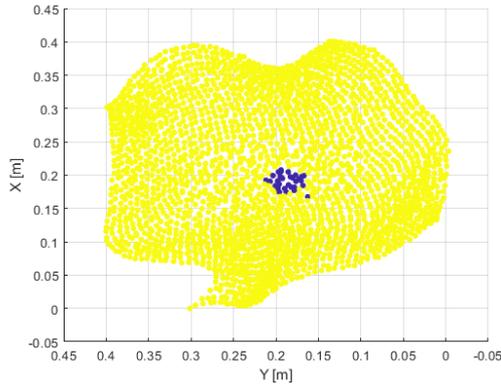
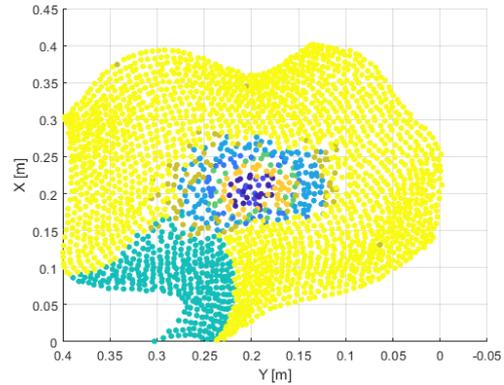
Figure 3.11.: Soft clustering: residuals to trend surface

### 3.6. Number of clusters

Undoubtedly, the choice of the number of clusters has a significant impact on the result. The purpose is to determine this number automatically for each coordinate per epoch. However, determining this number is the most challenging task in applying a clustering method to a data set [Aggarwal and Reddy, 2018] [Xu and WunschII, 2005]. In the literature, several possible methods for calculating the number of clusters are given. One selection is: Akaike Information Criterion (AIC), Bayes Information Criterion (BIC), Calinski-Harabasz Index, Gap Statistic, Silhouette Coefficient [Aggarwal and Reddy, 2018] [Xu and WunschII, 2005] [The MathWorks, Inc., 2022d].

Several methods for choosing a suitable number of clusters were tested. Unfortunately, the methods failed for consistency in the three coordinate directions. The biggest challenge was the x-direction. Since the deformations are lowest there, the number of clusters was usually underestimated. On the other hand, the number in the z-direction was sometimes overestimated in epoch 3. The figures 3.12 and 3.13 show examples of poor clustering results.

During a literature search for finding more methods, the *subclust* [The MathWorks, Inc., 2022k] method was discovered. This MATLAB<sup>®</sup> method uses a subtractive clustering algorithm. First, the likelihood of being a cluster center for each point is calculated. It is referred to the density of surrounding data points. Then the first cluster center is chosen by the highest potential being

Figure 3.12.: Number of clusters: underestimated  
x-directionFigure 3.13.: Number of clusters: overestimated  
z-direction

one. Afterwards, the neighbours are deleted as potential cluster centres. The criterion defining the neighbourhood is called *clusterInfluenceRange*. It was empirically set to the value 0.6 in this work. In the remaining points the next cluster center is found and the neighbours are excluded. The algorithm is finished when all data point have a cluster center in their influence range.

In addition, it should be mentioned that the Gaussian mixture model method reduces the number of clusters when it is to high. This characteristic was suppressed by the above described method. The clustering results was also improved as a result. The determination on this method was made in the progression of the investigations on the feature vector and the transition area of the deformed area. These are described in the following two sections (3.7 and 3.8).

### 3.7. Feature vector adjustment

Another aspect, which was considered in the course of the work, is the adjustment of the feature vector for clustering. While searching for a suitable solution for calculating the number of clusters, it became apparent how strongly this was related to the change in the weights in the feature vector. After defining the clustering process with the combination of spectral clustering and the Gaussian mixture model (section 3.4), it became clear that this makes the weighting in the feature vector obsolete.

But the clustering results in combination with the innovations of the deformation analysis approach were not satisfying. Therefore, other constellations of the vector were tested. Since the deformation analysis approach separates the coordinates and the epochs a feature vector which fulfills this criterion has to be found. The equation (3.7) shows the feature vector which was used in [Harmening and Neuner, 2020]. The basic principle is to cluster the residuals to the trend surface. To localize the data point onto the surface the three coordinates vectors were added. During the tests it was found that the coordinate vector of the corresponding direction to the residual vector has the main impact on the clustering result. Consequently, only the supplementary coordinates vectors were added to the feature vector. The comparison is shown in the figures 3.14 and 3.15.

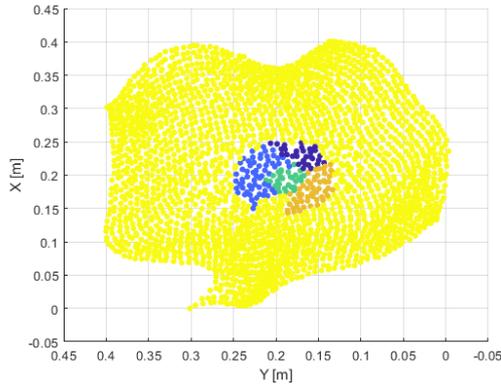


Figure 3.14.: Feature vector: all three coordinates directions are used

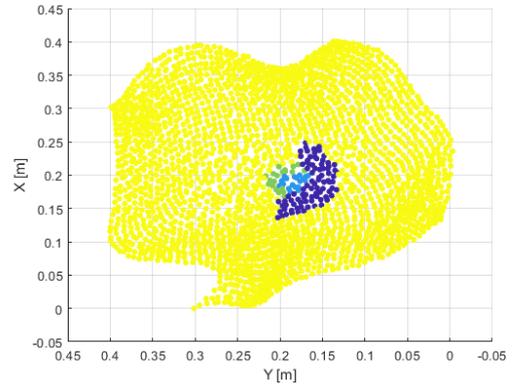


Figure 3.15.: Feature vector: the supplementary coordinates directions are used

Obviously, the number of clusters are changing with the different feature vectors. In addition, the clustering result reflects the progression of the residuals better in figure 3.15. There the cluster in the center is divided which increases the knowledge of the progression of the residuals. However, the darkblue cluster is not describing the distribution of the residuals to the trend surface. According to that an additional variable has to be found which leads to a more accurate similarity graph for the spectral clustering.

Figure 3.16 shows the clustering result with the final constellation of the feature vector. The variable used instead of the corresponding coordinate consists of the absolute values of the residuals to the trend surface. It was challenging to find it, but due to the fact the representative standard deviations are modelled according to the absolute value of the maximum residual it was then obvious to try out the variable. Furthermore, the absolute value can also be interpreted as the distance of the point to the trend surface. The final constellation of the feature vector for the x-direction is shown in equation (3.11). For comparison, the corresponding residuals to the trend surface are shown in figure 3.17.

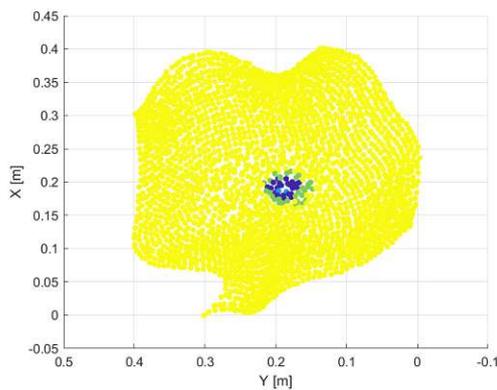


Figure 3.16.: Feature vector: final constellation

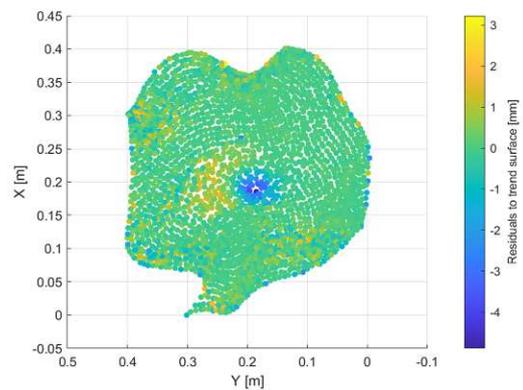


Figure 3.17.: Feature vector: residuals to trend surface

$$\underline{f}_x^e = \left[ \underline{r}_x^e \quad |\underline{r}_x^e| \quad \underline{y}^e \quad \underline{z}^e \right] \quad (3.11)$$

with

$\underline{f}_x$  ... feature vector in x-direction

$\underline{r}_x$  ... residuals to trend surface in x-direction

$\underline{x}$  ... vector of x-coordinates

$e$  ... epoch

$|\cdot|$  ... absolute value

### 3.8. Edge region of the deformed area

While setting up the new clustering approach the question about the edge region between the deformed and stable area raised. Because of the Gaussian mixture model which should find clusters according to the Gaussian distribution the points with high residuals were clustered across the whole data set. Although, the deformed part is located in the center of the object points at the edge of the object were also clustered as deformed (see figure 3.18). These higher values are related to the trend estimations in the first epoch. The aim was to filter these points which are away from the deformed part. The idea is to find the biggest influence range of the deformation according to the stable area. Since the x-direction causes the most clustered points at the edges of the object because of the smallest deformation in this direction. Thus, the focus was set on this coordinate direction.

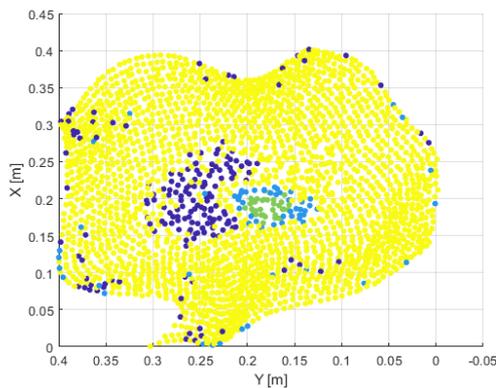


Figure 3.18.: Edge region: the whole data set is used for the clustering process

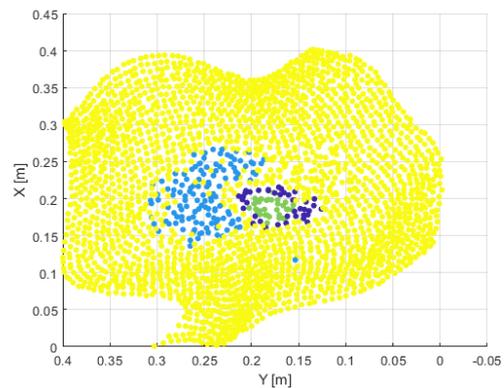


Figure 3.19.: Edge region: a part of the data set is used for the clustering process

The stable declared points which take part at the clustering process are found by a nearest neighbour search routine. For that, the neighbours of the deformed declared points are searched. The number of them are two times the number of all deformed points. This value has been determined purely empirically with focus on the x-coordinate. Figure 3.19 shows the clustering result.

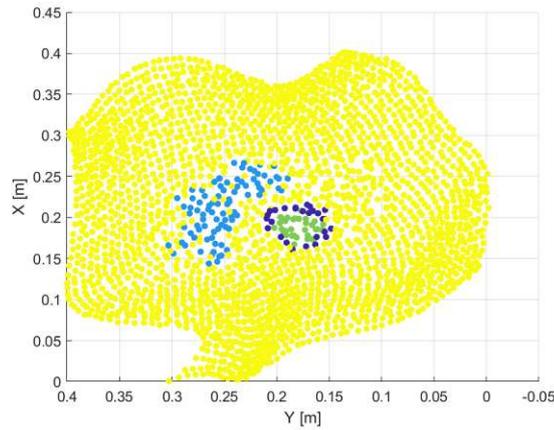


Figure 3.20.: Edge region: final clustering result

For the z-coordinate in epoch 2 the approach shows the worst results. Therefore, after the clustering process another nearest neighbour search is applied to delete clustered points in the surrounding. However, this filtering also affects the cluster edges of the deformed area (see figure 3.20). This improvement is part of future work.

### 3.9. Summary of the new approach

In this section the implemented algorithm is described. In addition, an overview of the selected parameters is given. Figure 3.21 shows the schematic structure of the new approach.

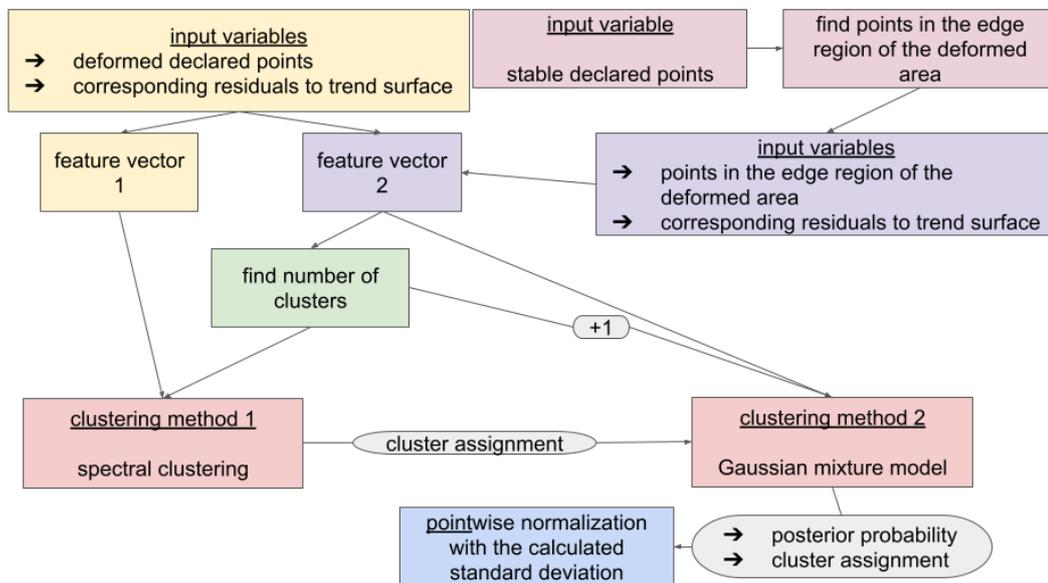


Figure 3.21.: New approach

First of all, the variables which contains the indices about the stable declared points and the deformed ones are taken over. The only additional variable needed contains the residuals to the trend surface.

The first step in the clustering process finds the neighbours of the deformed points in the stable area (light purple boxes). The number of these neighbours is defined to be two times the number of deformed points. As mentioned in section 3.8 the value is determined empirically. Next, the two feature vectors can be set up (yellow and purple boxes). One only with the values regarding to the deformed declared points. The second one is an extension of the first one with the neighbours of the deformed declared points. The form of the feature vectors is shown in equation (3.11). Afterwards, the number of clusters is found by the *subclust* method [The MathWorks, Inc., 2022k]. The extended feature vector (feature vector 2, purple box) and the *clusterInfluenceRange* with the value 0.6 are used as parameters. This procedure produces the most suitable clustering result with respect to the progression of the residuals to the trend surface.

For the spectral clustering method (see section 2.2, left red box) the feature vector describing the deformed part is used. The distance calculation is set to the Euclidean distance. The number of nearest neighbours is the same as above. Both influence the structure of the similarity graph. The type of the similarity graph is the k-nearest neighbour graph. The K-medoids method is chosen for clustering the eigenvectors of the Laplacian matrix. The number of clusters is that one from the *subclust* method. Parameters which are not mentioned are the default ones described in [The MathWorks, Inc., 2022j].

The following Gaussian mixture model (right red box) clusters all points to be clustered (feature vector 2). Since the selected stable points are now added to be clustered the number of clusters is increased by one. The clustering result of the spectral clustering is used as the first estimate of the cluster membership. The selected stable points are introduced as one cluster. Additional parameters are the *CoveType* which is set to *diagonal* and the *SharedCovariance* which is set to *true*. It follows, that the fitted covariance matrices in the algorithm is a diagonal matrix. Additionally, they are assumed to be identical. Thus, the data are assumed to be uncorrelated and equally weighted.

After the clustering process the result is filtered with a nearest neighbour search to find clustered points in the stable area, which are obviously surrounded by stable points. Then the representative standard deviations are calculated as described in section 3.5 (bottom gray and blue box). The following parts of the deformation analysis algorithm is taken over as described in [Harmening and Neuner, 2020]. Only the approximations of the autocorrelation functions and cross-correlation functions were adapted to the progression of the newly calculated ones.

## 4. Data and results

In this section the analysis results of the new approach (figure 3.21) with respect to the initial and updated approach (figure 3.5) are shown and discussed. In addition, the outcome is shown with further data sets. For reasons for better readability, only few representative figures are shown in this section. All remaining plots are listed in the appendices.

### 4.1. Data set D1

Data set (D1) is taken over from [Harmening and Neuner, 2020]. However, the figures are of an later execution of the algorithm. Therefore the results pictured in the figures referred to the initial approach are not equal to those in [Harmening and Neuner, 2020].

The simulated data set consists of three epochs. The first one has no deformed parts and is realised by a B-Spline surface. For the second and third epoch the control point  $P_{4,5}$  is shifted in the z-coordinate. The figure 4.1 shows the control point  $P_{4,5}$  red circled and its location in relation to the B-spline surface. Table 4.1 shows the shift of the control point per epoch with respect to the previous epoch. An illustration of the deformation in the middle of the point cloud is given in figure 4.2.

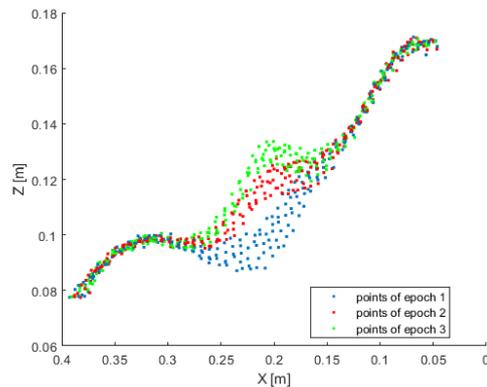
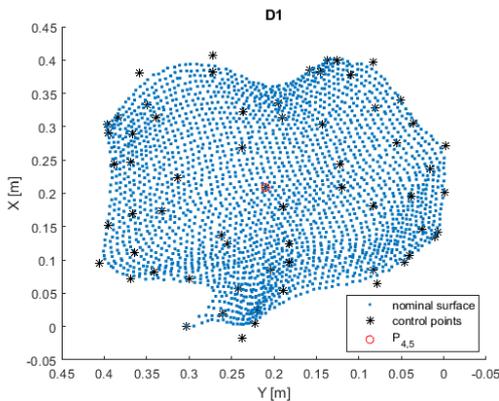


Figure 4.1.: D1: trend surface with control points      Figure 4.2.: D1: illustration of the deformation in the point cloud

Table 4.1.: Overview of the shift each epoch of  $P_{4,5}$  in the z-coordinate

| epoch | $P_{4,5}$ shift [mm] |
|-------|----------------------|
| 0     | 0                    |
| 1     | 50                   |
| 2     | 25                   |

## 4.2. Comparison of the clustering result

In figure 4.3 one clustering result of the initial approach is shown. All figures of the results of the dataset can be found in the appendix A. It was created by the application of the K-means clustering method. Only the deformed declared points with equation (3.6) were used for the clustering process. In the center of the clustered area the clusters form concentric circles. This reflects the deformation of the surface. However, the clusters at the edge region of the clusters only suggest a circular shape.

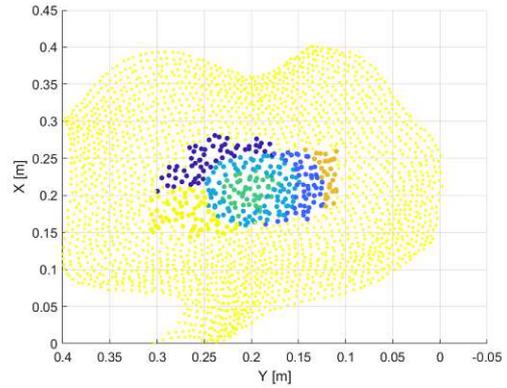
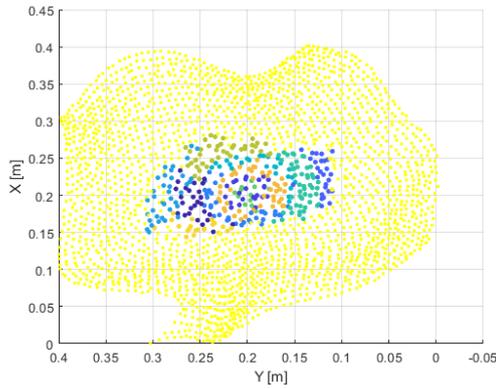


Figure 4.3.: Clustering result: z-direction of epoch 3 - initial approach  
Figure 4.4.: Clustering result: z-direction of epoch 3 - updated approach

In comparison, the clustering result of the updated approach can be seen in figure 4.4. For complete figures on the results of the dataset, see appendix B. Also for this only the deformed declared points are used. The division of the clusters is similar to the K-Means method but they have a more compact form. Also the cluster number decreased. This is due to the Gaussian mixture model, which attempts to create clusters with respect to Gaussian distributed posterior probabilities for the membership of a cluster per point.

