

Optimization on representative element volume (RVE) generation and evaluation process of Q&P980 steel

Trung Nguyen, Ben Nguyen, Byeongmin Oh

**Optimization on representative element
volume (RVE) generation and evaluation
process of Q&P980 steel**

Trung Nguyen, Ben Nguyen, Byeongmin Oh

Computational engineering project submitted in fulfillment of
the requirements for the degree of Bachelor of Science in
Technology.

Otaniemi, 15 December 2023

Supervisor: Junhe Lian

Advisor: Rongfei Juan

Author

Trung Nguyen, Ben Nguyen, Byeongmin Oh

Title

Optimization on representative element volume (RVE) generation and evaluation process of Q&P980 steel.

School School of Engineering**Degree programme** Bachelor's Programme in Science and Technology**Major** Computational Engineering**Code** ENG3082**Supervisor** Junhe Lian**Advisor** Rongfei Juan**Date** 15 December 2023**Pages** 49**Language** English**Abstract**

Quench and Partition (Q&P) steel is a promising material for pursuing advanced automotive and structural applications due to its combination of high strength and ductility. In order to optimize its mechanical properties and performance, it is important to understand the microstructural evolution in Q&P. However, achieving an accurate characterization of the steel microstructures is a significant challenge, primarily due to the complex nature of the material having 5 different phases.

This study utilize Representative Volume Elements (RVEs) in the analysis of Q&P980 steel microstructures. RVEs are fundamental in providing a means to capture statistically representative information within a finite sample volume for analysis. In the case of Q&P980 steel, the RVEs are characterized by a mixture of phases containing ferrite, austenite, martensite and bainite. In order to optimize the analysis, it is important to generate an RVE that have a minimal difference to the referenced data.

In this study, we have extended and modify the current pipeline for multi-phase generation by adding phase segmentation into the process, as well as create a new and improved pipeline that can evaluate the RVE more accurately. The previous pipeline is used for analysing AISI 439 and the modified version is used for Q&P980 microstructure, generated a batch of RVE and evaluated to find the most accurate RVE. The efficiency of the pipeline is described, as well as the result are discussed.

Although more things can be done to further improve the outcome of the project, we believe that what we have done is an important step toward more accurate and efficient multi-phase material modelling.

Keywords Microstructure modelling; RVE generation; Phase segmentation; Process evaluation.

urn <https://aaltodoc.aalto.fi>

Contents

Contents	iii
Symbol and Abbreviation	v
1. Introduction	1
2. Theoretical Background	3
2.1 RVE generation studies	3
2.2 Single phase material - AISI439	4
2.3 Multiple phases material - Q&P980	4
2.3.1 Quenching and Partitioning process	5
2.3.2 Target Material	7
3. Methodology	9
3.1 Single-phase RVE generation and evaluation	9
3.1.1 Single-phase RVE generation	9
3.1.2 Current RVE evaluation process	12
3.1.3 RVE evaluation by fitting ellipsoids to 3D grains	14
3.2 Pipeline extension to Q&P steel generation	17
3.2.1 Phase segmentation methods for Q&P steel . .	18
3.2.2 Multi-phase RVE generation	19
3.3 Multi-phase RVE evaluation process	20
3.3.1 Extension of slicing method to multi-phase RVEs	21
3.3.2 Extension of fitting ellipsoids method to multi- phase RVEs	21
4. Result	23
4.1 Single-phase RVE generation and evaluation using cur- rent pipeline	23
4.2 Multi-phase RVE generation	25

4.2.1	RVE generation output	26
4.2.2	Time and space complexity	27
4.3	Multi-phase RVE evaluation	27
4.3.1	RVE evaluation using slicing method	28
4.3.2	RVE evaluation using fitting ellipsoid method . .	28
4.3.3	Time complexity	30
5.	Discussion	32
5.1	Multi-phase RVE generation discussion	32
5.2	Multi-phase RVE evaluation discussion	34
6.	Self-evaluation	36
6.1	Ben Nguyen	36
6.2	Byeongmin Oh	37
6.3	Trung Nguyen	38
Reference		40
A.	Fitting an ellipsoid to an arbitrary three-dimensional shape	42
A.1	General equation of an ellipsoid in three dimensions . . .	42
A.2	Fitting an ellipsoid to a 3D region identified by points . .	44
B.	Multi-phase evaluation using fitting ellipsoids method: full results	46

Symbol and Abbreviation

AISI American Iron and Steel Institute

BCC Body Centered Cubic

CS Crystal Structure

EBSD Electron Backscatter Diffraction

FCC Face Centered Cubic

GOS Grain Orientation Spread

IQ Image Quality

KAM Kernel Average Misorientation

MDF Misorientation Distribution Function

MOE Modulus of Elasticity

ODF Orientation Distribution Function

Q&P Quench and Partition

RD Rolling Direction

RVE Representative Volume Element

TD Transverse Direction

TE Tensile Elongation

UE Ultimate Elongation

UTS Ultimate Tensile Strength

YS Yield Strength

1. Introduction

The optimization of Representative Volume Elements (RVEs) in the analysis of Quench and Partition (Q&P) steel microstructures is an important step to understand this innovative material. Q&P steel, known for its excellent combination of strength, ductility, and toughness, has received considerable attention in various engineering applications, including automotive manufacturing, structural engineering, and beyond. The microstructural in Q&P steel contain a complex mixture of martensite, retained austenite, bainite and ferrite, require a profound understanding at the microstructure[1, 2].