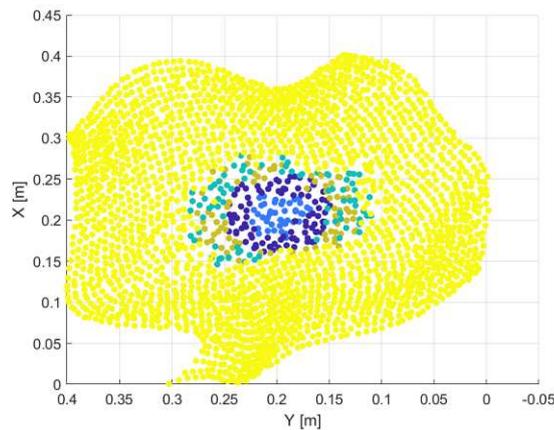


Figure 4.5.: Clustering result: z-direction of epoch 3 - new approach

In contrast, the clustering result of the new approach better reflects the progression of the residuals to be clustered (see figure 4.5 and appendix C). All clusters have the form of concentric circles. One reason of the blurred crossover between the clustered and not clustered (yellow points) area is the use of the Gaussian mixture model as second clustering method. A second reason is that points declared stable were also included in the clustering process. Overall, the clustering result in the new approach reflects the progression of the residuals to the trend surface in more detail.

### 4.3. Comparison of the calculated standard deviations

The aim of the clustering process is to reconstruct the progression of the residuals to the trend surface. These are shown for the z-direction in the third epoch in figure 4.6. In figure 4.7 the result of the initial approach is pictured. The main difference to the results of the updated approach (figure 4.8) is in the middle of the deformed area. It is caused by the varying number of clusters in this area. In the edge region to the stable region are no significant differences.

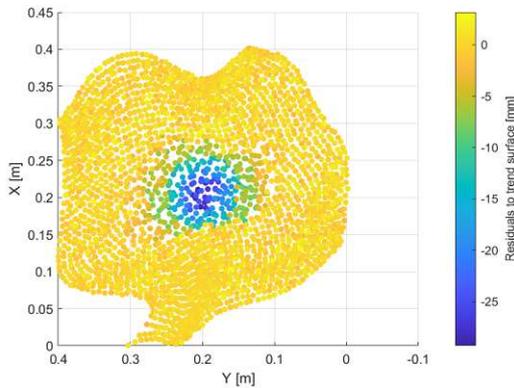


Figure 4.6.: Calculated standard deviations:  
z-direction of epoch 3 - residuals to  
trend surface

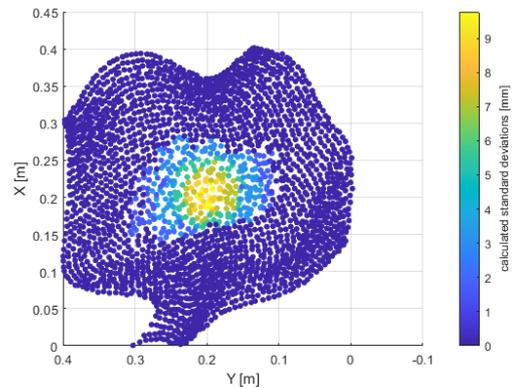


Figure 4.7.: Calculated standard deviations:  
z-direction of epoch 3 - initial approach

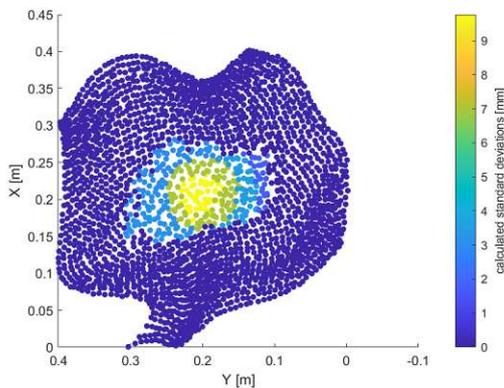


Figure 4.8.: Calculated standard deviations:  
z-direction of epoch 3 - updated  
approach

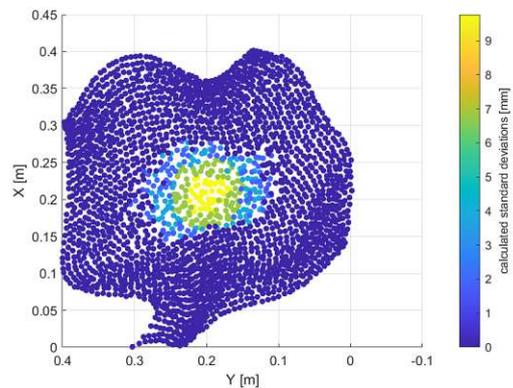


Figure 4.9.: Calculated standard deviations:  
z-direction of epoch 3 - new approach

Additionally, the result of the new approach is shown in figure 4.9. In contrast to the other results the edge region to the stable area better reflects the progression of the residuals. However, the area the highest deformation is also separated from the smooth region all around. In summary, the edge region is better approximated with the new approach. The initial approach, on the other hand, reproduces the maximum deformation better.

#### 4.4. Comparison of the analysis results

A quick look at the analysis results shows that the new approach basically works well. However, it is interesting to note that the z-direction with the largest deformation gives better results than the y-direction. One possible reason is that the approximation of the autocorrelation function is worse in the y-direction than in the z-direction. The figures 4.10 and 4.11 show the discrepancies to the nominal surface in comparison to the initial and new approach. In the results of the new approach structures describing the edge of the deformed area clearly can be seen. This additionally indicates that the normalization of the residuals with the representative standard deviations per data point is not yet fully developed. But also the approximation of the calculated autocorrelation functions is not yet optimized. In this case the approximation is worst in the y-direction, which is also reflected in the results.

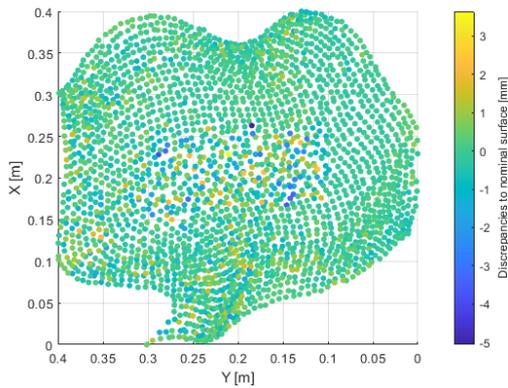


Figure 4.10.: D1: y-direction discrepancies to nominal surface - initial approach

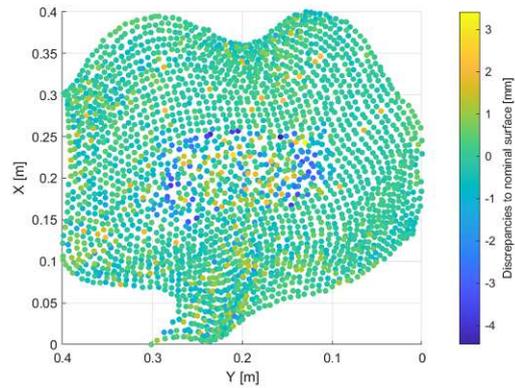


Figure 4.11.: D1: y-direction discrepancies to nominal surface - new approach

The **Cumulative Distribution Functions (CDF)** allow easier comparison between the three versions of the deformation analysis algorithm: initial approach, updated approach and new approach. They are shown in the figures 4.12 - 4.14 with respect to the residuals of the collocation.

Unsurprisingly, the x-direction has residuals within the smallest range. The majority of the residuals lies in the interval from -1 to 1. In the two other directions, the majority of the values are larger by 2 mm overall. The z-direction shows the largest absolute values with up to 8 mm. In contrast to the y-direction, these are symmetrically distributed around the mean value. The residuals of collocation in the y-direction of the new approach have larger negative values than positive. This can also be seen in the figure of the residuals of collocation (4.15). The structures appear weaker than in figure 4.11. The comparison shows that the deformation is still mapped in the noise.

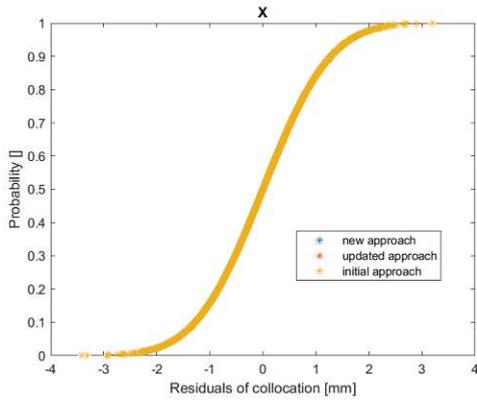


Figure 4.12.: Cdf: x-direction

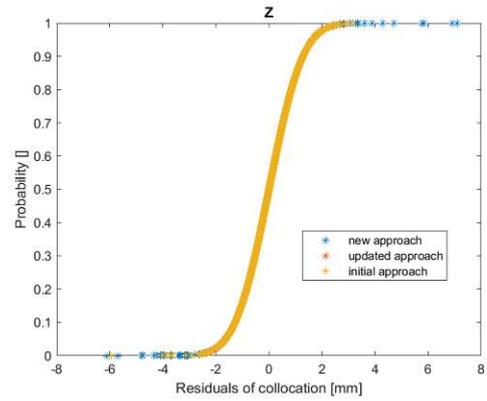


Figure 4.13.: Cdf: z-direction

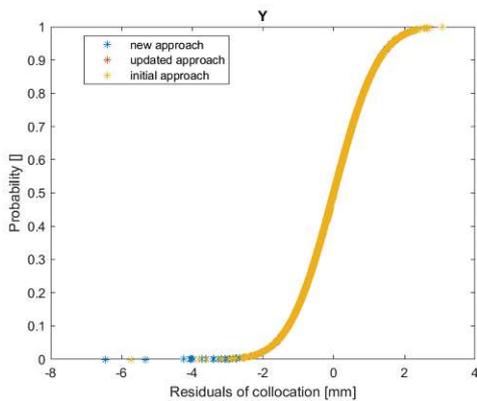


Figure 4.14.: Cdf: y-direction

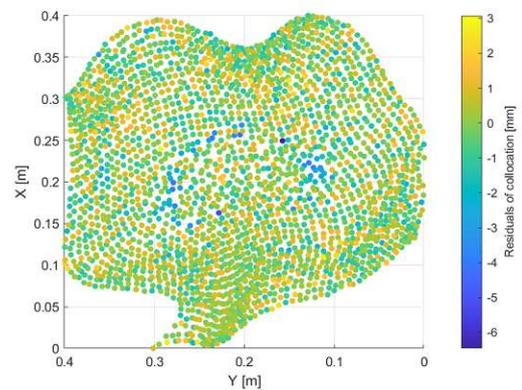


Figure 4.15.: D1: y-direction residuals of collocation - new approach

### 4.5. Further data sets

To ensure that the developed method does not only work on one data set, others were tested. For this purpose, deformations were simulated at other locations of the point cloud by moving control points of the trend surface. In general, the results show the same phenomena as with the original data set. In data set D2 a single control point was shifted with the same values used for the data set D1 (see table 4.1). Due to the different curvature of the surface, there is a different deformation pattern. This is better represented by the representative standard deviation with the initial approach. The comparison can be seen in the figures 4.16 - 4.18.

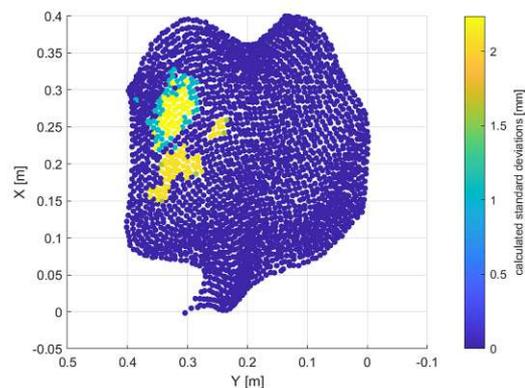
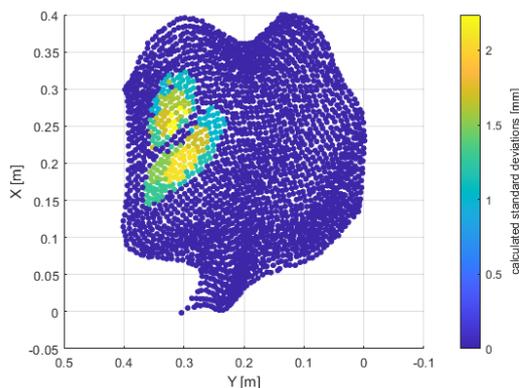


Figure 4.16.: D2: clustering result - initial approach Figure 4.17.: D2: clustering result - new approach

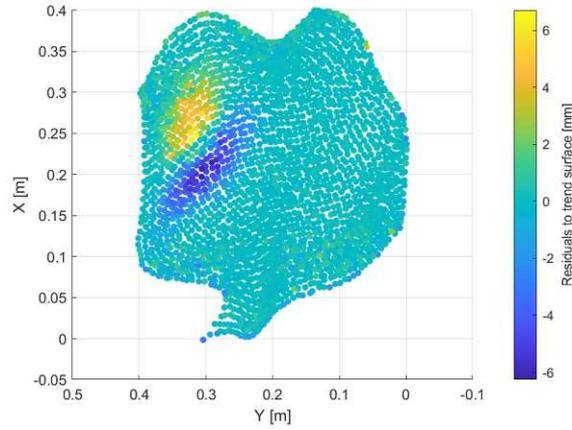


Figure 4.18.: D2: residuals to trend surface - x-direction epoch 3

The differences in the final results have the same behaviours as with the data set D1. Especially, structures describing the edge region of the deformed area occur also within the results of the additional data sets. The figures [4.19](#) and [4.20](#) show the difference between the initial approach and the new one in the residuals of collocation in z-direction. It is noticeable that they are mainly visible in the z-direction, where the greatest deformation takes place. It follows that the normalization of the residuals to the trend surface is not yet fully developed.

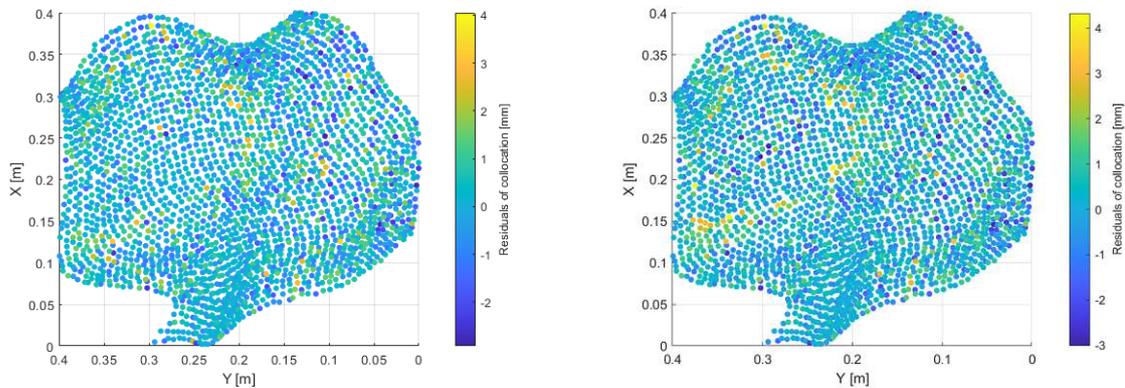


Figure 4.19.: D2: z-direction residuals of collocation - initial approach  
 Figure 4.20.: D2: z-direction residuals of collocation - new approach

In figures 4.21 - 4.25 a similar comparison is pictured. In this data set D3 two control points were shifted. The amount is equal to D1 and D2. However, the directions are opposite.

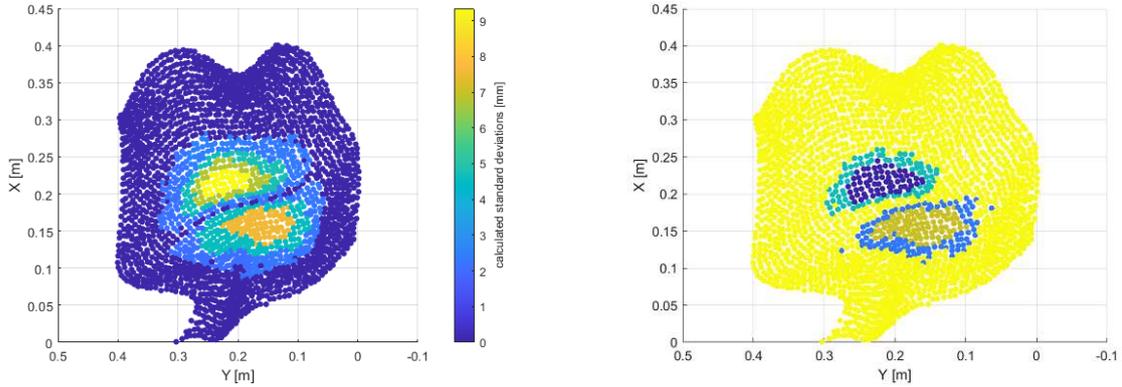


Figure 4.21.: D3: clustering result - initial approach    Figure 4.22.: D3: clustering result - new approach

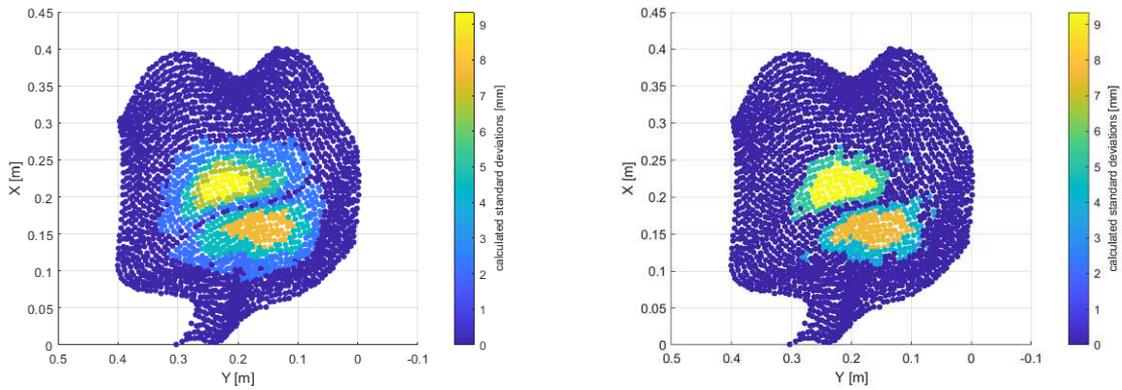


Figure 4.23.: D3: calculated standard deviations - initial approach    Figure 4.24.: D3: calculated standard deviations - new approach

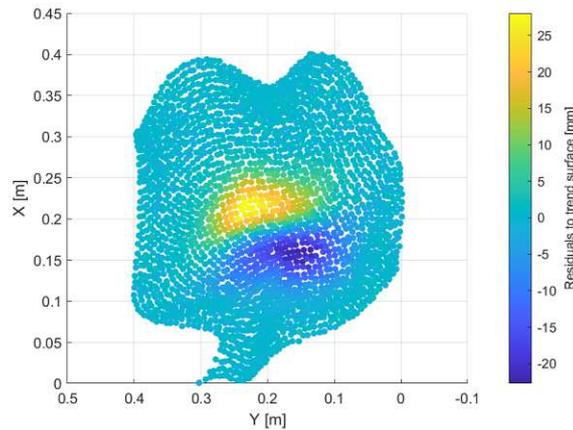


Figure 4.25.: D3: residuals to trend surface - z-direction epoch 3

A third additional data set D4 tested within the two approaches contains two control points shifted in the same direction. Again, the amount is the same as in D1. The comparison of the calculated standard deviations is shown in the figures 4.26 - 4.28. Even though the residuals are better represented by the new approach, the structures in the collocation residuals are clearly visible. These can be seen in figure 4.29.

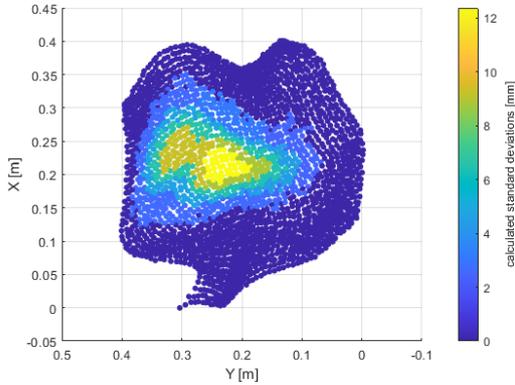


Figure 4.26.: D4: calculated standard deviations - initial approach

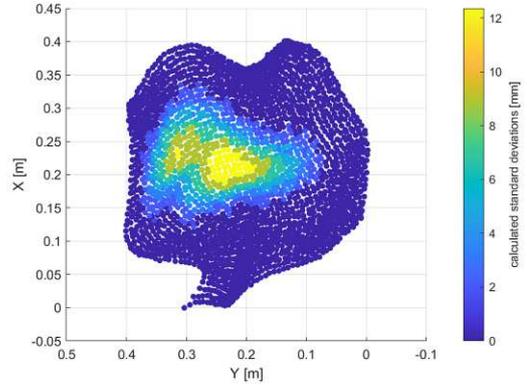


Figure 4.27.: D4: calculated standard deviations - new approach

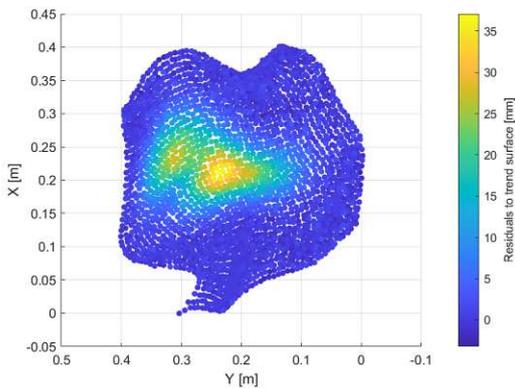


Figure 4.28.: D4: residuals to trend surface - z-direction epoch 3

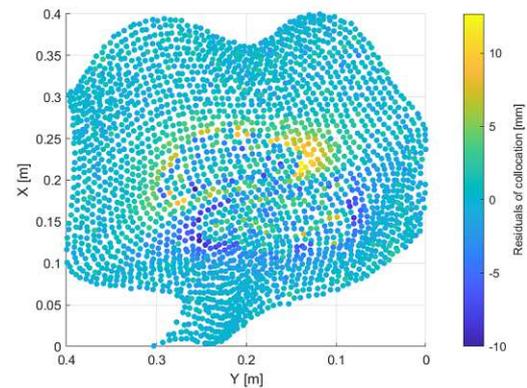


Figure 4.29.: D4: z-direction residuals of collocation - new approach

For completeness, it should be mentioned that smaller maximum deformations were also tested. It can be said that if the deformations get smaller, the approach works less effectively. This is especially true for the clustering results. In conclusion, the approach presented in this thesis provides reliable results. The phenomena, which still need to be improved, occur similarly in all tested data sets. Compared to the initial approach the calculation routine for standard deviations is not always reflecting the actual phenomena. This part of the algorithm still needs to be improved.