The concept of RVEs, which enables the representation of a material's complex microstructure within a finite sample volume, has emerged as a powerful tool in the study of heterogeneous materials[3]. When applied to Q&P steel, RVEs offer the promise of providing insights into the underlying mechanisms that govern its special mechanical properties. This is driven by the recognition that, while mechanical properties of the material is undeniably linked to the features within the material's microstructure [4, 5, 6].

However, achieving an accurate characterization of the steel microstructures is a significant challenge, primarily due to the complex nature of the material having 5 different phases [7].

This study utilize Representative Volume Elements (RVEs) in the analysis of Q&P steel microstructures. In the case of Q&P steel, the RVEs are characterized by a mixture of phases containing ferrite, austenite, martensite and bainite. In order to optimize the analysis, it is important to generate a RVE that have a minimal difference to the referenced data.

This study develop a pipeline that apply novel scientific finding in order to

increase the accuracy of RVE. The focus of the project is in generating and evaluating mono-phase and multi-phase material in different approach.

There are 4 sections in the remaining portion of the paper. In the second section, theoretical background related to the topic is provided and explained. Next, the methodology in generating the RVE, as well as RVE development is presented. The results are then presented and discussed.

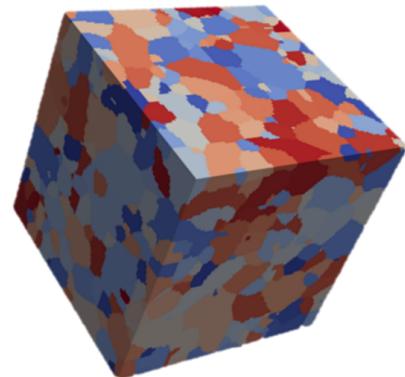


Figure 1.1. Example of an RVE we have generated during our project.

2. Theoretical Background

This section provide the definition of concepts and process discussed in the study. The material used in the project are described.

2.1 RVE generation studies

Representative Volume Element (RVE) is a material volume that can be use for representing the material [3]. The RVE contain information about the microstructure, which correlate with the properties of the material as a whole. The optimized RVE has a correct size, meaning that it is big enough to accurately represent the material, meaning that it have enough grains to perform the analysis correctly, while not being too big that it requires too much computational power for analysis [8]. There are two main properties for the RVE, dimension, which determine the size, and resolution, which determine the smallest grain size that can be captured.

In RVE generation, it is commonly used three types of generation methods for the RVEs of composites: inclusion non-intersection based method, inclusion intersection removal based method and three-dimensional (3-D) image reconstruction based method. In our project, Dream 3D is utilized in RVE generation.

The generation process typically use grains data. The grain data can varies, but usually the grains size, shape, texture, misorientation and tilt angle are considered for generation. The distribution for these parameters can be obtain using Electron Back-scatter Diffraction (EBSD) on the chosen material [9]. EBSD provide crystallographic information, which can be use for analysis and collect the data needed.

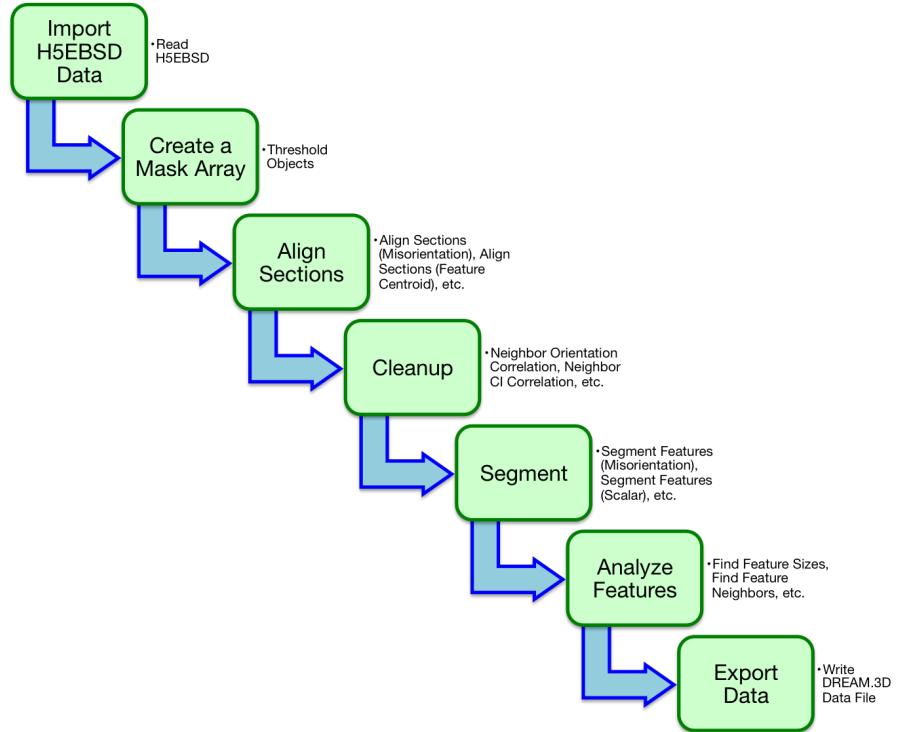


Figure 2.1. EBSD reconstruction steps in Dream3D

2.2 Single phase material - AISI439

The material used for single phase is AISI439, a ferritic chromium stainless steel, stabilized with chromium. The material has a good corrosion resistance and can be used in mildly corrosive environment. AISI439 is used mostly in automotive exhaust system, sugar industry equipment or household application [10]. The chemical composition of the steel is presented in Table 2.1 and the mechanical properties is presented in Table 2.2 [11].