## 5. Conclusion

This section summarizes the results and discussions during this work. In addition, an outlook for future work is given.

### 5.1. Summary

In [\[Harmening and Neuner, 2020\]](#) a new deformation analysis approach was published. This work discussed the part of finding homogeneous areas in the deformed region of the point cloud. Since the residuals to the trend surface based on the first measurement epoch contain information about the deformation of the point cloud, their modelling is a main component of the solution approach. Up to now, the K-means clustering method was used to define the homogeneous areas. However, the algorithm involves a random process, which has a significant impact on the final result of the deformation analysis algorithm. The main objective of this work was to find a suitable alternative for the clustering method.

This was found by combining two clustering methods. Firstly, the spectral clustering, which finds the membership of a point to a cluster based on a similarity graph of data points. The second method is called Gaussian mixture model. Its clustering results consist of clusters related to a Gaussian distribution.

In the search for a new clustering method, the calculation of the representative standard deviations was also considered. Until now, each cluster was represented by a single locally stationary standard deviation. However, this modelling did not reflect the underlying process. In this case, a slowly varying scale factor is assumed. But the modelled values change abruptly at the cluster edges and in the crossover to the stable region. Using the Gaussian mixture model, a soft clustering approach could be implemented. Consequently, the modelling of the locally stationary standard deviations could be adapted.

Another influence on the result was generated by the choice of the number of clusters. Since, there is no explicit method for this, to find a suitable one for this purpose was challenging. However, a MATLAB<sup>®</sup> supported function could be found, which provides appropriate results. During this investigation the influence of the feature vector and the high values of the residuals at the edges of the object became clear. However, a suitable approach which connects these three recognized factors is provided in this work.

Finally, the deformation analysis method introduced in [Harmening and Neuner, 2020] could be further developed. Now the final results remains the same, if the algorithm is executed repeatedly. Even if the discrepancies to the nominal surface are not improved related to the values in [Harmening and Neuner, 2020], multiple factors with an influence on them could be improved. Furthermore, there are other factors left, which are not discussed in this work.

## 5.2. Outlook

As mentioned above this work only deals with a part of the deformation analysis approach introduced in [Harmening and Neuner, 2020]. During this work the connection between each step of the algorithm became clearer. However, not all phenomena could be explained. Also the fact that the estimation of the autocorrelation functions and the cross-correlation functions are not investigated in dependence of the clustering result and the modelling of the standard deviations and their standardization so far, makes it difficult to classify the influence of them to the final results.

The new approach was tested with additional data sets. However, this was just a brief look if the clustering methods produce useful results. Therefore, more detailed tests with additional data sets need to be performed. Also a test with real measured data is missing up to now.

In conclusion, this work could improve some parts of the deformation analysis algorithm. However, the connection to the steps, which were not investigated, and their improvement is part of future work.

# List of Figures

|  |    |
|--|----|
| 2.1. K-means (number of clusters: 6)   | 10 |
| 2.2. K-means (number of clusters: 6) retry   | 10 |
| 2.3. K-means (number of clusters: 12)  | 10 |
| 2.4. K-means (number of clusters: 12) retry  | 10 |
| 2.5. K-medoids (number of clusters: 6)   | 11 |
| 2.6. K-medoids (number of clusters: 12)  | 11 |
| 2.7. Spectral clustering number of clusters: 3 distance metric: Euclidean similarity<br>graph: k-nearest neighbour Laplacian matrix: normalized random-walk                  | 13 |
| 2.8. Spectral clustering number of clusters: 3 distance metric: Euclidean similarity<br>graph: $\epsilon$ -neighbourhood (radius = 1) Laplacian matrix: normalized symmetric | 13 |
| 2.9. Gaussian mixture model (number of clusters: 3)  | 15 |
| 2.10. Gaussian mixture model with contours (number of clusters: 3)   | 15 |
| 2.11. k-distance graph ( $MinPts = 50$ )   | 16 |
| 2.12. DBSCAN clustering ( $\epsilon = 1.2$ )   | 16 |
| 2.13. Dendrogram with Euclidean distance   | 17 |
| 2.14. Hierarchical clustering (number of clusters: 6)  | 17 |
| 3.1. Clusters DBSCAN (number of clusters = 10)   | 24 |
| 3.2. Hard clustering: clustering result  | 26 |
| 3.3. Hard clustering: calculated standard deviations   | 27 |
| 3.4. Hard clustering: residuals to trend surface   | 27 |
| 3.5. Initial and updated approach  | 27 |
| 3.6. Updated clustering process: spectral + GMM method (number of clusters: 6)   | 29 |
| 3.7. Updated clustering process: hierarchical + GMM method (number of clusters: 10)  | 29 |
| 3.8. Initial clustering process: K-means method  | 29 |
| 3.9. Soft clustering: clustering result  | 31 |
| 3.10. Soft clustering: calculated standard deviations  | 32 |
| 3.11. Soft clustering: residuals to trend surface  | 32 |
| 3.12. Number of clusters: underestimated x-direction   | 33 |
| 3.13. Number of clusters: overestimated z-direction  | 33 |
| 3.14. Feature vector: all three coordinates directions are used  | 34 |
| 3.15. Feature vector: the supplementary coordinates directions are used  | 34 |
| 3.16. Feature vector: final constellation  | 34 |
| 3.17. Feature vector: residuals to trend surface   | 34 |
| 3.18. Edge region: the whole data set is used for the clustering process   | 35 |
| 3.19. Edge region: a part of the data set is used for the clustering process   | 35 |
| 3.20. Edge region: final clustering result   | 36 |
| 3.21. New approach   | 36 |

|  |      |
|--|------|
| 4.1. D1: trend surface with control points . . . . .   | 38   |
| 4.2. D1: illustration of the deformation in the point cloud . . . . .                              | 38   |
| 4.3. Clustering result: z-direction of epoch 3 - initial approach . . . . .                        | 39   |
| 4.4. Clustering result: z-direction of epoch 3 - updated approach . . . . .                        | 39   |
| 4.5. Clustering result: z-direction of epoch 3 - new approach . . . . .                            | 39   |
| 4.6. Calculated standard deviations: z-direction of epoch 3 - residuals to trend surface . . . . . | 40   |
| 4.7. Calculated standard deviations: z-direction of epoch 3 - initial approach . . . . .           | 40   |
| 4.8. Calculated standard deviations: z-direction of epoch 3 - updated approach . . . . .           | 40   |
| 4.9. Calculated standard deviations: z-direction of epoch 3 - new approach . . . . .               | 40   |
| 4.10. D1: y-direction discrepancies to nominal surface - initial approach . . . . .                | 41   |
| 4.11. D1: y-direction discrepancies to nominal surface - new approach . . . . .                    | 41   |
| 4.12. Cdf: x-direction . . . . .   | 42   |
| 4.13. Cdf: z-direction . . . . .   | 42   |
| 4.14. Cdf: y-direction . . . . .   | 42   |
| 4.15. D1: y-direction residuals of collocation - new approach . . . . .                            | 42   |
| 4.16. D2: clustering result - initial approach . . . . .   | 42   |
| 4.17. D2: clustering result - new approach . . . . .   | 42   |
| 4.18. D2: residuals to trend surface - x-direction epoch 3 . . . . .                               | 43   |
| 4.19. D2: z-direction residuals of collocation - initial approach . . . . .                        | 43   |
| 4.20. D2: z-direction residuals of collocation - new approach . . . . .                            | 43   |
| 4.21. D3: clustering result - initial approach . . . . .   | 44   |
| 4.22. D3: clustering result - new approach . . . . .   | 44   |
| 4.23. D3: calculated standard deviations - initial approach . . . . .                              | 44   |
| 4.24. D3: calculated standard deviations - new approach . . . . .                                  | 44   |
| 4.25. D3: residuals to trend surface - z-direction epoch 3 . . . . .                               | 44   |
| 4.26. D4: calculated standard deviations - initial approach . . . . .                              | 45   |
| 4.27. D4: calculated standard deviations - new approach . . . . .                                  | 45   |
| 4.28. D4: residuals to trend surface - z-direction epoch 3 . . . . .                               | 45   |
| 4.29. D4: z-direction residuals of collocation - new approach . . . . .                            | 45   |
|  |      |
| A.1. D1: x-direction clustering result epoch 2 . . . . .   | X    |
| A.2. D1: x-direction calculated standard deviations epoch 2 . . . . .                              | X    |
| A.3. D1: x-direction residuals to trend surface epoch 2 . . . . .                                  | X    |
| A.4. D1: y-direction clustering result epoch 2 . . . . .   | XI   |
| A.5. D1: y-direction calculated standard deviations epoch 2 . . . . .                              | XI   |
| A.6. D1: y-direction residuals to trend surface epoch 2 . . . . .                                  | XI   |
| A.7. D1: z-direction clustering result epoch 2 . . . . .   | XII  |
| A.8. D1: z-direction calculated standard deviations epoch 2 . . . . .                              | XII  |
| A.9. D1: z-direction residuals to trend surface epoch 2 . . . . .                                  | XII  |
| A.10.D1: x-direction clustering result epoch 3 . . . . .   | XIII |
| A.11.D1: x-direction calculated standard deviations epoch 3 . . . . .                              | XIII |
| A.12.D1: x-direction residuals to trend surface epoch 3 . . . . .                                  | XIII |
| A.13.D1: y-direction clustering result epoch 3 . . . . .   | XIV  |
| A.14.D1: y-direction calculated standard deviations epoch 3 . . . . .                              | XIV  |

|  |       |
|--|-------|
| A.15.D1: y-direction residuals to trend surface epoch 3        | XIV   |
| A.16.D1: z-direction clustering result epoch 3                 | XV    |
| A.17.D1: z-direction calculated standard deviations epoch 3    | XV    |
| A.18.D1: z-direction residuals to trend surface epoch 3        | XV    |
| A.19.D1: x-direction observations before and after collocation | XVI   |
| A.20.D1: x-direction residuals of collocation                  | XVI   |
| A.21.D1: x-direction discrepancies to nominal surface          | XVI   |
| A.22.D1: x-direction noise                                     | XVI   |
| A.23.D1: y-direction observations before and after collocation | XVI   |
| A.24.D1: y-direction residuals of collocation                  | XVI   |
| A.25.D1: y-direction discrepancies to nominal surface          | XVII  |
| A.26.D1: y-direction noise                                     | XVII  |
| A.27.D1: z-direction observations before and after collocation | XVII  |
| A.28.D1: z-direction residuals of collocation                  | XVII  |
| A.29.D1: z-direction discrepancies to nominal surface          | XVII  |
| A.30.D1: z-direction noise                                     | XVII  |
|  |       |
| B.1. D1: x-direction clustering result epoch 2                 | XVIII |
| B.2. D1: x-direction calculated standard deviations epoch 2    | XVIII |
| B.3. D1: x-direction residuals to trend surface epoch 2        | XVIII |
| B.4. D1: y-direction clustering result epoch 2                 | XIX   |
| B.5. D1: y-direction calculated standard deviations epoch 2    | XIX   |
| B.6. D1: y-direction residuals to trend surface epoch 2        | XIX   |
| B.7. D1: z-direction clustering result epoch 2                 | XX    |
| B.8. D1: z-direction calculated standard deviations epoch 2    | XX    |
| B.9. D1: z-direction residuals to trend surface epoch 2        | XX    |
| B.10.D1: x-direction clustering result epoch 3                 | XXI   |
| B.11.D1: x-direction calculated standard deviations epoch 3    | XXI   |
| B.12.D1: x-direction residuals to trend surface epoch 3        | XXI   |
| B.13.D1: y-direction clustering result epoch 3                 | XXII  |
| B.14.D1: y-direction calculated standard deviations epoch 3    | XXII  |
| B.15.D1: y-direction residuals to trend surface epoch 3        | XXII  |
| B.16.D1: z-direction clustering result epoch 3                 | XXIII |
| B.17.D1: z-direction calculated standard deviations epoch 3    | XXIII |
| B.18.D1: z-direction residuals to trend surface epoch 3        | XXIII |
| B.19.D1: x-direction observations before and after collocation | XXIV  |
| B.20.D1: x-direction residuals of collocation                  | XXIV  |
| B.21.D1: x-direction discrepancies to nominal surface          | XXIV  |
| B.22.D1: x-direction noise                                     | XXIV  |
| B.23.D1: y-direction observations before and after collocation | XXIV  |
| B.24.D1: y-direction residuals of collocation                  | XXIV  |
| B.25.D1: y-direction discrepancies to nominal surface          | XXV   |
| B.26.D1: y-direction noise                                     | XXV   |
| B.27.D1: z-direction observations before and after collocation | XXV   |

|   |        |
|---|--------|
| <b>B.28.D1: z-direction residuals of collocation</b> . . . . .                              | XXV    |
| <b>B.29.D1: z-direction discrepancies to nominal surface</b> . . . . .                      | XXV    |
| <b>B.30.D1: z-direction noise</b> . . . . .   | XXV    |
| <b>C.1. D1: x-direction pre-selection of points to be clustered areas epoch 2</b> . . . . . | XXVI   |
| <b>C.2. D1: x-direction clustering result epoch 2 unfiltered</b> . . . . .                  | XXVI   |
| <b>C.3. D1: x-direction clustering result epoch 2 filtered</b> . . . . .                    | XXVI   |
| <b>C.4. D1: x-direction posterior probability epoch 2 cluster 1 unfiltered</b> . . . . .    | XXVII  |
| <b>C.5. D1: x-direction posterior probability epoch 2 cluster 2 unfiltered</b> . . . . .    | XXVII  |
| <b>C.6. D1: x-direction posterior probability epoch 2 cluster 3 unfiltered</b> . . . . .    | XXVII  |
| <b>C.7. D1: x-direction posterior probability epoch 2 cluster 4 unfiltered</b> . . . . .    | XXVII  |
| <b>C.8. D1: x-direction calculated standard deviations epoch 2</b> . . . . .                | XXVII  |
| <b>C.9. D1: x-direction residuals to trend surface epoch 2</b> . . . . .                    | XXVII  |
| <b>C.10.D1: y-direction pre-selection of points to be clustered epoch 2</b> . . . . .       | XXVIII |
| <b>C.11.D1: y-direction clustering result epoch 2 unfiltered</b> . . . . .                  | XXVIII |
| <b>C.12.D1: y-direction clustering result epoch 2 filtered</b> . . . . .                    | XXVIII |
| <b>C.13.D1: y-direction posterior probability epoch 2 cluster 1 unfiltered</b> . . . . .    | XXVIII |
| <b>C.14.D1: y-direction posterior probability epoch 2 cluster 2 unfiltered</b> . . . . .    | XXVIII |
| <b>C.15.D1: y-direction posterior probability epoch 2 cluster 3 unfiltered</b> . . . . .    | XXIX   |
| <b>C.16.D1: y-direction posterior probability epoch 2 cluster 4 unfiltered</b> . . . . .    | XXIX   |
| <b>C.17.D1: y-direction calculated standard deviations epoch 2</b> . . . . .                | XXIX   |
| <b>C.18.D1: y-direction residuals to trend surface epoch 2</b> . . . . .                    | XXIX   |
| <b>C.19.D1: z-direction pre-selection of points to be clustered epoch 2</b> . . . . .       | XXX    |
| <b>C.20.D1: z-direction clustering result epoch 2 unfiltered</b> . . . . .                  | XXX    |
| <b>C.21.D1: z-direction clustering result epoch 2 filtered</b> . . . . .                    | XXX    |
| <b>C.22.D1: z-direction posterior probability epoch 2 cluster 1 unfiltered</b> . . . . .    | XXX    |
| <b>C.23.D1: z-direction posterior probability epoch 2 cluster 2 unfiltered</b> . . . . .    | XXX    |
| <b>C.24.D1: z-direction posterior probability epoch 2 cluster 3 unfiltered</b> . . . . .    | XXXI   |
| <b>C.25.D1: z-direction posterior probability epoch 2 cluster 4 unfiltered</b> . . . . .    | XXXI   |
| <b>C.26.D1: z-direction posterior probability epoch 2 cluster 5 unfiltered</b> . . . . .    | XXXI   |
| <b>C.27.D1: z-direction calculated standard deviations epoch 2</b> . . . . .                | XXXI   |
| <b>C.28.D1: z-direction residuals to trend surface epoch 2</b> . . . . .                    | XXXI   |
| <b>C.29.D1: x-direction pre-selection of points to be clustered epoch 3</b> . . . . .       | XXXII  |
| <b>C.30.D1: x-direction clustering result epoch 3 unfiltered</b> . . . . .                  | XXXII  |
| <b>C.31.D1: x-direction clustering result epoch 3 filtered</b> . . . . .                    | XXXII  |
| <b>C.32.D1: x-direction posterior probability epoch 3 cluster 1 unfiltered</b> . . . . .    | XXXII  |
| <b>C.33.D1: x-direction posterior probability epoch 3 cluster 2 unfiltered</b> . . . . .    | XXXII  |
| <b>C.34.D1: x-direction posterior probability epoch 3 cluster 3 unfiltered</b> . . . . .    | XXXIII |
| <b>C.35.D1: x-direction posterior probability epoch 3 cluster 4 unfiltered</b> . . . . .    | XXXIII |
| <b>C.36.D1: x-direction calculated standard deviations epoch 3</b> . . . . .                | XXXIII |
| <b>C.37.D1: x-direction residuals to trend surface epoch 3</b> . . . . .                    | XXXIII |
| <b>C.38.D1: y-direction pre-selection of points to be clustered epoch 3</b> . . . . .       | XXXIV  |
| <b>C.39.D1: y-direction clustering result epoch 3 unfiltered</b> . . . . .                  | XXXIV  |
| <b>C.40.D1: y-direction clustering result epoch 3 filtered</b> . . . . .                    | XXXIV  |

|   |         |
|---|---------|
| C.41.D1: y-direction posterior probability epoch 3 cluster 1 unfiltered | XXXIV   |
| C.42.D1: y-direction posterior probability epoch 3 cluster 2 unfiltered | XXXIV   |
| C.43.D1: y-direction posterior probability epoch 3 cluster 3 unfiltered | XXXV    |
| C.44.D1: y-direction posterior probability epoch 3 cluster 4 unfiltered | XXXV    |
| C.45.D1: y-direction calculated standard deviations epoch 3             | XXXV    |
| C.46.D1: y-direction residuals to trend surface epoch 3                 | XXXV    |
| C.47.D1: z-direction pre-selection of points to be clustered epoch 3    | XXXVI   |
| C.48.D1: z-direction clustering result epoch 3 unfiltered               | XXXVI   |
| C.49.D1: z-direction clustering result epoch 3 filtered                 | XXXVI   |
| C.50.D1: z-direction posterior probability epoch 3 cluster 1 unfiltered | XXXVI   |
| C.51.D1: z-direction posterior probability epoch 3 cluster 2 unfiltered | XXXVI   |
| C.52.D1: z-direction posterior probability epoch 3 cluster 3 unfiltered | XXXVII  |
| C.53.D1: z-direction posterior probability epoch 3 cluster 4 unfiltered | XXXVII  |
| C.54.D1: z-direction posterior probability epoch 3 cluster 5 unfiltered | XXXVII  |
| C.55.D1: z-direction calculated standard deviations epoch 3             | XXXVII  |
| C.56.D1: z-direction residuals to trend surface epoch 3                 | XXXVII  |
| C.57.D1: x-direction observations before and after collocation          | XXXVIII |
| C.58.D1: x-direction residuals of collocation                           | XXXVIII |
| C.59.D1: x-direction discrepancies to nominal surface                   | XXXVIII |
| C.60.D1: x-direction noise  | XXXVIII |
| C.61.D1: y-direction observations before and after collocation          | XXXVIII |
| C.62.D1: y-direction residuals of collocation                           | XXXVIII |
| C.63.D1: y-direction discrepancies to nominal surface                   | XXXIX   |
| C.64.D1: y-direction noise  | XXXIX   |
| C.65.D1: z-direction observations before and after collocation          | XXXIX   |
| C.66.D1: z-direction residuals of collocation                           | XXXIX   |
| C.67.D1: z-direction discrepancies to nominal surface                   | XXXIX   |
| C.68.D1: z-direction noise  | XXXIX   |