Table 2.1. Chemical composition of AISI439, % by mass

C (<=)	Si (<=)	Mn (<=)	P (<=)	S (<=)	Cr	Ti (<=)
0.05	1.00	1.00	0.04	0.015	16.0-18.00	4x(C+N) + 0.15

Table 2.2. Mechanical properties of AISI439

Hardness, HB	YS, MPa	UTS, MPa	TE, %	MOE, GPa
185	270	450-600	20	220

2.3 Multiple phases material - Q&P980

In this part, we will introduce the main target material of our project: multi-phase Q&P980 steel. The overall material process is given, as well

as an overview of the material's mechanical properties.

2.3.1 Quenching and Partitioning process

Quench and Partition (Q&P) steel is known for being a high performance steel, having superior strength and ductility compare to dual phase (DP) steel [12]. The manufacture process consist 2 process: quenching (cooling the material rapidly) and partitioning (heating the material for diffusion), shown in Figure 2.2.

The quenching process means the rapid cooling of a heated material, typically a metal, by immersing it in a quenching medium, such as water, oil, or air. During rapid cooling, the atoms within the material are locked in a certain arrangement, increasing strength and hardness, leading to the formation of a desired microstructure, such as martensite [13, 14].

Quenching has an important roles in the production of various industrial components, from automotive gears to cutting tools, as it create the necessary hardness required for their respective applications. During the process, there are multiple different variables that affect the final result. These include cooling rates, quenching medium and temperature control.

The partitioning process is the second step within the Quench and Partition (Q&P) heat treatment technique, primarily used in enhancing the mechanical properties of steel alloys. During the partitioning process, the previously formed metastable phase, typically austenite retained from the quenching stage, is allowed to distribute its carbon content within the microstructure. This transformation is achieved through a controlled heat treatment at elevated temperatures, stabilizing atoms in the steel's matrix [13, 14].

Partitioning stimulate the development of a unique microstructure characterized by a balance between strength and ductility. This innovative approach has led to the production of advanced high-strength steels with exceptional mechanical properties, making them highly sought-after materials for a wide range of applications, from automotive components to structural elements in construction.

Steel going through these two process contain 5 different phases [15]:

- Ferrite is an alpha steel phase, meaning that it has the crystal

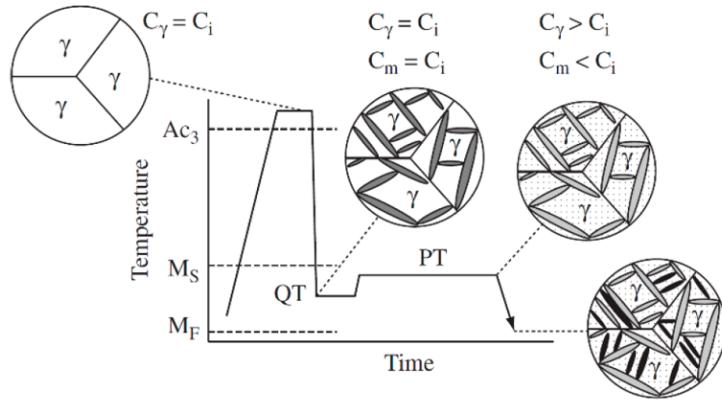


Figure 2.2. Quenching and Partitioning process

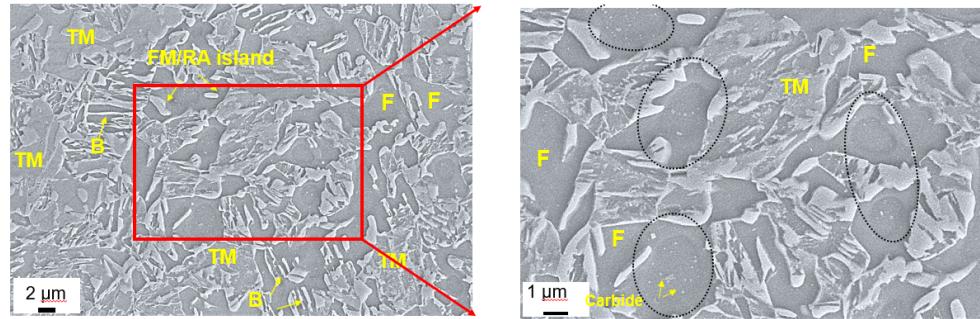


Figure 2.3. QP steel contains different steel phases

structure of Body Centered Cubic (BCC). Ferrite is soft and ductile compare to martensite, which help with formability.

- Retained metastable austenite (RA) is a gamma steel phase, having Face Centered Cubic (FCC) crystal structure. Austenite is stable in high temperature (over 912 Celsius degree) and in room temperature, most of austenite convert to other phases. Austenite is soft and ductile compare to the other iron phases.
- Fresh Martensite is an alpha steel phase (BCC) created from the quenching process. Because of rapid cooling, the microstructure stop carbon from diffusing, generating internal stress, creating a hard and compressed brittle structure.
- Tempered Martensite is an alpha steel phase (BCC) created from Fresh Martensite after the partitioning process. By heating martensite and slow cooling, relieving internal stress, martensite hardness and strength is slightly reduce while the microstructure stability and material ductility is significantly increased.
- Bainite is an alpha steel phase (BCC) created during the phase change of austenite. Bainite consists thin needle-like ferrite plate intercept with fine particle of cementite.