# List of Tables

|  |    |
|--|----|
| <b>3.1. Overview of the first exclusion</b> . . . . .  | 21 |
| <b>3.2. Overview of results part 1</b> . . . . .   | 22 |
| <b>3.3. Overview of results part 2</b> . . . . .   | 23 |
| <b>3.4. Overview of results part 3</b> . . . . .   | 23 |
| <b>3.5. Overview of the second exclusion</b> . . . . .   | 24 |
| <b>3.6. Overview of the final exclusion</b> . . . . .  | 25 |
| <b>3.7. Overview of results with combined clustering methods</b> . . . . .                         | 28 |
| <b>4.1. Overview of the shift each epoch of <math>P_{4,5}</math> in the z-coordinate</b> . . . . . | 38 |

# Bibliography

- [Aggarwal and Reddy, 2018] Aggarwal, C. C. and Reddy, C. K. (2018). *Data Clustering Algorithms and Applications*. Taylor & Francis Group.
- [Amjad and Shah, 2020] Amjad, F. and Shah, L. A. (2020). Identification and assessment of sites for solar farms development using GIS and density based clustering technique- a case of pakistan. *Renewable Energy*, 155:761–769.
- [Asgharbeygi and Maleki, 2008] Asgharbeygi, N. and Maleki, A. (2008). Geodesic k-means clustering. In *2008 19th International Conference on Pattern Recognition*. IEEE.
- [Azran and Ghahramani, ] Azran, A. and Ghahramani, Z. Spectral methods for automatic multiscale data clustering. In *2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition - Volume 1 (CVPR'06)*. IEEE.
- [Everitt, 2011] Everitt, B. (2011). *Cluster Analysis*. Wiley.
- [FISHER, 1936] FISHER, R. A. (1936). THE USE OF MULTIPLE MEASUREMENTS IN TAXONOMIC PROBLEMS. *Annals of Eugenics*, 7(2):179–188.
- [Granat et al., 2021] Granat, R., Donnellan, A., Heflin, M., Lyzenga, G., Glasscoe, M., Parker, J., Pierce, M., Wang, J., Rundle, J., and Ludwig, L. G. (2021). Clustering analysis methods for GNSS observations: A data-driven approach to identifying california's major faults. *Earth and Space Science*, 8(11).
- [Häme et al., 2020] Häme, T., Sirro, L., Kilpi, J., Seitsonen, L., Andersson, K., and Melkas, T. (2020). A hierarchical clustering method for land cover change detection and identification. *Remote Sensing*, 12(11):1751.
- [Han, 2011] Han, J. (2011). *Data mining concepts and techniques*. Elsevier.
- [Harmening and Neuner, 2020] Harmening, C. and Neuner, H. (2020). A spatio-temporal deformation model for laser scanning point clouds. *Journal of Geodesy*, 94(2).
- [Higginson et al., 2015] Higginson, S., Thompson, K. R., Woodworth, P. L., and Hughes, C. W. (2015). The tilt of mean sea level along the east coast of north america. *Geophysical Research Letters*, 42(5):1471–1479.
- [Jain et al., 1999] Jain, A. K., Murty, M. N., and Flynn, P. J. (1999). Data clustering. *ACM Computing Surveys*, 31(3):264–323.
- [Li et al., 2022] Li, Y., Yan, H., Yang, W., Wang, S., and Su, X. (2022). OPTIMAL LINEAR COMBINATIONS SELECTION BASED ON FUZZY CLUSTERING ANALYSIS FOR MULTI-FREQUENCY GNSS. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XLVI-3/W1-2022:97–103.

- [Lv et al., 2010] Lv, Z., Hu, Y., Zhong, H., Wu, J., Li, B., and Zhao, H. (2010). Parallel k-means clustering of remote sensing images based on MapReduce. In *Web Information Systems and Mining*, pages 162–170. Springer Berlin Heidelberg.
- [Morsdorf et al., 2003] Morsdorf, F., Meier, E., Allgöwer, B., and Uesch, D. (2003). Clustering in airborne laser scanning raw data for segmentation of single trees. 34.
- [Nagy, 1968] Nagy, G. (1968). State of the art in pattern recognition. *Proceedings of the IEEE*, 56(5):836–863.
- [Ng et al., 2001] Ng, A., Jordan, M., and Weiss, Y. (2001). On spectral clustering: Analysis and an algorithm. In Dietterich, T., Becker, S., and Ghahramani, Z., editors, *Advances in Neural Information Processing Systems*, volume 14. MIT Press.
- [Pokonieczny et al., 2017] Pokonieczny, K., Bielecka, E., and Kaminski, P. (2017). Analysis of geodetic control points density depending on the land cover and relief – the opoczno district case study. In *Proceedings of 10th International Conference "Environmental Engineering"*. VGTU Technika.
- [Praene et al., 2019] Praene, J. P., Malet-Damour, B., Radanielina, M. H., Fontaine, L., and Rivière, G. (2019). GIS-based approach to identify climatic zoning: A hierarchical clustering on principal component analysis. *Building and Environment*, 164:106330.
- [Schubert et al., 2017] Schubert, E., Sander, J., Ester, M., Kriegel, H. P., and Xu, X. (2017). DBSCAN revisited, revisited. *ACM Transactions on Database Systems*, 42(3):1–21.
- [Shi and Malik, 2000] Shi, J. and Malik, J. (2000). Normalized cuts and image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(8):888–905.
- [Singh et al., 2021] Singh, S. K., Raval, S., and Banerjee, B. P. (2021). Automated structural discontinuity mapping in a rock face occluded by vegetation using mobile laser scanning. *Engineering Geology*, 285:106040.
- [The MathWorks, Inc., 2022a] The MathWorks, Inc. (1994-2022a). cluster - matlab documentation - mathworks deutschland.
- [The MathWorks, Inc., 2022b] The MathWorks, Inc. (1994-2022b). clusterdata - matlab documentation - mathworks deutschland.
- [The MathWorks, Inc., 2022c] The MathWorks, Inc. (1994-2022c). dbscan - matlab documentation - mathworks deutschland.
- [The MathWorks, Inc., 2022d] The MathWorks, Inc. (1994-2022d). evalclusters - matlab documentation - mathworks deutschland.
- [The MathWorks, Inc., 2022e] The MathWorks, Inc. (1994-2022e). fcm - matlab documentation - mathworks deutschland.
- [The MathWorks, Inc., 2022f] The MathWorks, Inc. (1994-2022f). fitgmdist - matlab documentation - mathworks deutschland.

- [The MathWorks, Inc., 2022g] The MathWorks, Inc. (1994-2022g). kmeans - matlab documentation - mathworks deutschland.
- [The MathWorks, Inc., 2022h] The MathWorks, Inc. (1994-2022h). kmedoids - matlab documentation - mathworks deutschland.
- [The MathWorks, Inc., 2022i] The MathWorks, Inc. (1994-2022i). Matlab documentation - mathworks deutschland.
- [The MathWorks, Inc., 2022j] The MathWorks, Inc. (1994-2022j). spectralcluster - matlab documentation - mathworks deutschland.
- [The MathWorks, Inc., 2022k] The MathWorks, Inc. (1994-2022k). subclust - matlab documentation - mathworks deutschland.
- [von Luxburg, 2007] von Luxburg, U. (2007). A tutorial on spectral clustering. *Statistics and Computing*, 17(4):395–416.
- [Wunderlich et al., 2016] Wunderlich, T., Niemeier, W., Wujanz, D., Holst, C., Neitzel, F., and Kuhlmann, H. (2016). Areal deformation analysis from tls point clouds – the challenge. *AVN Allgemeine Vermessungs-Nachrichten*, 123:340–351.
- [Xu and WunschII, 2005] Xu, R. and WunschII, D. (2005). Survey of clustering algorithms. *IEEE Transactions on Neural Networks*, 16(3):645–678.
- [Xu et al., 2016] Xu, Y., Tuttas, S., Heogner, L., and Stilla, U. (2016). CLASSIFICATION OF PHOTOGRAMMETRIC POINT CLOUDS OF SCAFFOLDS FOR CONSTRUCTION SITE MONITORING USING SUBSPACE CLUSTERING AND PCA. *ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XLI-B3:725–732.

# A. Results D1 (initial approach)

Appendix A lists the results of the clustering process, the calculated standard deviations and the final results of the deformation analysis approach using the initial approach. They are sorted by coordinate direction and epoch.