2.3.2 Target Material

In this study, the target material for analysis is QP980 steel. QP980 is a third generation advance high strength steel (AHSS). The chemical compositing of the material is shown in Table 2.3 [12].

Table 2.3. Chemical composition of QP980, % by mass

C	Si	Mn	Al	Cr	Co	Mo+Ni+Cu	P+S+Nb
0.15	1.67	2.42	0.074	0.02	0.026	<0.01	<0.006

The mechanical properties of QP980 steel is summarized in Table 2.4. [12].

Table 2.4. Mechanical properties of QP980

YS, MPa	UTS, MPa	UE, %	TE, %	R-value at UE, -
752.2	1090.20	18.09	23.57	0.75

The engineering stress-strain curve and true stress-strain curve of the material is present in Figure 2.4, 2.5.

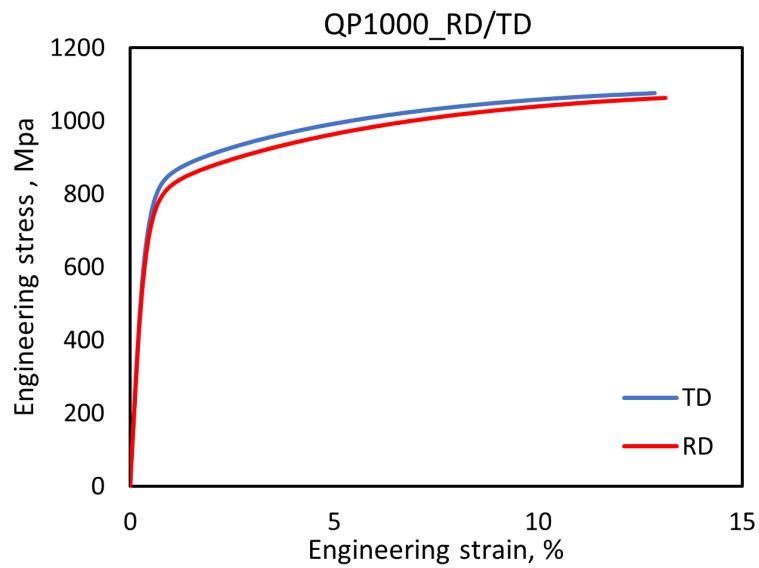


Figure 2.4. QP980 Engineering Stress-Strain Curve in RD and TD

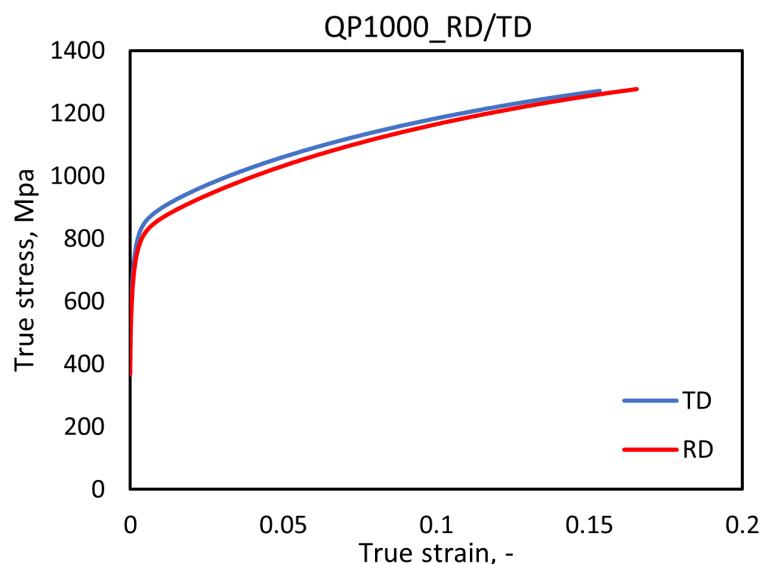


Figure 2.5. QP980 True Stress-Strain Curve in RD and TD

3. Methodology

In this part, we propose an automatic pipeline for generating RVEs for Q&P980 steel with high accuracy. First, a general pipeline for generating general RVE and post-evaluating is discussed. The pipeline is then extended to Q&P980 steel by including phase segmentation for independent phase generation, and finally an extra step is presented to evaluate the spatial phase distribution of the generated multi-phase RVEs.

3.1 Single-phase RVE generation and evaluation

An existing pipeline for automatically generating and evaluating *single-phase RVEs* is presented. Here, our main focus will be on the structures of the pipelines, possible computational optimizations, and the inclusion of the distribution curves of the material.

3.1.1 Single-phase RVE generation

Given a known single-phase EBSD data, a pipeline for generating RVEs is given in Figure 3.3. The first step of the generation process is to extract the statistical data of the material's features from the EBSD. We will divide this into two parallel branches: the first one analyzes the grains' size and shape distribution of the material, and the second one extracts the ODF data from the EBSD. In the first branch, the grains are first fitted into equivalent ellipses: the corresponding ellipse has the same area as the grain's area. These ellipses have two important characteristics: their equivalent diameters (Figure 3.1) which tell us the size of the grains, and their aspect ratios (Figure 3.2) which tell us the shape of the grains (values near 0 being extremely elongated and values near 1 being more equiaxed).

The grains' equivalent diameters d are then fitted into a *log-normal*

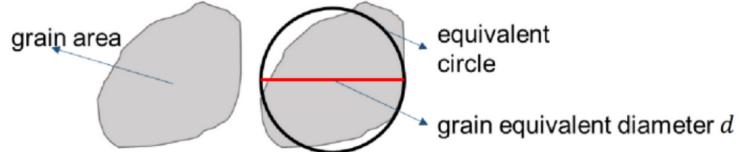


Figure 3.1. Fit a circle to the points making up a grain based on the equivalent area criterion

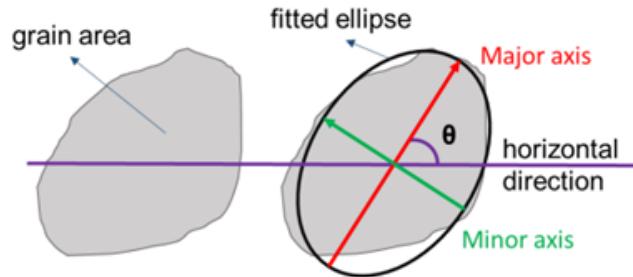


Figure 3.2. Fit an ellipse to the points making up a grain based on the equivalent area criterion

distribution with the formula:

$$f(d; \mu, \sigma)dy = \frac{1}{d\sigma\sqrt{2\pi}} e^{-\frac{(\ln d - \mu)^2}{2\sigma^2}} dy \quad (3.1)$$

By using maximum likelihood estimation, we can find the corresponding value μ and σ for the distribution of our empirical equivalent diameters data. The average size (diameter) can be calculated from the distribution as:

$$\bar{d} = e^{\left(\mu + \frac{\sigma^2}{2}\right)} \quad (3.2)$$

Similarly, we also fit the aspect ratio data a using a *beta distribution* with the formula:

$$f(a; \alpha, \beta)dy = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} a^{\alpha-1} (1-a)^{\beta-1} dy \quad (3.3)$$

and get the distribution parameters α and β using maximum likelihood estimation, and the empirical average aspect ratio can be calculated as:

$$\bar{a} = \frac{\alpha}{\alpha + \beta} \quad (3.4)$$

The fitting of equivalent ellipses and estimation of distribution curves can be done in MTEX - an open-source toolbox for analyzing and modeling crystallographic textures by means of EBSDs. In the second branch, we utilize DREAM3D - a collection of data analysis tools for materials science research - to build a separate pipeline and extract the ODF data into a

separate file. An existing pipeline *Export StatsGenerator ODF Angle File* would export a file with Euler angles data ($\varphi_1, \phi, \varphi_2$), weight w , and σ , and this file will be used as an input for the ODF data of the RVE generation.

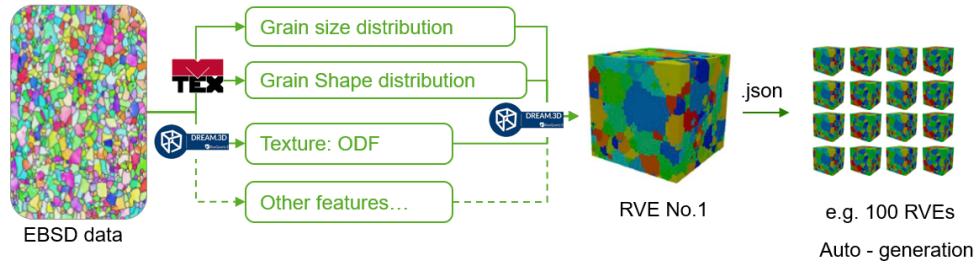


Figure 3.3. RVE automatic generation pipeline for single-phase materials.

The next step in our generation process is building a pipeline for RVE generation on DREAM3D. The pipeline includes the following filters:

- StatsGenerator: this filter takes in μ and σ for log-normal size distribution, α and β for beta shape distribution, and the ODF file produced earlier as input for RVE's ODF.
- Initialize Synthetic Volume: creates an empty volume as a base for inserting features to create a synthetic microstructure. Here, the dimension of our RVE will be $100 \times 100 \times 100$ to capture both small and large grains, and the resolution will be set to $5 \times 5 \times 5$.
- Establish Shape Types: here we assign the *ellipsoid* shape type to each ensemble of a synthetic structure.
- Pack Primary Phases: place precipitate features with the sizes, shapes, physical orientations and locations corresponding to the goal statistics.
- Find Feature Neighbors: determines, for each feature, the number of other Features that are in contact with it.
- Match Crystallography: match a defined orientation distribution function (ODF) to a set of features using a "switch or swap" approach.
- Generate IPF Colors: generate *inverse pole figure (IPF)* colors for our cubic crystal structures.
- Write DREAM.3D Data File: write the contents of the current data structure to an HDF5 based file with the file extension .dream3d. This file will be used for future access to the RVE.
- Export Feature Data as CSV File: writes the data associated with each Feature to a file in CSV format. This file will be used for RVE evaluation process.