## A.1. x-direction epoch 2

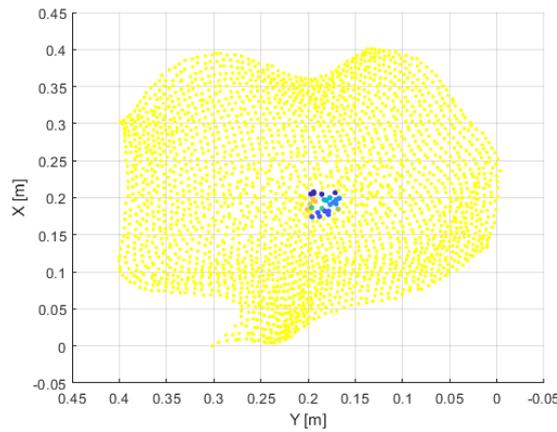


Figure A.1.: D1: x-direction clustering result epoch 2

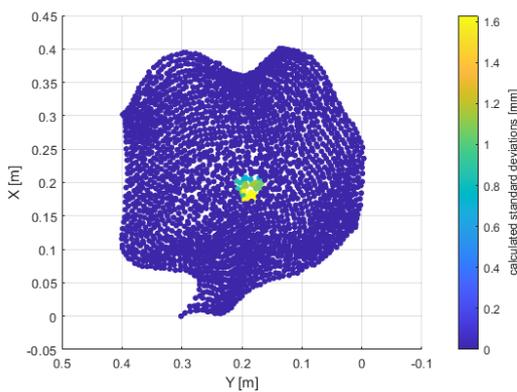


Figure A.2.: D1: x-direction calculated standard deviations epoch 2

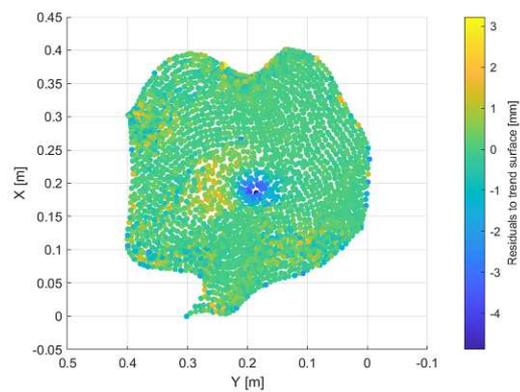


Figure A.3.: D1: x-direction residuals to trend surface epoch 2

## A.2. y-direction epoch 2

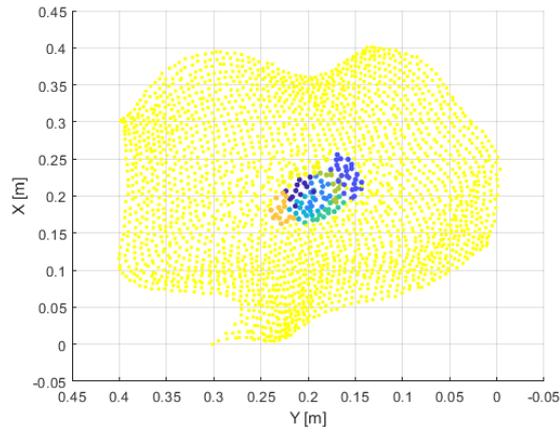


Figure A.4.: D1: y-direction clustering result epoch 2

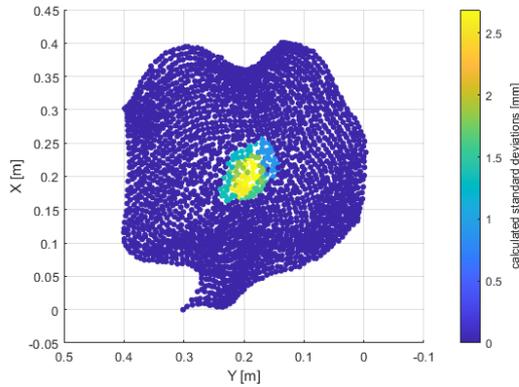


Figure A.5.: D1: y-direction calculated standard deviations epoch 2

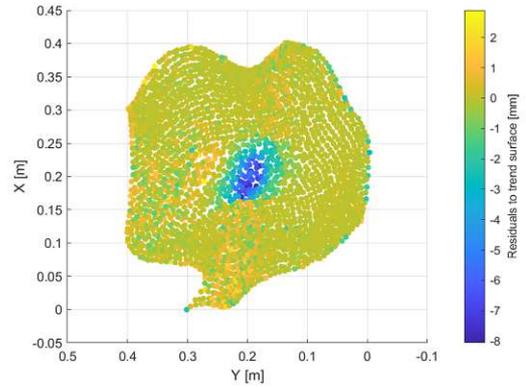


Figure A.6.: D1: y-direction residuals to trend surface epoch 2

### A.3. z-direction epoch 2

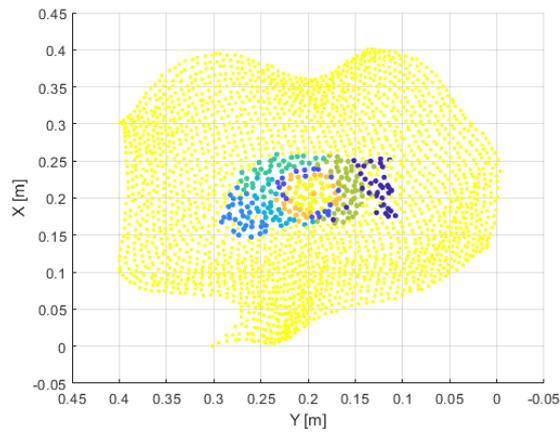


Figure A.7.: D1: z-direction clustering result epoch 2

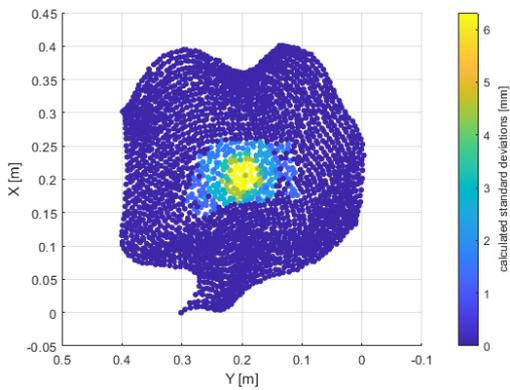


Figure A.8.: D1: z-direction calculated standard deviations epoch 2

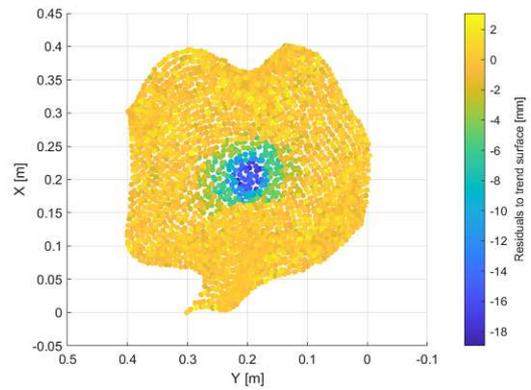


Figure A.9.: D1: z-direction residuals to trend surface epoch 2

### A.4. x-direction epoch 3

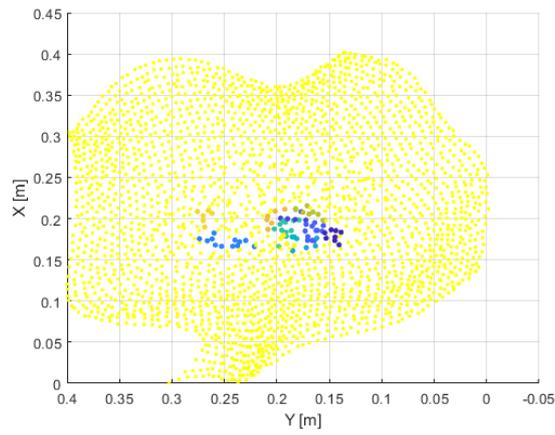


Figure A.10.: D1: x-direction clustering result epoch 3

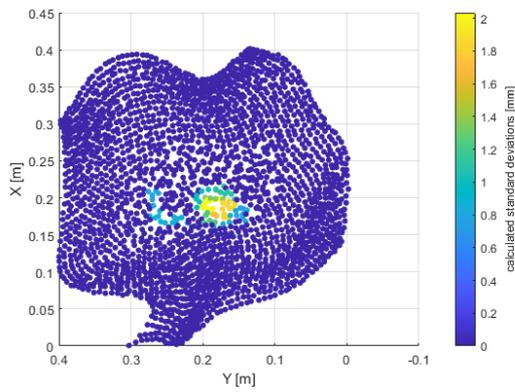


Figure A.11.: D1: x-direction calculated standard deviations epoch 3

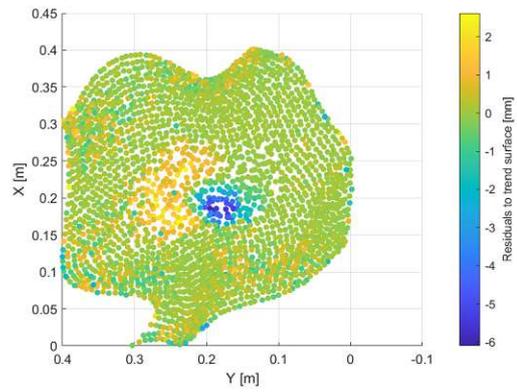


Figure A.12.: D1: x-direction residuals to trend surface epoch 3

### A.5. y-direction epoch 3

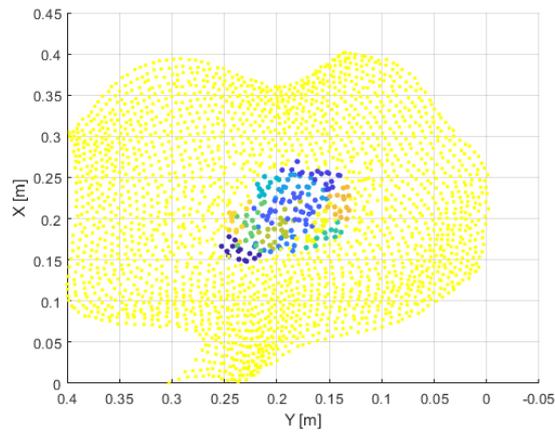


Figure A.13.: D1: y-direction clustering result epoch 3

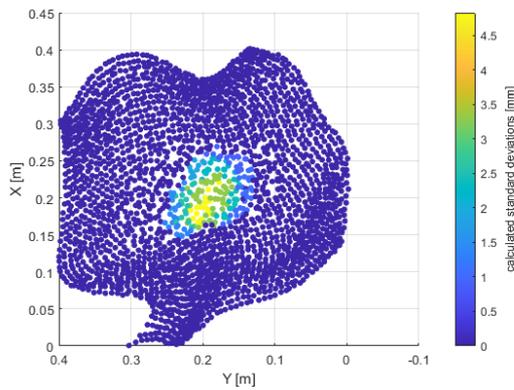


Figure A.14.: D1: y-direction calculated standard deviations epoch 3

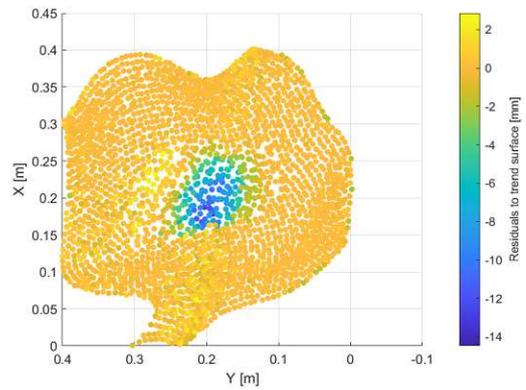


Figure A.15.: D1: y-direction residuals to trend surface epoch 3

### A.6. z-direction epoch 3

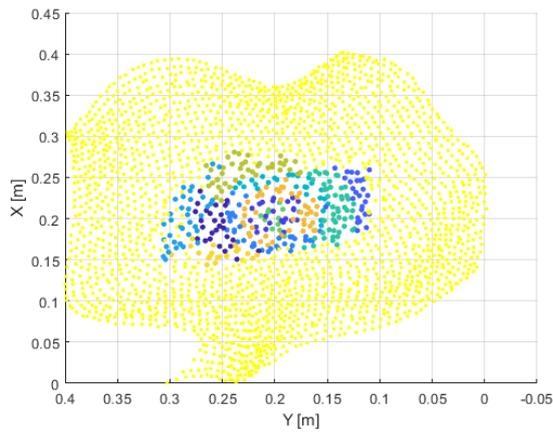


Figure A.16.: D1: z-direction clustering result epoch 3

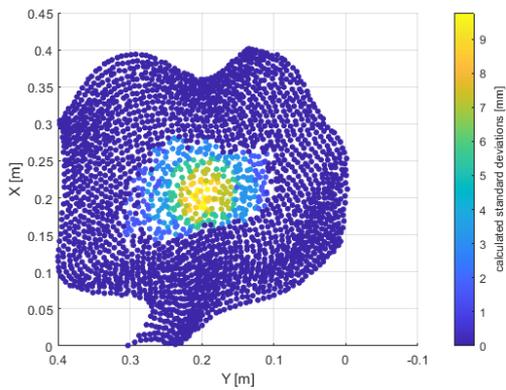


Figure A.17.: D1: z-direction calculated standard deviations epoch 3

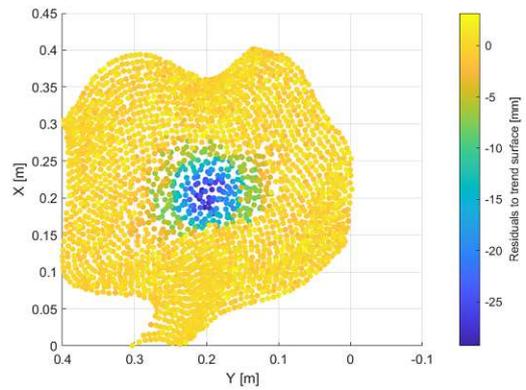


Figure A.18.: D1: z-direction residuals to trend surface epoch 3

### A.7. final results

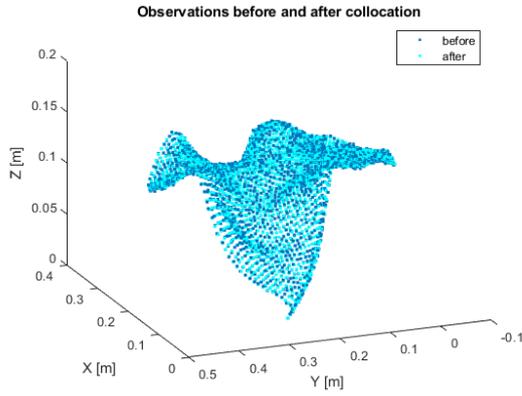


Figure A.19.: D1: x-direction observations before and after collocation

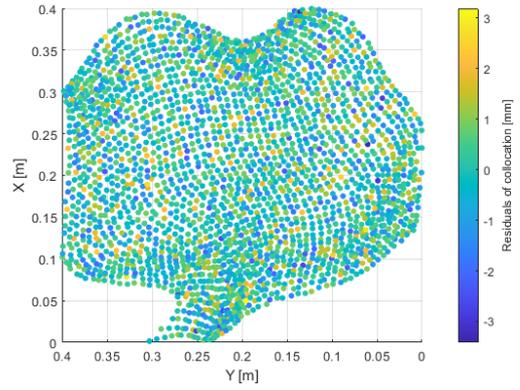


Figure A.20.: D1: x-direction residuals of collocation

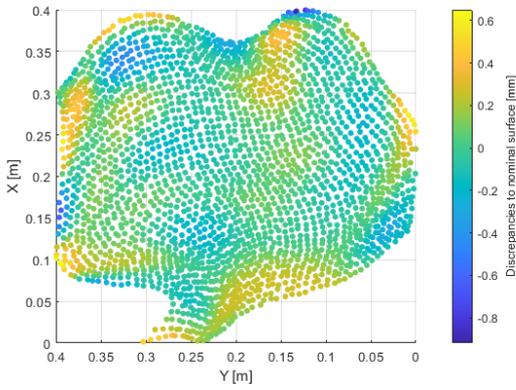


Figure A.21.: D1: x-direction discrepancies to nominal surface

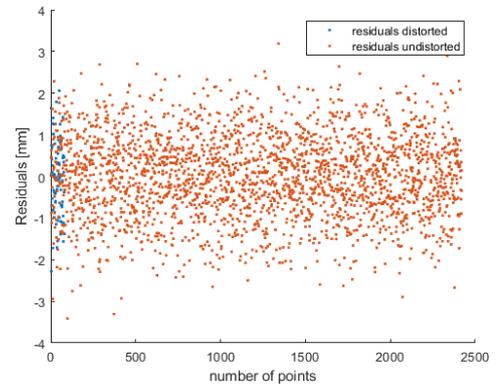


Figure A.22.: D1: x-direction noise

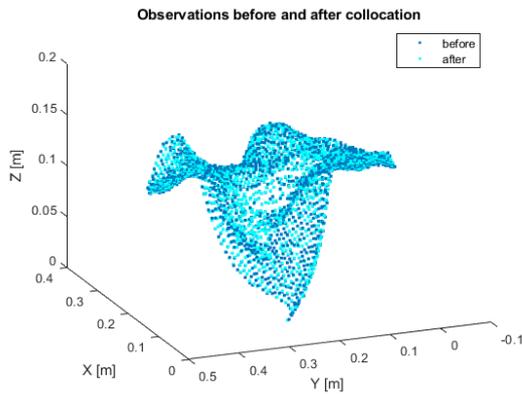


Figure A.23.: D1: y-direction observations before and after collocation

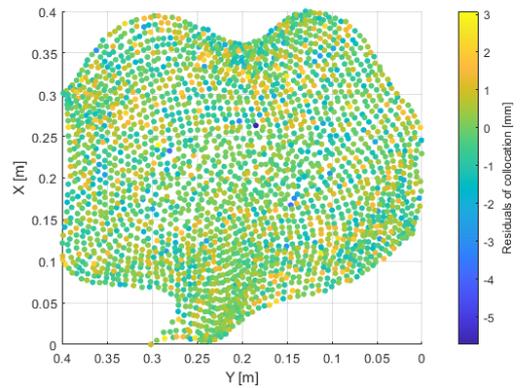


Figure A.24.: D1: y-direction residuals of collocation

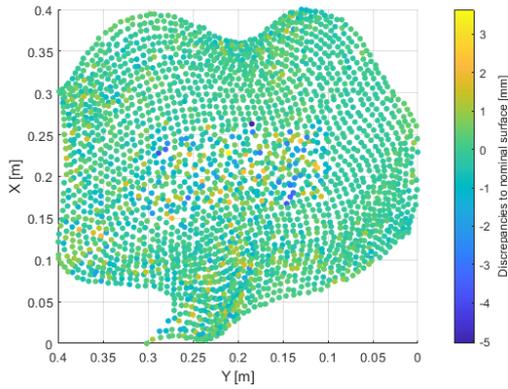


Figure A.25.: D1: y-direction discrepancies to nominal surface

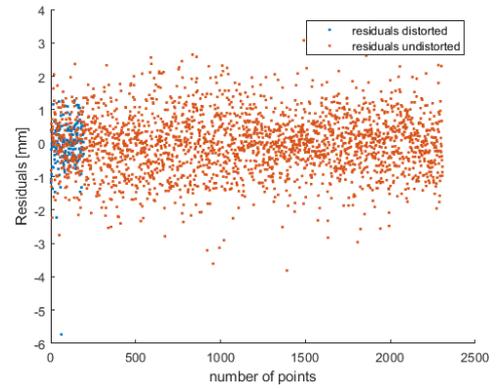


Figure A.26.: D1: y-direction noise

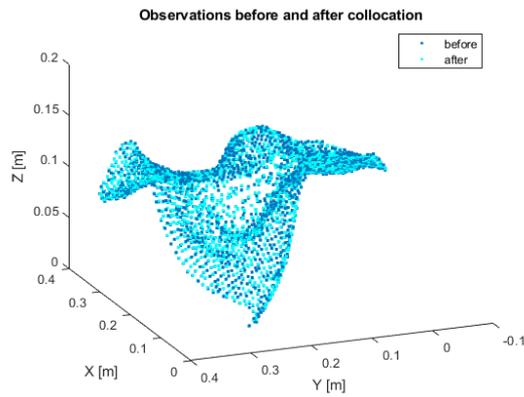


Figure A.27.: D1: z-direction observations before and after collocation

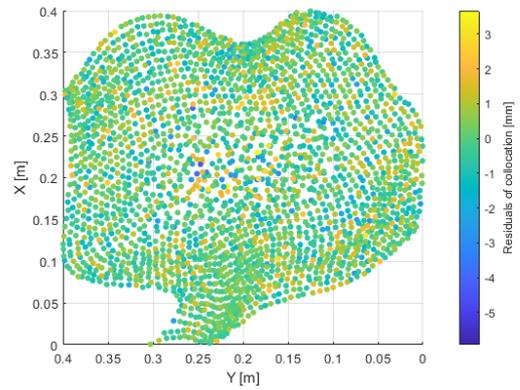


Figure A.28.: D1: z-direction residuals of collocation

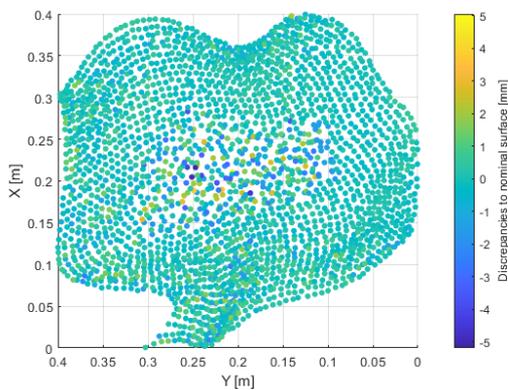


Figure A.29.: D1: z-direction discrepancies to nominal surface

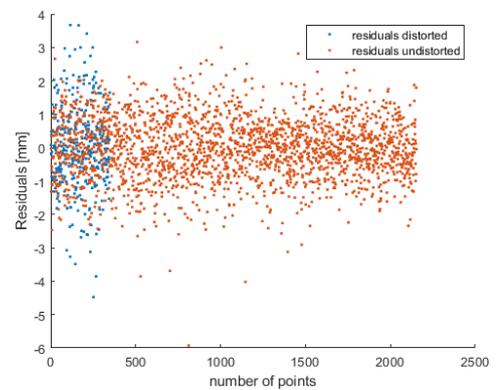


Figure A.30.: D1: z-direction noise

## B. Results D1 (updated approach)

Appendix B lists the results of the clustering process, the calculated standard deviations and the final results of the deformation analysis approach using the updated approach. They are sorted by coordinate direction and epoch.