- Export Los Alamos FFT File: writes out cell data from an image geometry to a TXT file. This file will be used for RVE evaluation process.

Finally, we will use the pipeline to create multiple RVEs at the same time. This can be done by utilizing Puhti supercomputer from CSC - IT Center for Science Ltd., a company providing IT support and modeling, computing, and information services for academia, research institutes, and companies in Finland. By applying core-based HPC (High Performance Computing) to run several DREAM3D simulations, the generation time of our RVEs can be reduced significantly for evaluation. The algorithms for RVE generation used by DREAM3D is discussed at [16] [17].

The code used in this project is available at <https://github.com/LianjGroup/RVE-Micromechanics-Project>.

3.1.2 Current RVE evaluation process

An important step in RVE generation is the post-evaluation process, where the generated RVE is compared to the original material on several characteristics: the smaller the difference, the more representative of the material our RVE is. This is not a simple task however, since comparing a 3D structure with a 2D EBSD scan requires some extra preprocessing, and the choices of features for evaluation are also restricted. We will first discuss an existing process for evaluating RVEs regarding their size and shape distribution, and from here we extend the process by including misorientation into the evaluation.

An existing pipeline for evaluating RVEs is given in Figure 3.4.

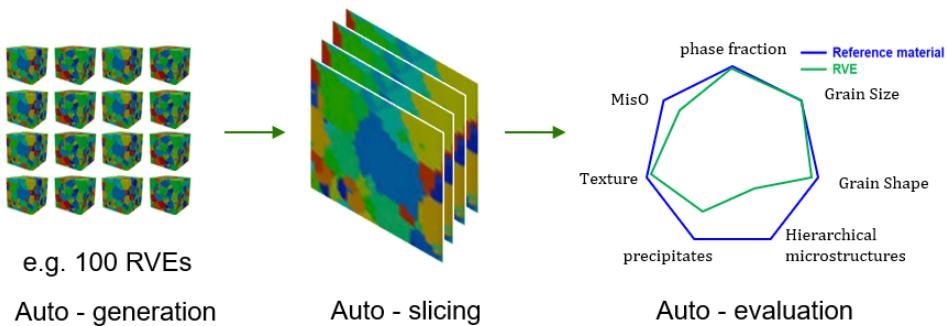


Figure 3.4. RVE automatic evaluation pipeline: an extra step includes slicing the generated RVE along z-axis [18].

a) RVE size evaluation

Using the Feature Data CSV file exported from Dream3D, the volume data from each grain are collected. The volume are used to calculate the grain equivalent diameter d , and this can be done by first explicitly define the formula for a sphere of volume V .

$$V = \frac{4}{3}\pi r^3 = \frac{4\pi}{3} \left(\frac{d}{2}\right)^3 \quad (3.5)$$

Then, the equivalent diameter can be calculated as:

$$d = 2 \left(\frac{3V}{4\pi}\right)^{\frac{1}{3}} \quad (3.6)$$

The average equivalent diameter is then compared with the experimental data analysed from MATLAB, providing the **size deviation**.

b) RVE shape and texture evaluation by slicing method

The Los Alamos FFT File exported from Dream3D is used for shape evaluation. The file provide the data for each voxel in the RVE, containing the angular rotation, the position of the voxel, the grain ID and the phase that is containing the grain. The file is imported and processed with MATLAB using MTEX.

The processing of each RVE layer is described as the following. First, the txt file containing the data from the layer is imported and processed such that data from each column are assigned to its correct meaning, providing an EBSD data variable.

The EBSD data is first used for grains reconstruction with grain boundary is 15 degree. Then, the grains data is used to eliminate the very small grains which can come from measurement error and reconstructed the grains again using the EBSD data after the elimination process. The boundary grains are removed. For the shape evaluation, the remaining grains are fitted with ellipse using the build in command from Mtex, which provide the major axis and minor axis needed to calculated the grain aspect ratio. Another way to find this value is use the build in function which provide grain aspect ratio for all the grains for one phase. For the texture, the ODF is calculated and then used to find the texture index.

Grain rotation	Grain spartial	ID	Phase
325.908	41.571	11.309	15 1 1
325.908	41.571	11.309	2132 4
325.908	41.571	11.309	16 1 1
325.908	41.571	11.309	2132 4
325.908	41.571	11.309	17 1 1
325.908	41.571	11.309	2132 4
325.908	41.571	11.309	18 1 1
325.908	41.571	11.309	2132 4
325.908	41.571	11.309	19 1 1
325.908	41.571	11.309	2132 4
325.908	41.571	11.309	20 1 1
325.908	41.571	11.309	2132 4
325.908	41.571	11.309	21 1 1
325.908	41.571	11.309	2132 4
325.908	41.571	11.309	22 1 1
325.908	41.571	11.309	2132 4
325.908	41.571	11.309	23 1 1
325.908	41.571	11.309	2132 4

Figure 3.5. Alamos FFT File

These two values: grain aspect ratio and texture index from each grain are taken average of and compare to the experimental data, giving the **shape deviation** and **texture deviation**.

The average of the three deviation is use to determine the goodness of an RVE,

However, this method have some disadvantages:

- Because there are multiple phases inside the data file, removing the boundary grains became difficult to implement. As the file is sliced, some grains from low fraction phase became miniatures and was deleted, hence doesn't provide enough information for accurate evaluation.
- The resulting size ratio is incorrect, as the longest and smallest axis of the grain can be rotated to the slicing direction.