### B.1. x-direction epoch 2

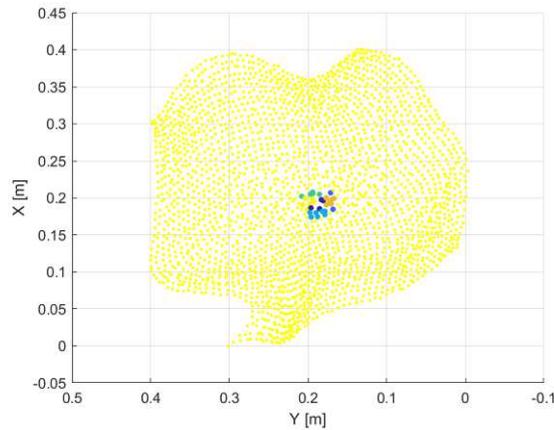


Figure B.1.: D1: x-direction clustering result epoch 2

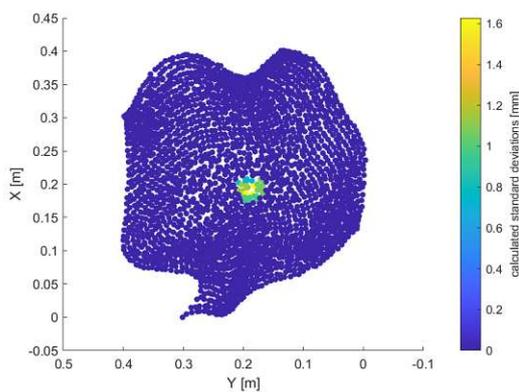


Figure B.2.: D1: x-direction calculated standard deviations epoch 2

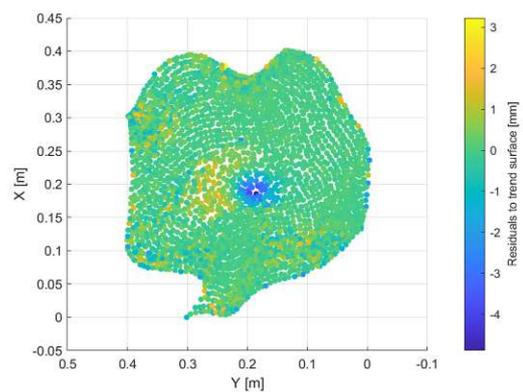


Figure B.3.: D1: x-direction residuals to trend surface epoch 2

## B.2. y-direction epoch 2

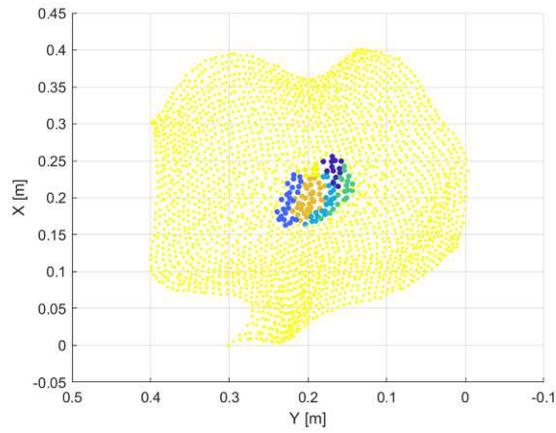


Figure B.4.: D1: y-direction clustering result epoch 2

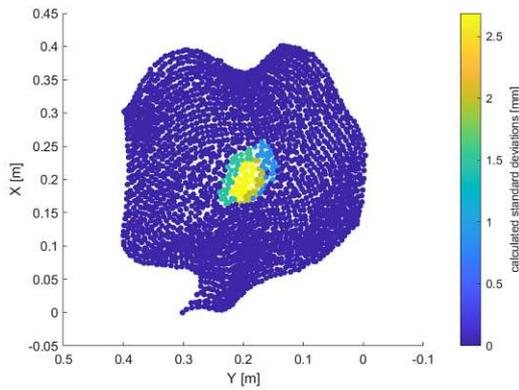


Figure B.5.: D1: y-direction calculated standard deviations epoch 2

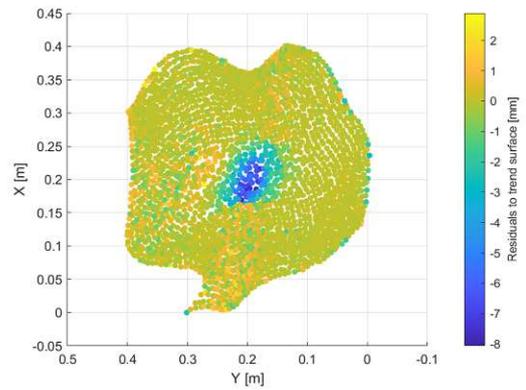


Figure B.6.: D1: y-direction residuals to trend surface epoch 2

### B.3. z-direction epoch 2

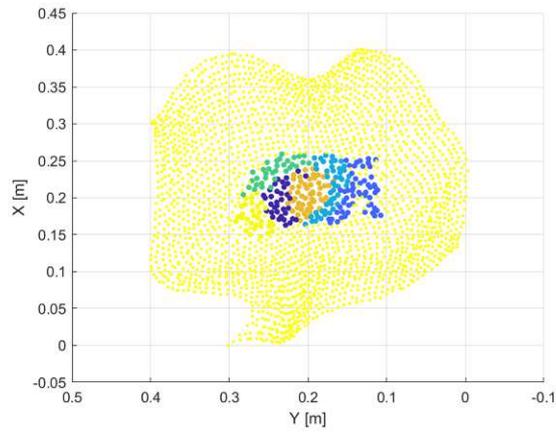


Figure B.7.: D1: z-direction clustering result epoch 2

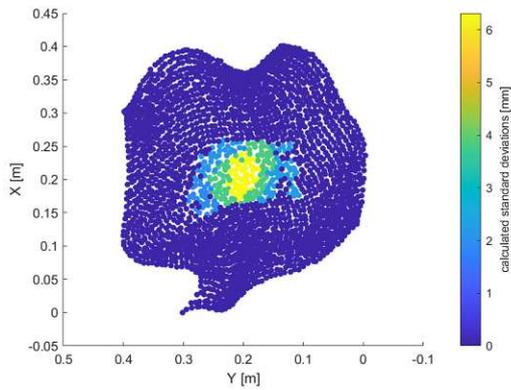


Figure B.8.: D1: z-direction calculated standard deviations epoch 2

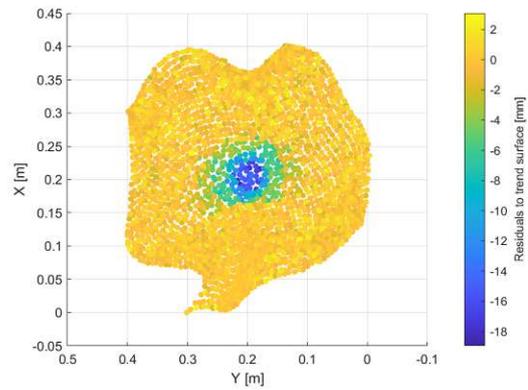


Figure B.9.: D1: z-direction residuals to trend surface epoch 2

### B.4. x-direction epoch 3

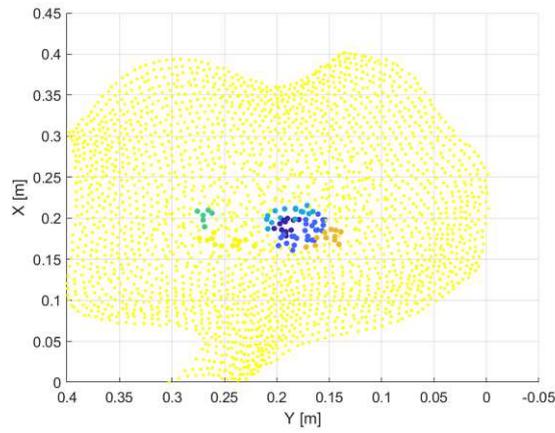


Figure B.10.: D1: x-direction clustering result epoch 3

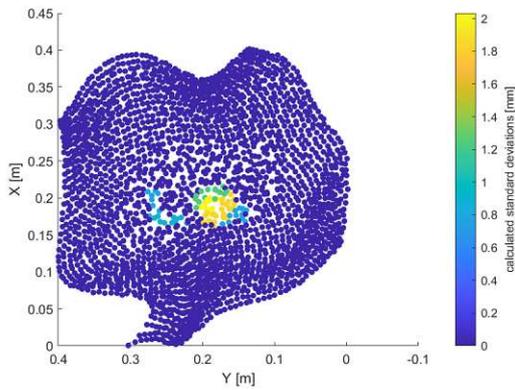


Figure B.11.: D1: x-direction calculated standard deviations epoch 3

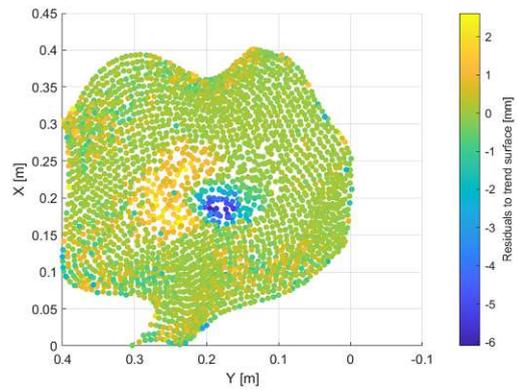


Figure B.12.: D1: x-direction residuals to trend surface epoch 3

### B.5. y-direction epoch 3

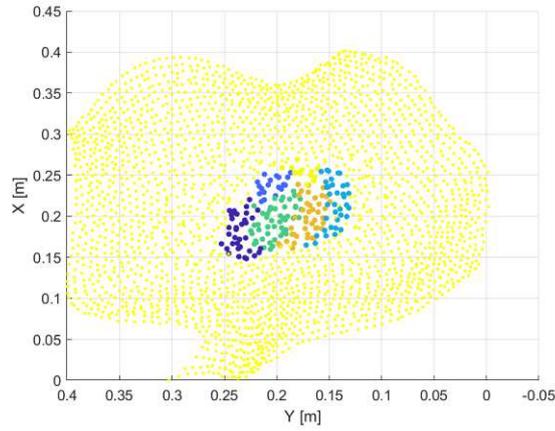


Figure B.13.: D1: y-direction clustering result epoch 3

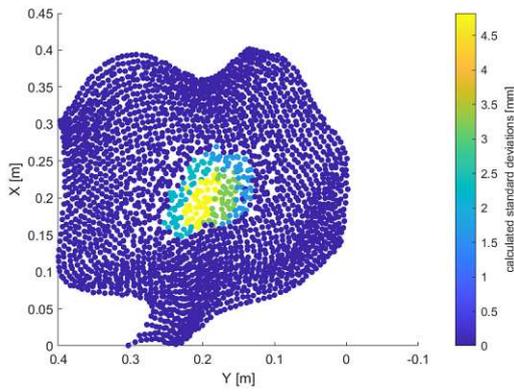


Figure B.14.: D1: y-direction calculated standard deviations epoch 3

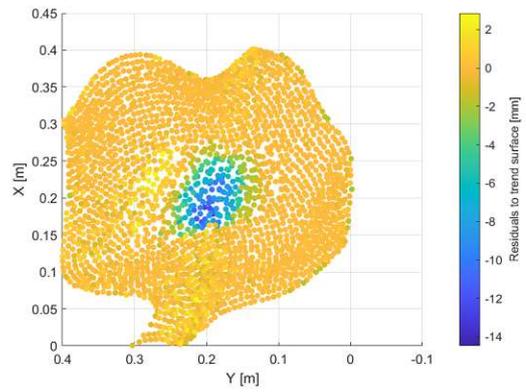


Figure B.15.: D1: y-direction residuals to trend surface epoch 3

### B.6. z-direction epoch 3

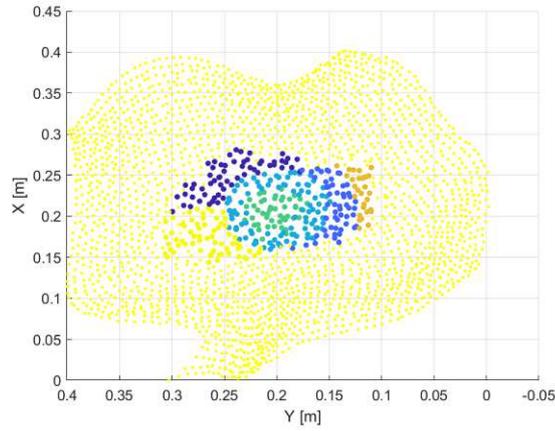


Figure B.16.: D1: z-direction clustering result epoch 3

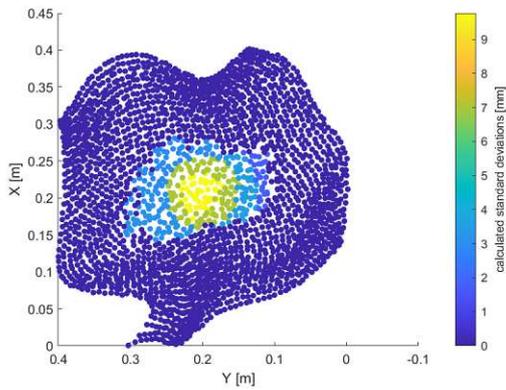


Figure B.17.: D1: z-direction calculated standard deviations epoch 3

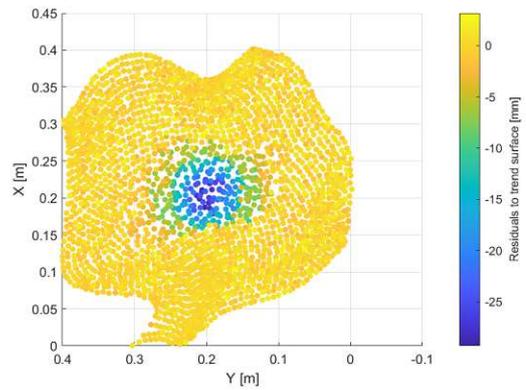


Figure B.18.: D1: z-direction residuals to trend surface epoch 3

## B.7. final results

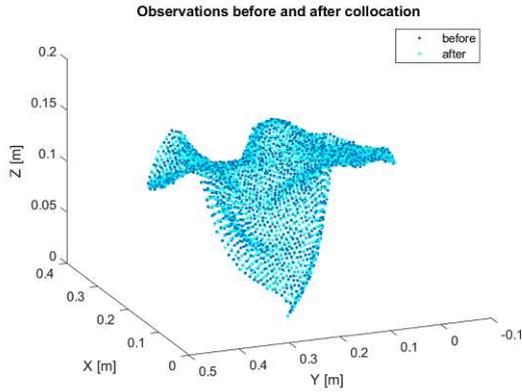


Figure B.19.: D1: x-direction observations before and after collocation

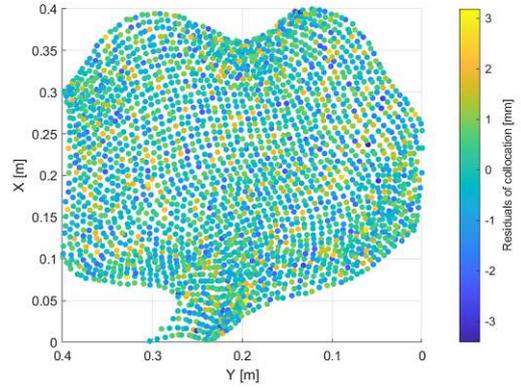


Figure B.20.: D1: x-direction residuals of collocation

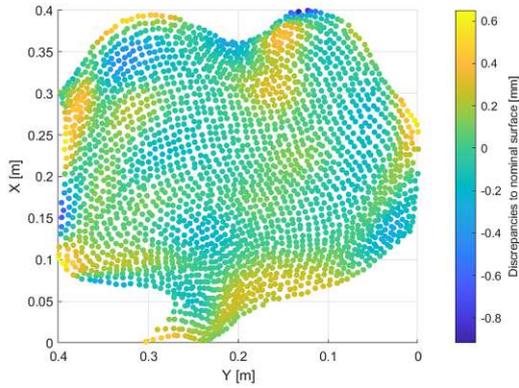


Figure B.21.: D1: x-direction discrepancies to nominal surface

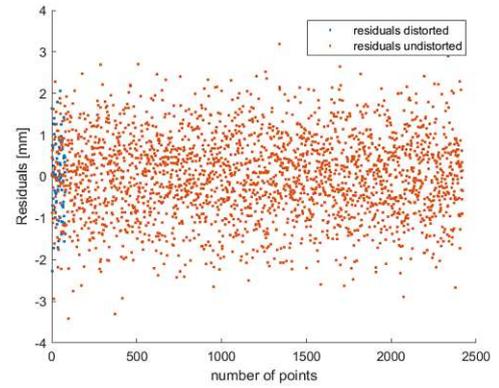


Figure B.22.: D1: x-direction noise

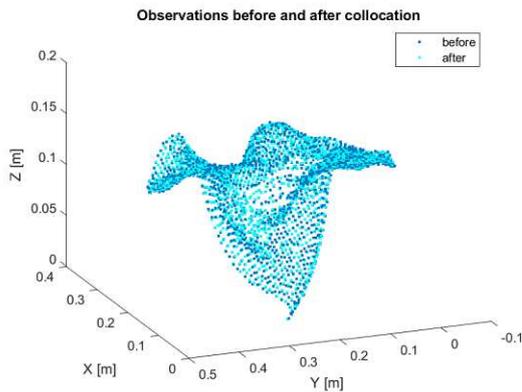


Figure B.23.: D1: y-direction observations before and after collocation

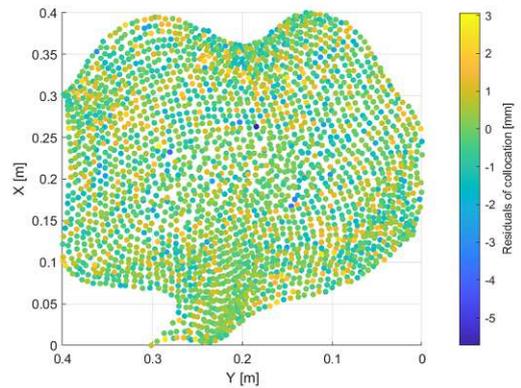


Figure B.24.: D1: y-direction residuals of collocation

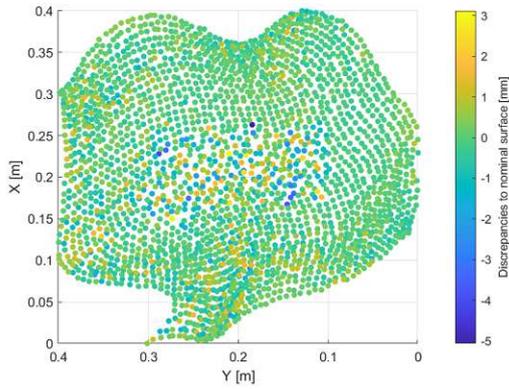


Figure B.25.: D1: y-direction discrepancies to nominal surface

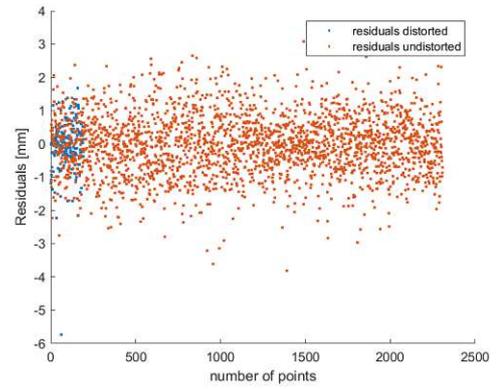


Figure B.26.: D1: y-direction noise

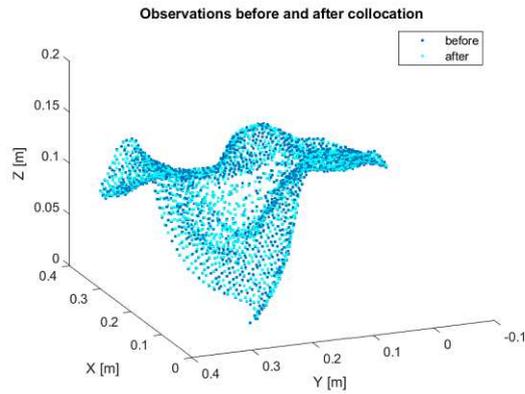


Figure B.27.: D1: z-direction observations before and after collocation

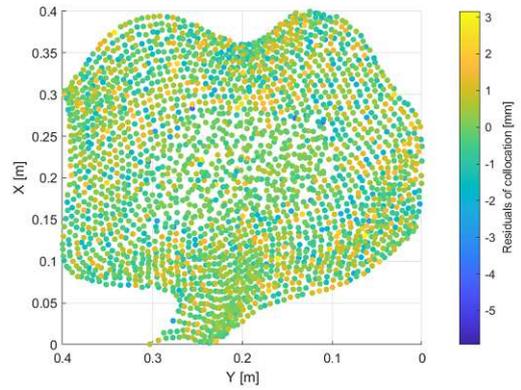


Figure B.28.: D1: z-direction residuals of collocation

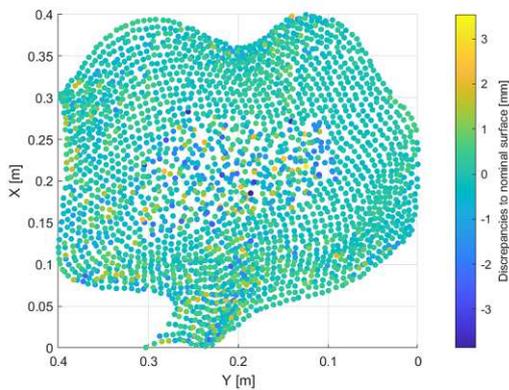


Figure B.29.: D1: z-direction discrepancies to nominal surface

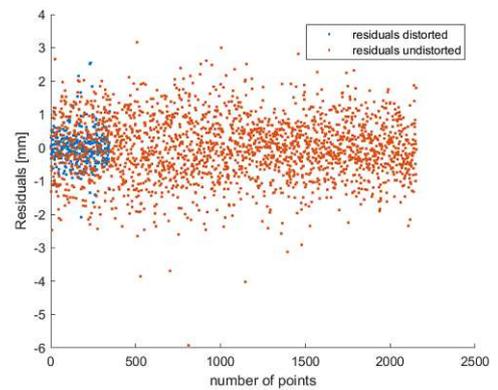


Figure B.30.: D1: z-direction noise

## C. Results D1 (new approach)

Appendix C lists the results of the clustering process, the calculated standard deviations and the final results of the deformation analysis approach using the new approach. They are sorted by coordinate direction and epoch.

### C.1. x-direction epoch 2

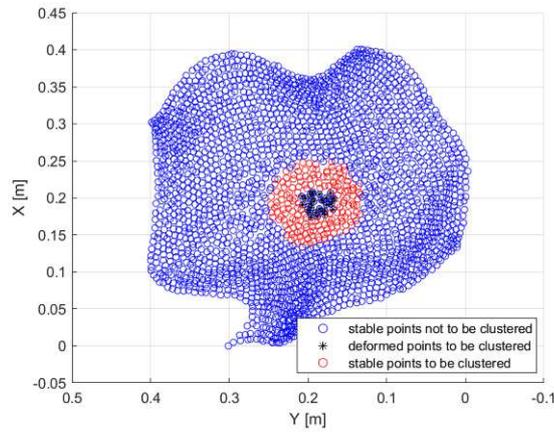


Figure C.1.: D1: x-direction pre-selection of points to be clustered areas epoch 2

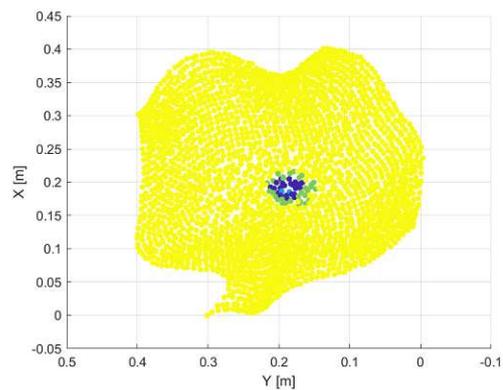
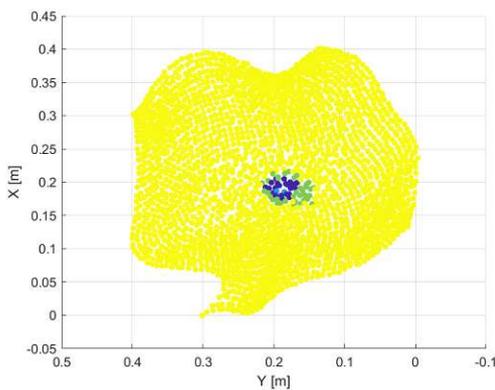


Figure C.2.: D1: x-direction clustering result epoch 2 unfiltered  
Figure C.3.: D1: x-direction clustering result epoch 2 filtered