3.1.3 RVE evaluation by fitting ellipsoids to 3D grains

Another approach to measure how accurately our RVE represents a material is by directly evaluating the three-dimensional data given to us without slicing them up into 2D layers. This approach is inspired by MTEX algorithm on fitting ellipses to 2D grains in EBSD data [19] to get both the equivalent diameters and the grain aspect ratios, and here our aim is to fit a 3D ellipsoid to a 3D region p of n points to get the best-fitted ellipsoid of a 3D grain. First, we have the general equation of an ellipsoid in three

dimensions is given as:

$$(\mathbf{p} - \mu)^T (R^{-1})^T L^{-1} R^{-1} (\mathbf{p} - \mu) = 1 \quad (3.7)$$

where \mathbf{p} is the point on the surface of the ellipsoid, μ is the center of the ellipsoid, L is the diagonal matrix of squared semi-axes' lengths, and R is the rotational matrix that contains the Euler angles data of the ellipsoid's orientation. Let $\Sigma = R^T L R$ and by using parameter estimation of *trivariate Gaussian distribution* [20], we can see that if a grain in the RVE consists of n points $\mathbf{p} = (\mathbf{x}, \mathbf{y}, \mathbf{z}) = (x_i, y_i, z_i), i = 1, 2, \dots, n$, then the center μ and Σ can be evaluated by:

$$\mu = \begin{bmatrix} \mu_x & \mu_y & \mu_z \end{bmatrix}^T = \begin{bmatrix} \frac{1}{n} \sum_{i=1}^n x_i & \frac{1}{n} \sum_{i=1}^n y_i & \frac{1}{n} \sum_{i=1}^n z_i \end{bmatrix}^T \quad (3.8)$$

$$\Sigma = \begin{bmatrix} \text{Var}(\mathbf{x}) & \text{Cov}(\mathbf{x}, \mathbf{y}) & \text{Cov}(\mathbf{x}, \mathbf{z}) \\ \text{Cov}(\mathbf{y}, \mathbf{x}) & \text{Var}(\mathbf{y}) & \text{Cov}(\mathbf{y}, \mathbf{z}) \\ \text{Cov}(\mathbf{z}, \mathbf{x}) & \text{Cov}(\mathbf{z}, \mathbf{y}) & \text{Var}(\mathbf{z}) \end{bmatrix} \quad (3.9)$$

Using matrix diagonalization, we can decompose Σ into L and R , and from here we can see that the grain aspect ratio can be calculated as:

$$a = \frac{l_{\min}}{l_{\max}} = \sqrt{\frac{\min(L)}{\max(L)}} = \sqrt{\frac{\lambda_{\min}(\Sigma)}{\lambda_{\max}(\Sigma)}} \quad (3.10)$$

The size of a single grain can be easily calculated by counting the total number of voxels n inside the grain and fit a sphere with equivalent diameter d to it:

$$d = 2 \left(\frac{3n}{4\pi} \right)^{\frac{1}{3}} \quad (3.11)$$

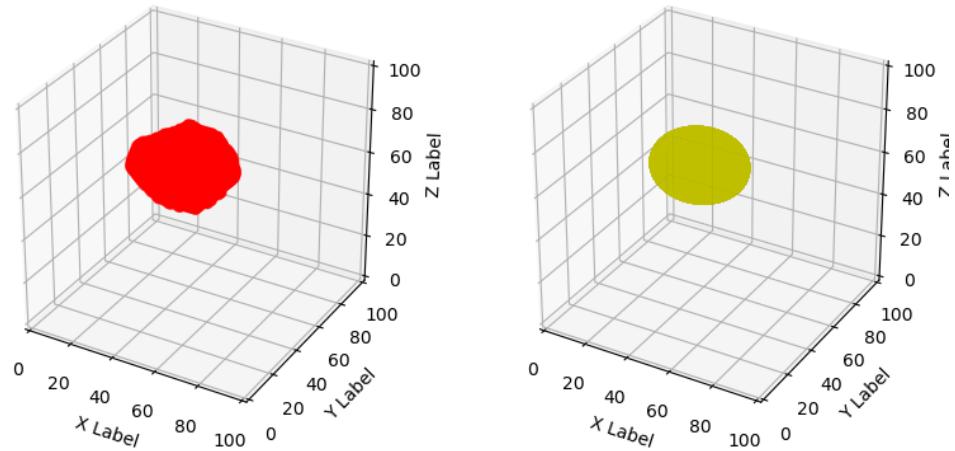


Figure 3.6. Fitting an ellipsoid to a 3D grain in the RVE.