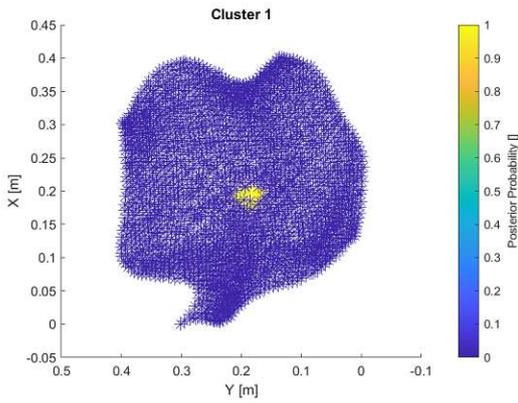


Figure C.4.: D1: x-direction posterior probability epoch 2 cluster 1 unfiltered

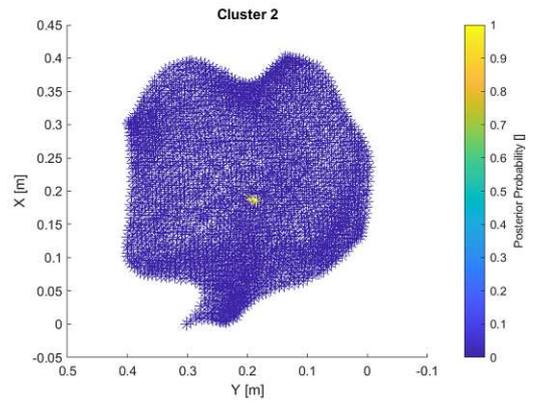


Figure C.5.: D1: x-direction posterior probability epoch 2 cluster 2 unfiltered

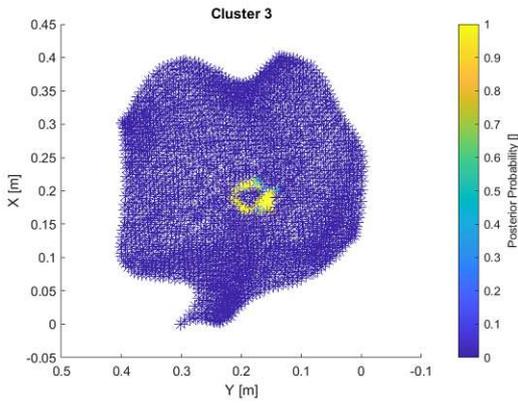


Figure C.6.: D1: x-direction posterior probability epoch 2 cluster 3 unfiltered

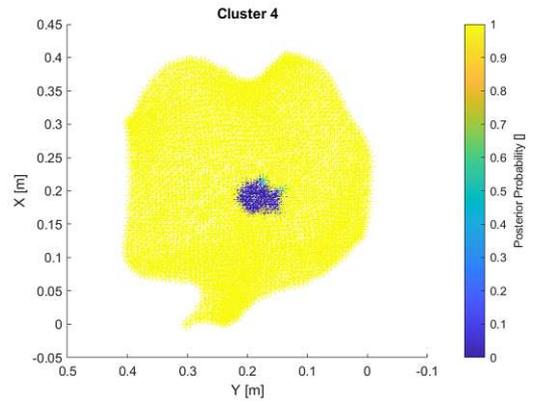


Figure C.7.: D1: x-direction posterior probability epoch 2 cluster 4 unfiltered

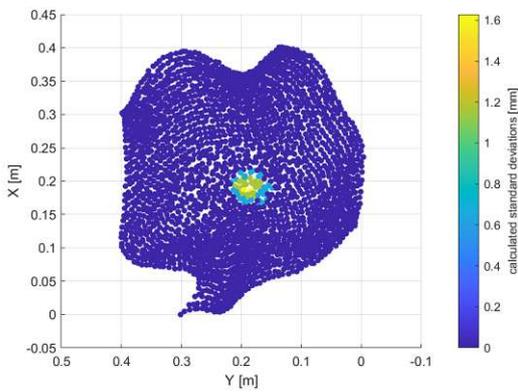


Figure C.8.: D1: x-direction calculated standard deviations epoch 2

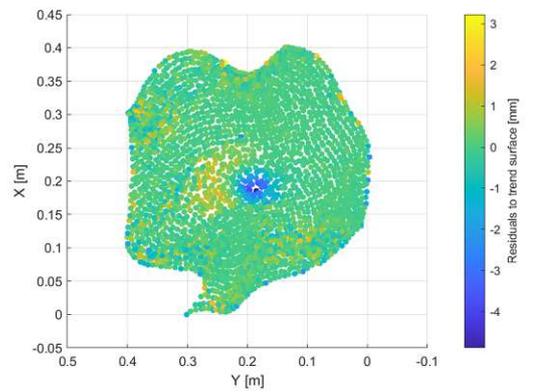


Figure C.9.: D1: x-direction residuals to trend surface epoch 2

## C.2. y-direction epoch 2

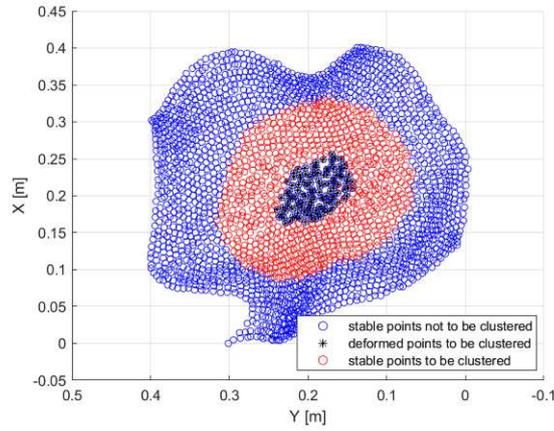


Figure C.10.: D1: y-direction pre-selection of points to be clustered epoch 2

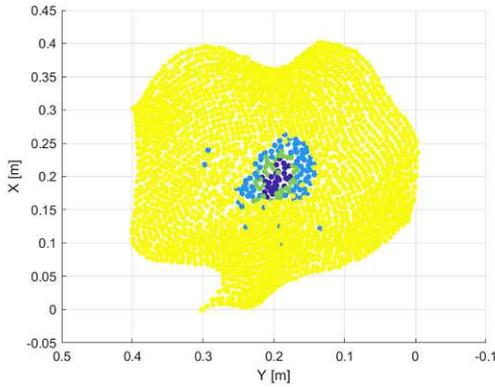


Figure C.11.: D1: y-direction clustering result epoch 2 unfiltered

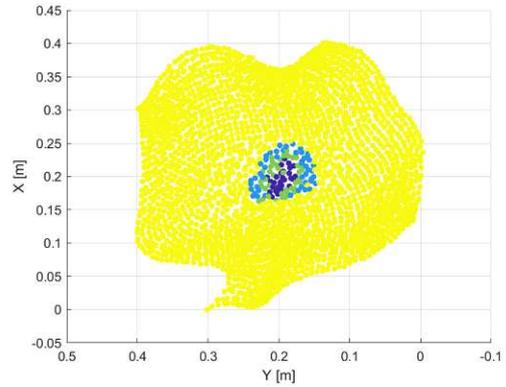


Figure C.12.: D1: y-direction clustering result epoch 2 filtered

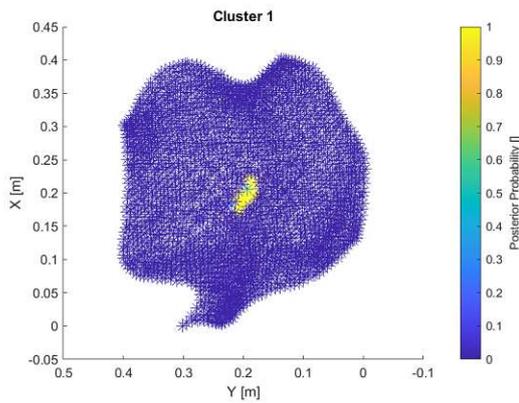


Figure C.13.: D1: y-direction posterior probability epoch 2 cluster 1 unfiltered

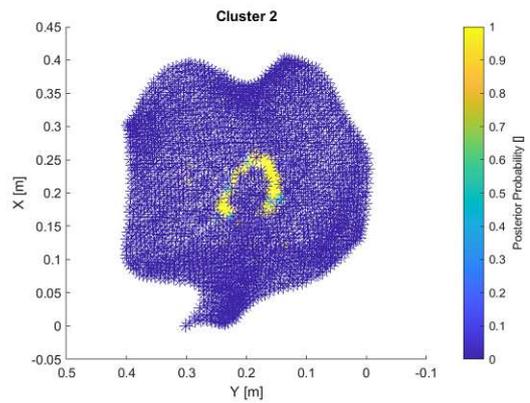


Figure C.14.: D1: y-direction posterior probability epoch 2 cluster 2 unfiltered

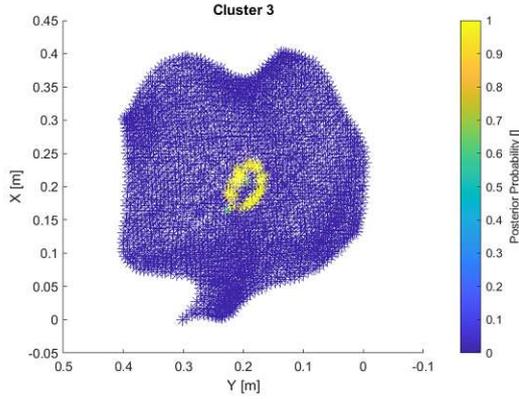


Figure C.15.: D1: y-direction posterior probability epoch 2 cluster 3 unfiltered

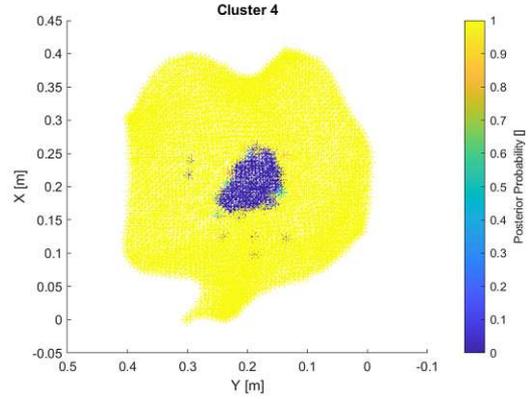


Figure C.16.: D1: y-direction posterior probability epoch 2 cluster 4 unfiltered

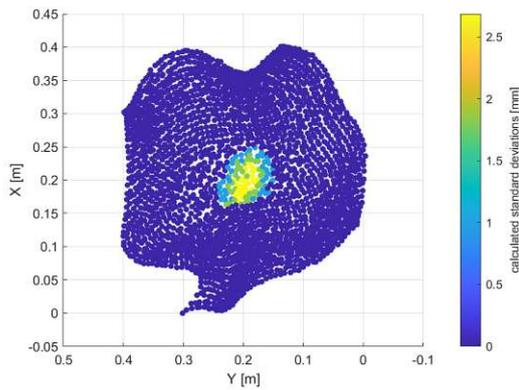


Figure C.17.: D1: y-direction calculated standard deviations epoch 2

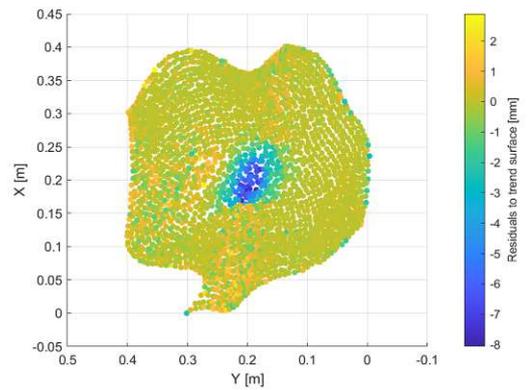


Figure C.18.: D1: y-direction residuals to trend surface epoch 2

### C.3. z-direction epoch 2

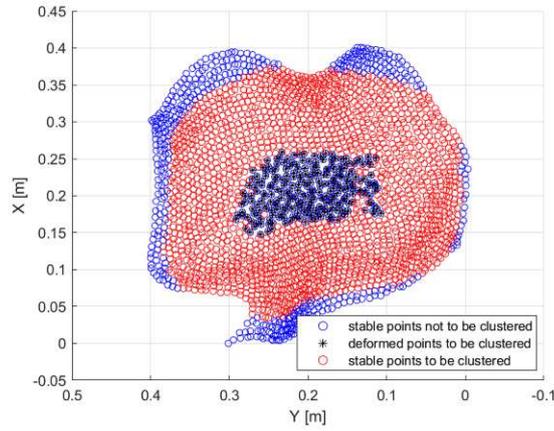


Figure C.19.: D1: z-direction pre-selection of points to be clustered epoch 2

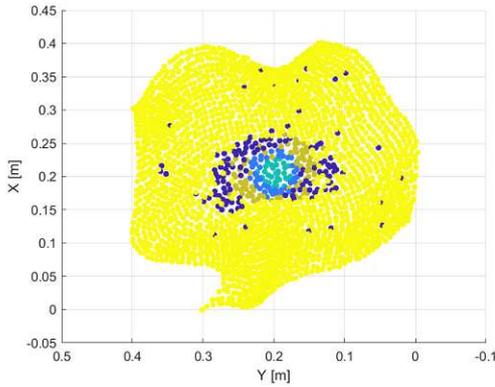


Figure C.20.: D1: z-direction clustering result epoch 2 unfiltered

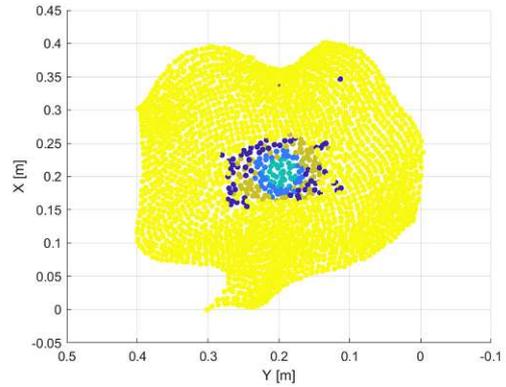


Figure C.21.: D1: z-direction clustering result epoch 2 filtered

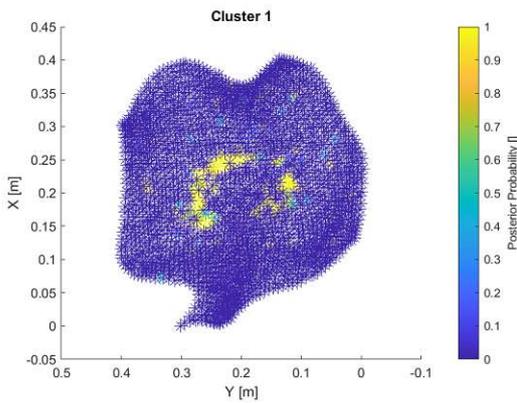


Figure C.22.: D1: z-direction posterior probability epoch 2 cluster 1 unfiltered

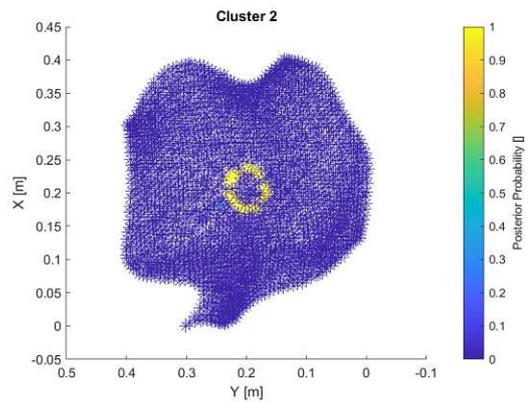


Figure C.23.: D1: z-direction posterior probability epoch 2 cluster 2 unfiltered

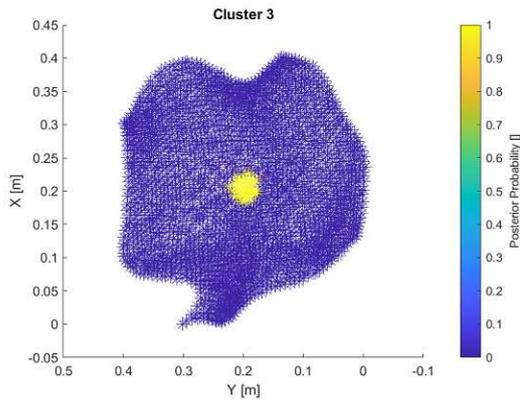


Figure C.24.: D1: z-direction posterior probability epoch 2 cluster 3 unfiltered

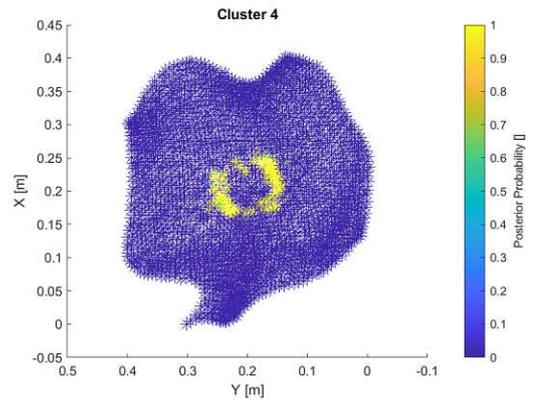


Figure C.25.: D1: z-direction posterior probability epoch 2 cluster 4 unfiltered

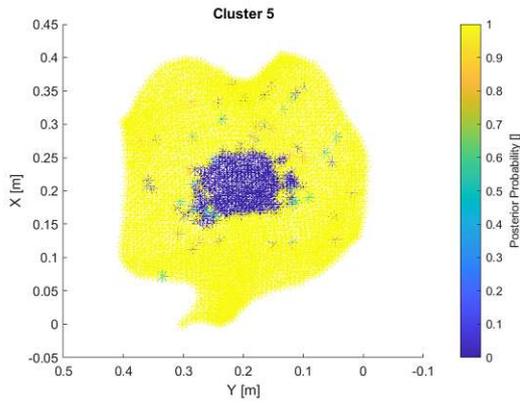


Figure C.26.: D1: z-direction posterior probability epoch 2 cluster 5 unfiltered

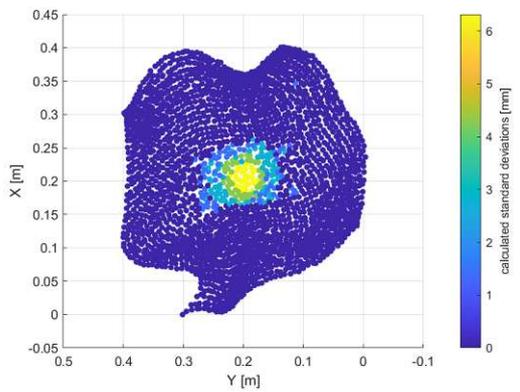


Figure C.27.: D1: z-direction calculated standard deviations epoch 2

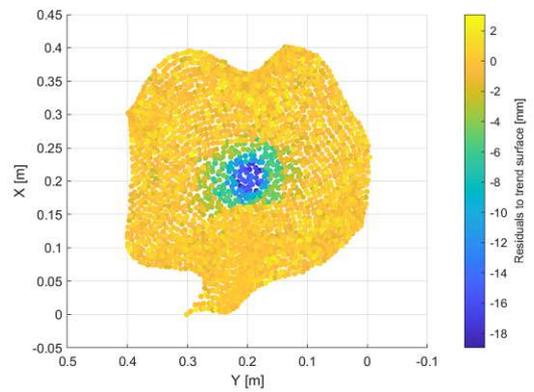


Figure C.28.: D1: z-direction residuals to trend surface epoch 2

### C.4. x-direction epoch 3

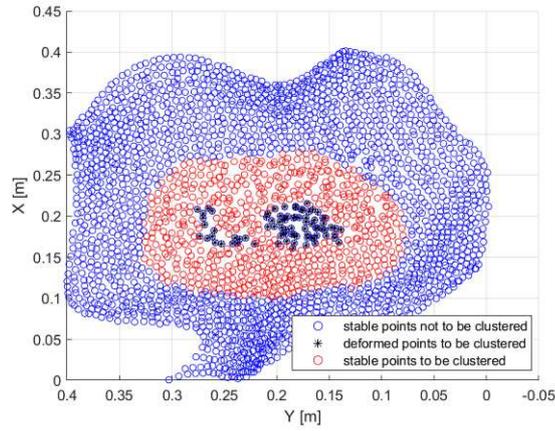


Figure C.29.: D1: x-direction pre-selection of points to be clustered epoch 3

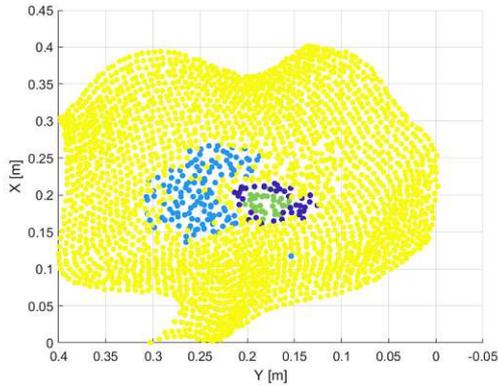


Figure C.30.: D1: x-direction clustering result epoch 3 unfiltered

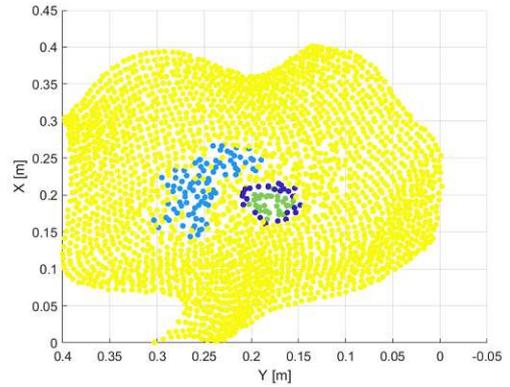


Figure C.31.: D1: x-direction clustering result epoch 3 filtered

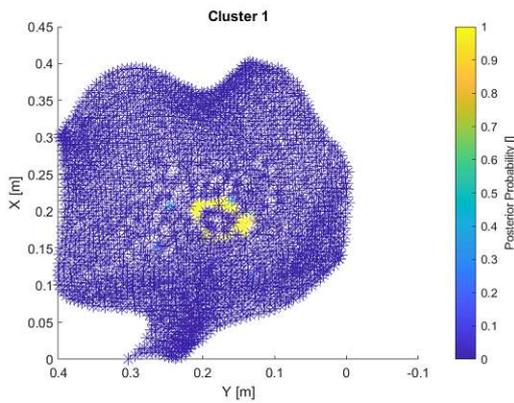


Figure C.32.: D1: x-direction posterior probability epoch 3 cluster 1 unfiltered

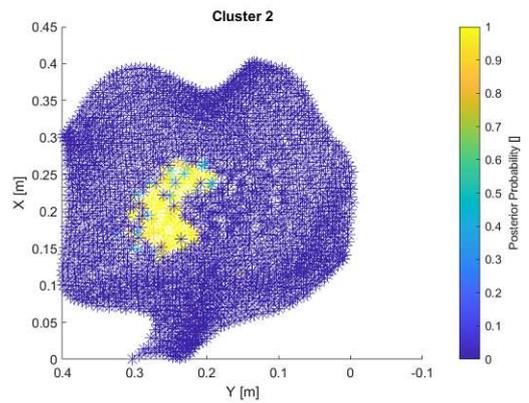


Figure C.33.: D1: x-direction posterior probability epoch 3 cluster 2 unfiltered

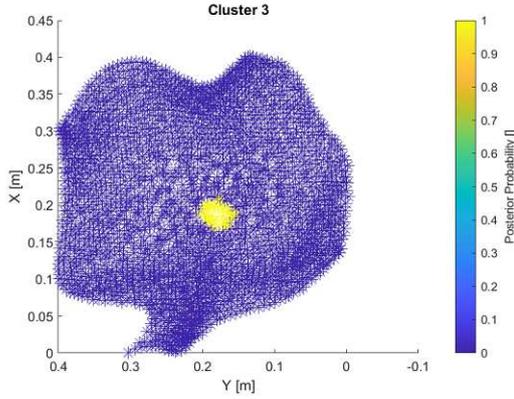


Figure C.34.: D1: x-direction posterior probability epoch 3 cluster 3 unfiltered

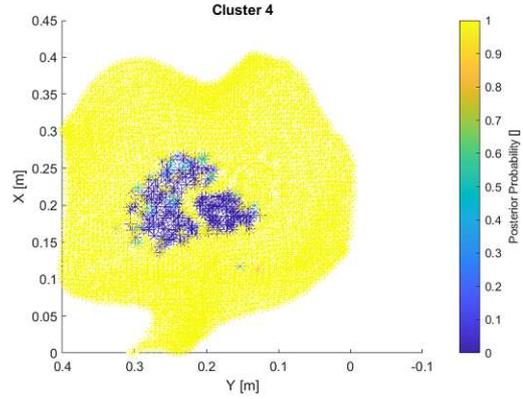


Figure C.35.: D1: x-direction posterior probability epoch 3 cluster 4 unfiltered

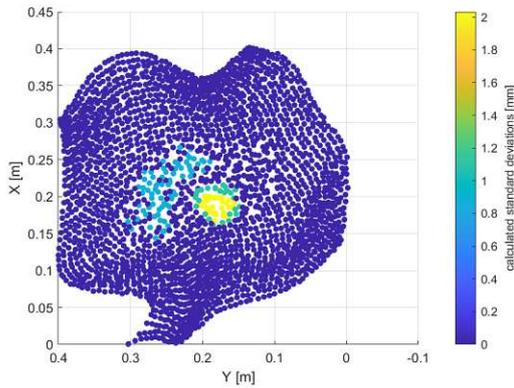


Figure C.36.: D1: x-direction calculated standard deviations epoch 3

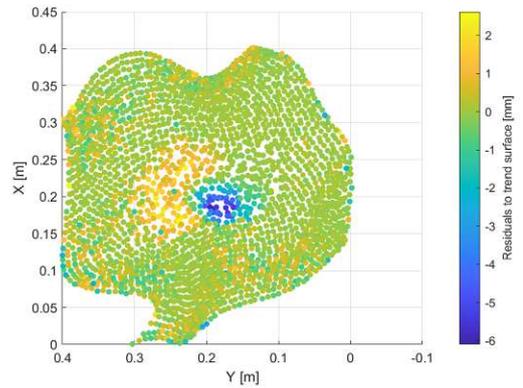


Figure C.37.: D1: x-direction residuals to trend surface epoch 3

### C.5. y-direction epoch 3

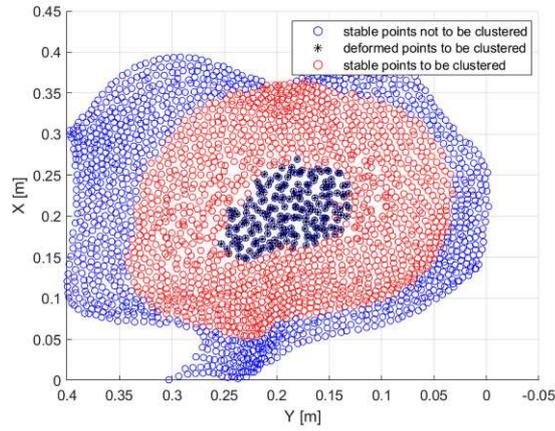


Figure C.38.: D1: y-direction pre-selection of points to be clustered epoch 3

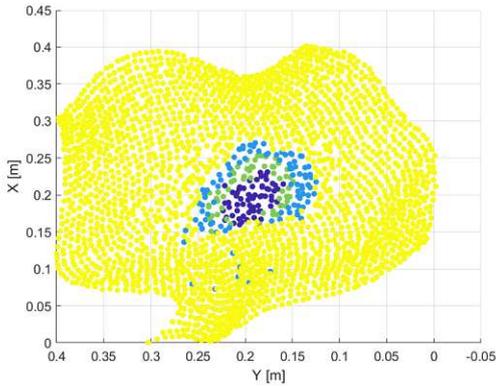


Figure C.39.: D1: y-direction clustering result epoch 3 unfiltered

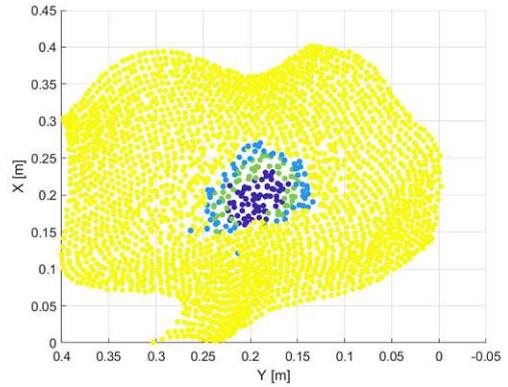


Figure C.40.: D1: y-direction clustering result epoch 3 filtered

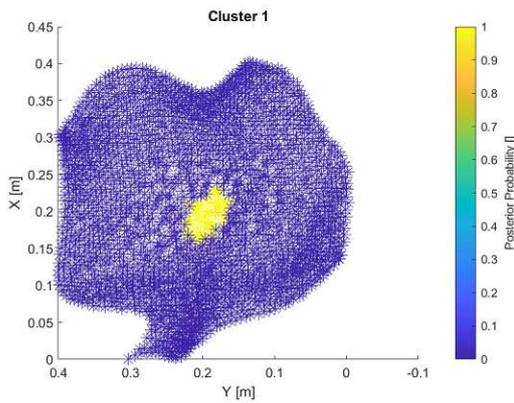


Figure C.41.: D1: y-direction posterior probability epoch 3 cluster 1 unfiltered

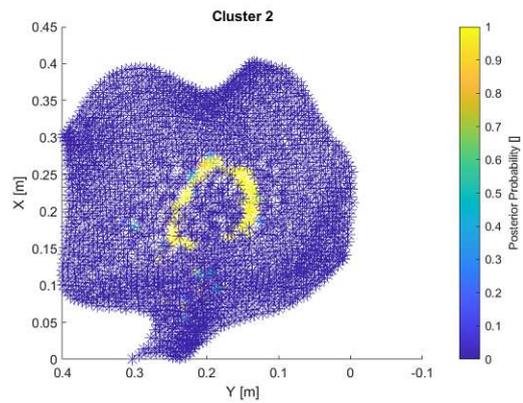


Figure C.42.: D1: y-direction posterior probability epoch 3 cluster 2 unfiltered

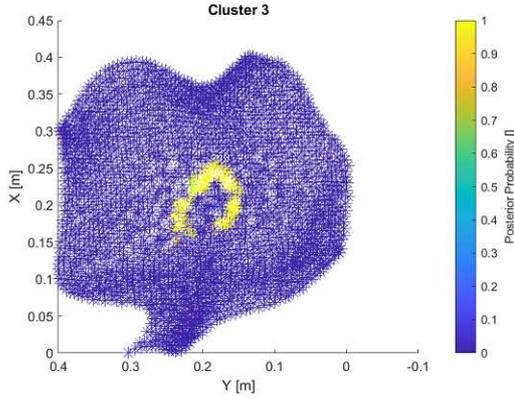


Figure C.43.: D1: y-direction posterior probability epoch 3 cluster 3 unfiltered

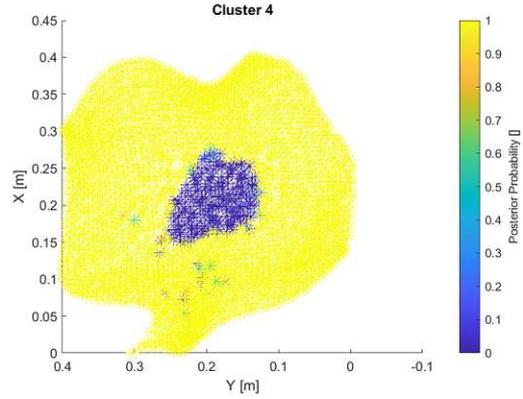


Figure C.44.: D1: y-direction posterior probability epoch 3 cluster 4 unfiltered

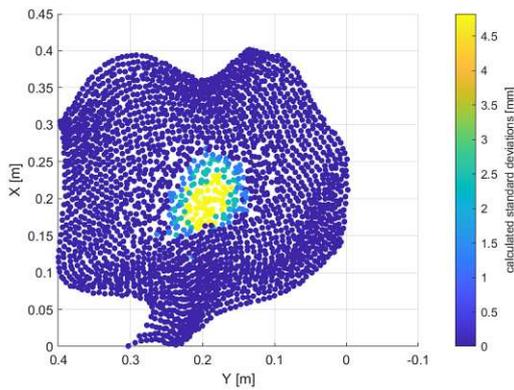


Figure C.45.: D1: y-direction calculated standard deviations epoch 3

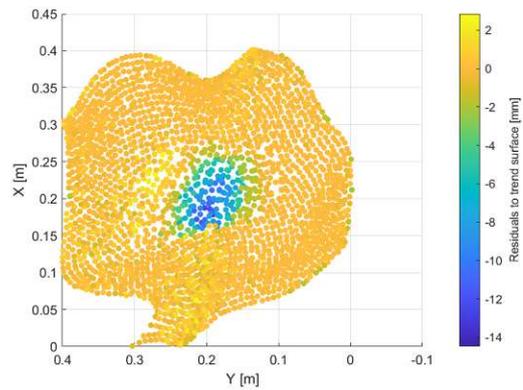


Figure C.46.: D1: y-direction residuals to trend surface epoch 3

### C.6. z-direction epoch 3

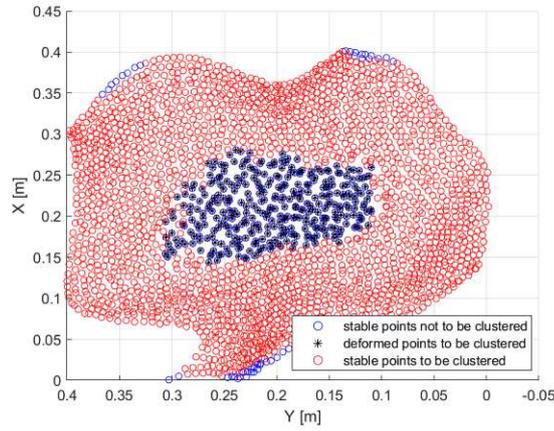


Figure C.47.: D1: z-direction pre-selection of points to be clustered epoch 3

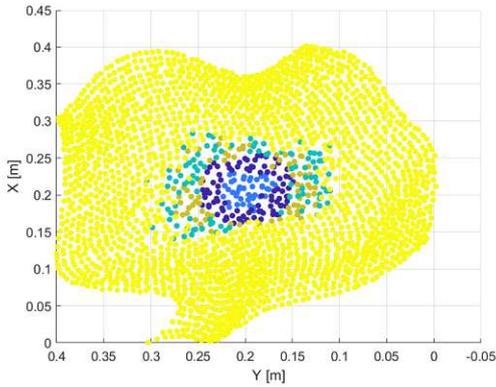


Figure C.48.: D1: z-direction clustering result epoch 3 unfiltered

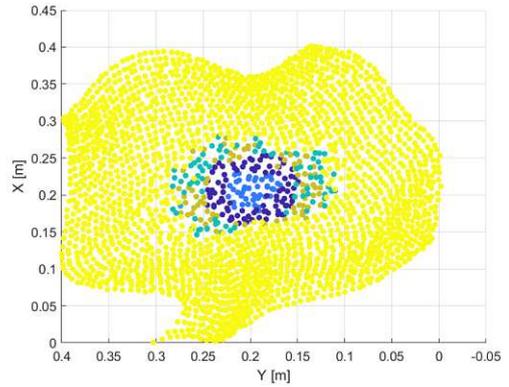


Figure C.49.: D1: z-direction clustering result epoch 3 filtered

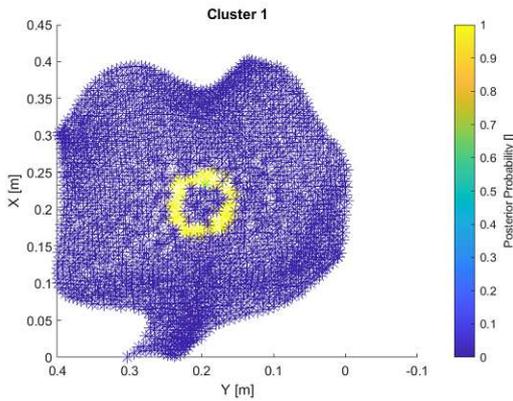


Figure C.50.: D1: z-direction posterior probability epoch 3 cluster 1 unfiltered

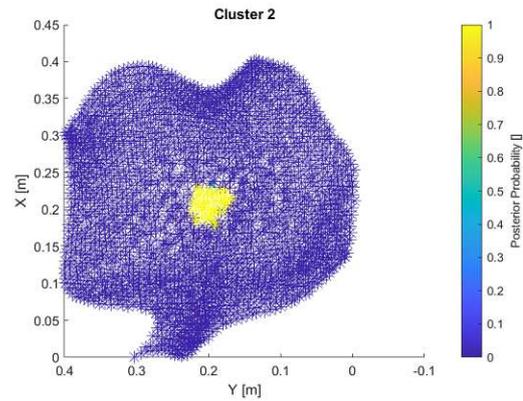


Figure C.51.: D1: z-direction posterior probability epoch 3 cluster 2 unfiltered

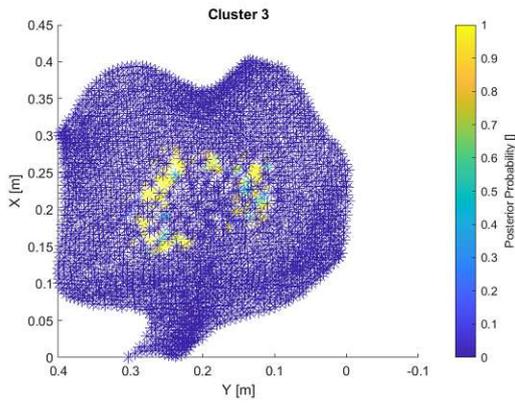


Figure C.52.: D1: z-direction posterior probability epoch 3 cluster 3 unfiltered

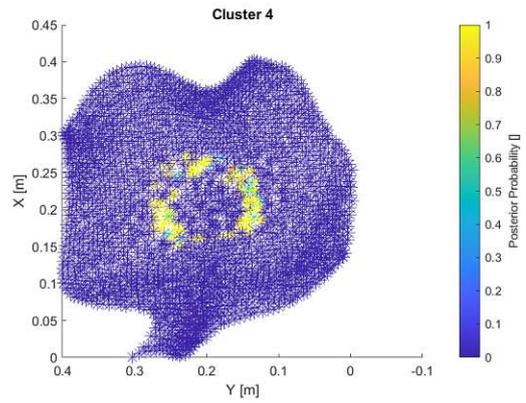


Figure C.53.: D1: z-direction posterior probability epoch 3 cluster 4 unfiltered

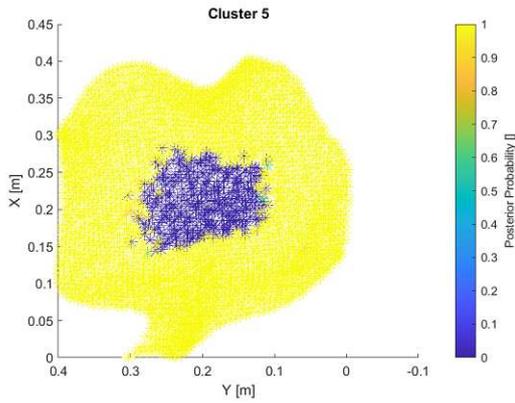


Figure C.54.: D1: z-direction posterior probability epoch 3 cluster 5 unfiltered

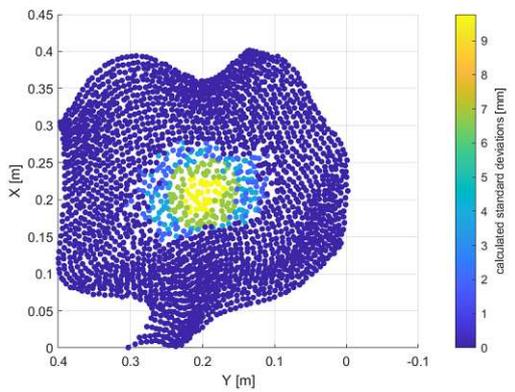


Figure C.55.: D1: z-direction calculated standard deviations epoch 3

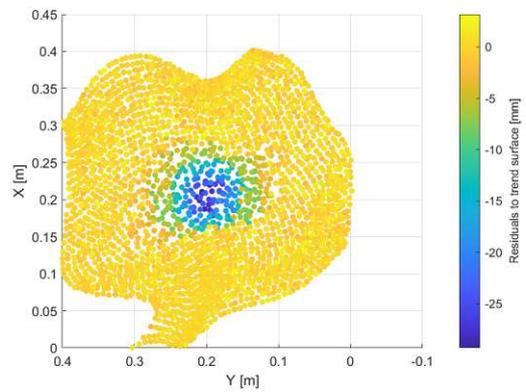


Figure C.56.: D1: z-direction residuals to trend surface epoch 3

### C.7. final results

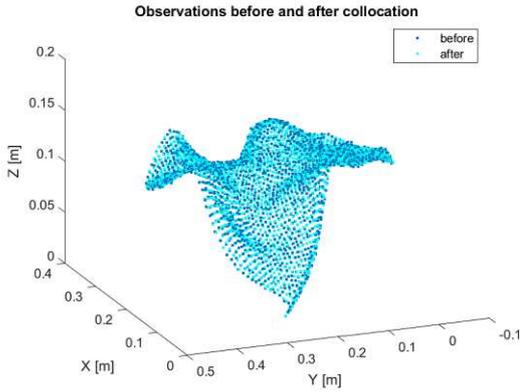


Figure C.57.: D1: x-direction observations before and after collocation

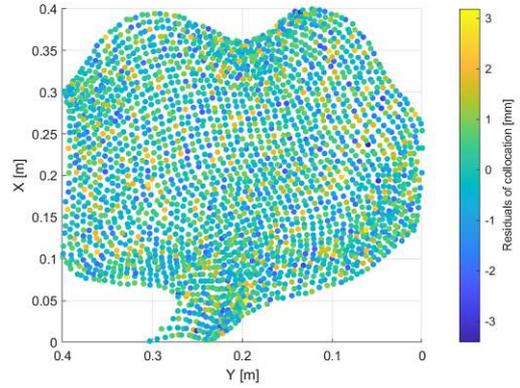


Figure C.58.: D1: x-direction residuals of collocation

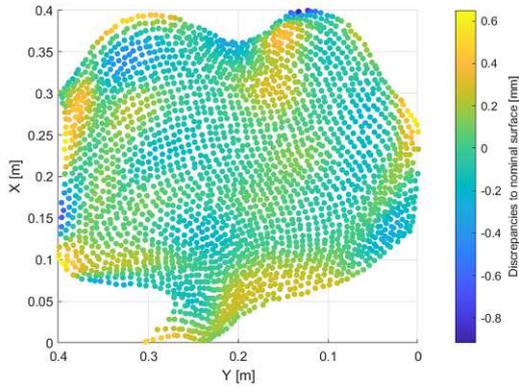


Figure C.59.: D1: x-direction discrepancies to nominal surface

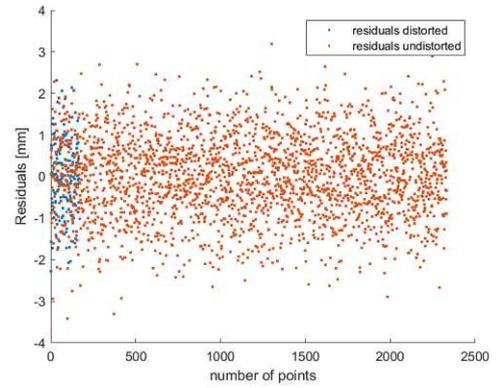


Figure C.60.: D1: x-direction noise

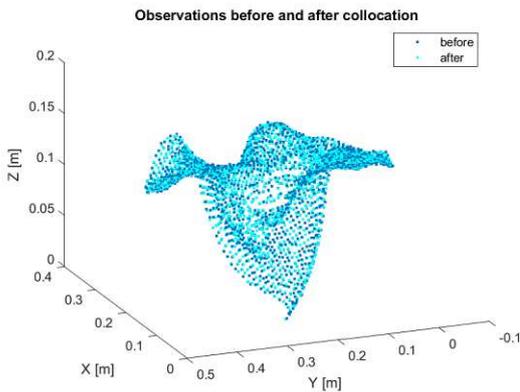


Figure C.61.: D1: y-direction observations before and after collocation

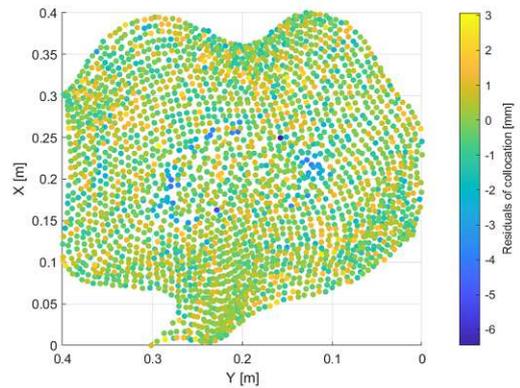


Figure C.62.: D1: y-direction residuals of collocation

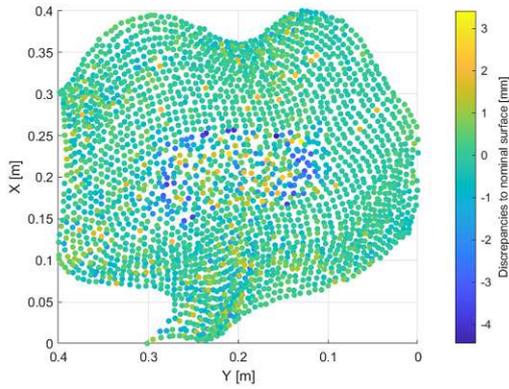


Figure C.63.: D1: y-direction discrepancies to nominal surface

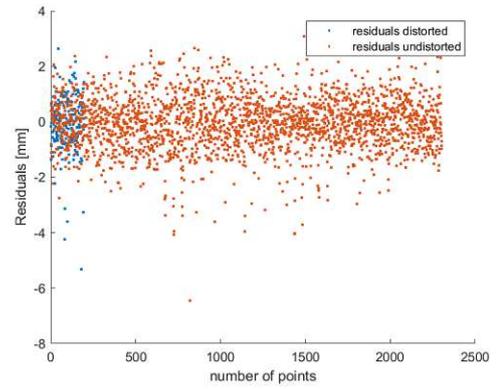


Figure C.64.: D1: y-direction noise

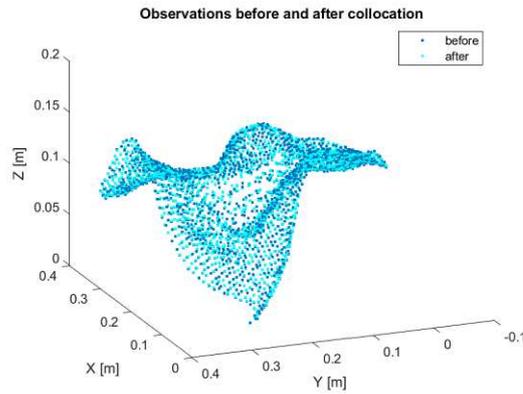


Figure C.65.: D1: z-direction observations before and after collocation

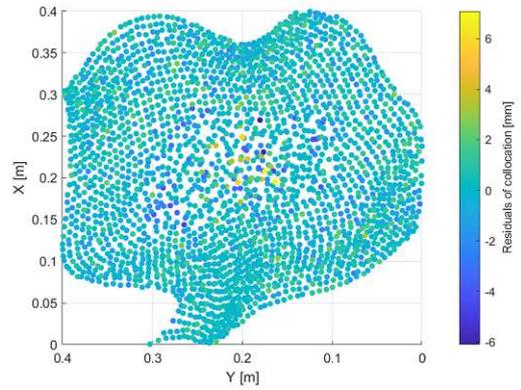


Figure C.66.: D1: z-direction residuals of collocation

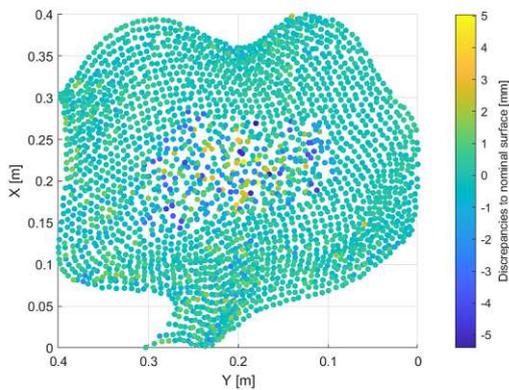


Figure C.67.: D1: z-direction discrepancies to nominal surface

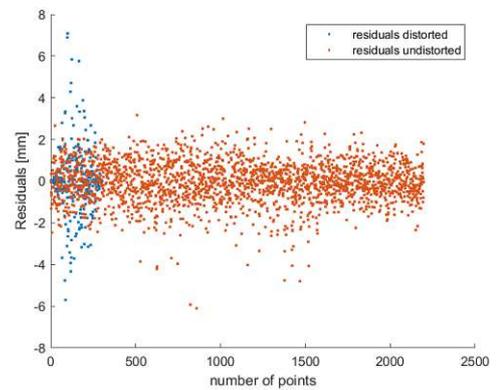


Figure C.68.: D1: z-direction noise