Detail calculation is provided in the Appendix. After fitting all grains in the RVE with their corresponding ellipsoids, we would obtain the size (equivalent diameters) and shape (aspect ratios) distribution of the grains inside the RVE. Using the Maximum Likelihood Estimator we can fit the lognormal distribution to the RVE size distribution to obtain parameters μ and σ , and similarly we can fit the beta distribution to the RVE shape distribution to obtain parameters α and β . The final step is to compare the difference between RVE's distribution and the experimental distribution, and for that we will utilize *Hellinger distance* between two probability distributions. Specifically, if our experimental size distribution is lognormal with parameters μ_{exp} and σ_{exp} , then the difference between our RVE size data and the experimental size data is:

$$H_d = \sqrt{1 - \sqrt{\frac{2\sigma_{\text{exp}}\sigma}{\sigma_{\text{exp}}^2 + \sigma^2}} e^{-\frac{1}{4} \frac{(\mu_{\text{exp}} - \mu)^2}{\sigma_{\text{exp}}^2 + \sigma^2}}} \quad (3.12)$$

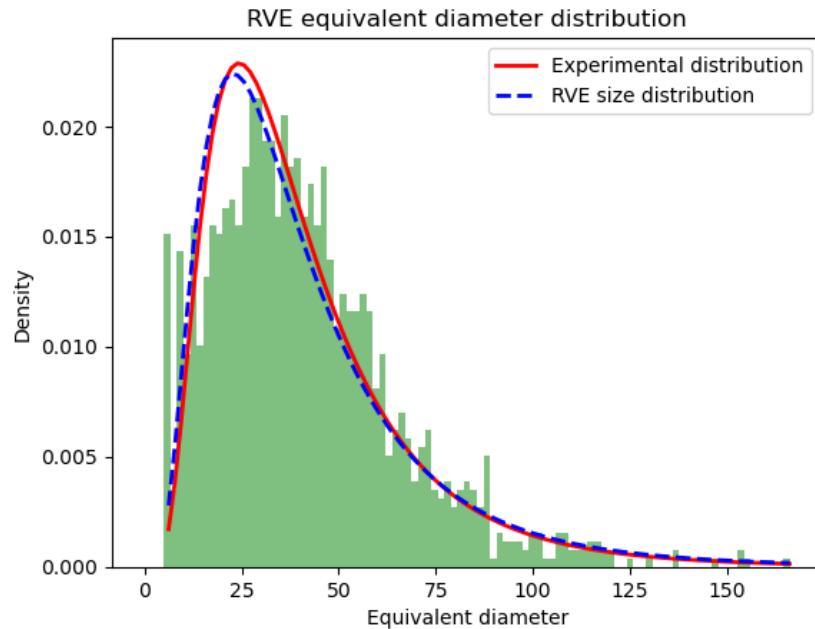


Figure 3.7. Grain size distribution of a generated single-phase RVE, plotted with experimental distribution.

The value ranges from 0 (being the same distribution) to 1 (d assigns probability zero to every set to which d_{exp} assigns a positive probability, and vice versa). Respectively, we can also find the Hellinger distance of beta distribution for the difference between our RVE shape data and the experimental shape data with parameters α_{exp} and β_{exp} :

$$H_a = \sqrt{1 - \frac{B\left(\frac{\alpha_{\text{exp}} + \alpha}{2}, \frac{\beta_{\text{exp}} + \beta}{2}\right)}{\sqrt{B(\alpha_{\text{exp}}, \beta_{\text{exp}})B(\alpha, \beta)}}} \quad (3.13)$$

where:

$$B(\alpha, \beta) = \int_0^1 t^{\alpha-1} (1-t)^{\beta-1} dt$$

is the beta function. Finally, the error function can be calculated by considering the size and shape distribution equally and obtain an arithmetic mean for the error as:

$$E_{\text{single}} = \frac{H_d + H_a}{2} \quad (3.14)$$

The value of E_{single} is bounded between 0 and 1, and during evaluation the RVE that yields the smallest value of E_{single} will be our chosen best RVE.

3.2 Pipeline extension to Q&P steel generation

The introduction of the pipeline to Q&P steel is not straightforward, since Q&P steel is a multi-phase material: different distributions are required for different phases, and the distribution of phases in the microstructure is also of great importance. Therefore, a separate process for accurately identifying different phases (namely ferrite, austenite, fresh martensite, tempered martensite with bainite) inside Q&P steel will be introduced.

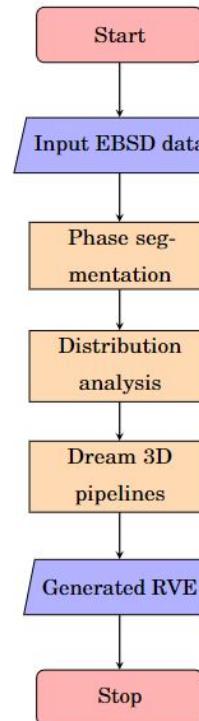


Figure 3.8. Generation workflow chart

3.2.1 Phase segmentation methods for Q&P steel

In order to accurately represent the steel microstructure, the model need to contain different phases inside the material. Generating different phases in RVE require the microstructure information from different phases. Hence, phases identification and segmentation is needed for generating the RVE.

The current phase segmentation criteria is presented in Table 3.1.

Table 3.1. Segmentation Criteria for Different Steel Phases

Phases	CS	IQ	GOS	KAM	Grain size
Austenite	FCC	-	-	-	Small
Ferrite	BCC	High	Low	Low	Large
Fresh Martensite	BCC	Low	Low	Low	Small
Tempered Martensite	BCC	High	High	High	Large
Bainite	BCC	Medium/High	High	High	Small

Having different properties, different phases of steel can be separated. Austenite can be detected as it have a FCC crystal structure, which is different from the other phases' BCC crystal structure. Because of the rough surface, fresh martensite can be identify having a lower image quality (IQ). Another variable can be consider in order to divide the phases are also grain orientation spread (GOS), detecting tempered martensite and bainite.

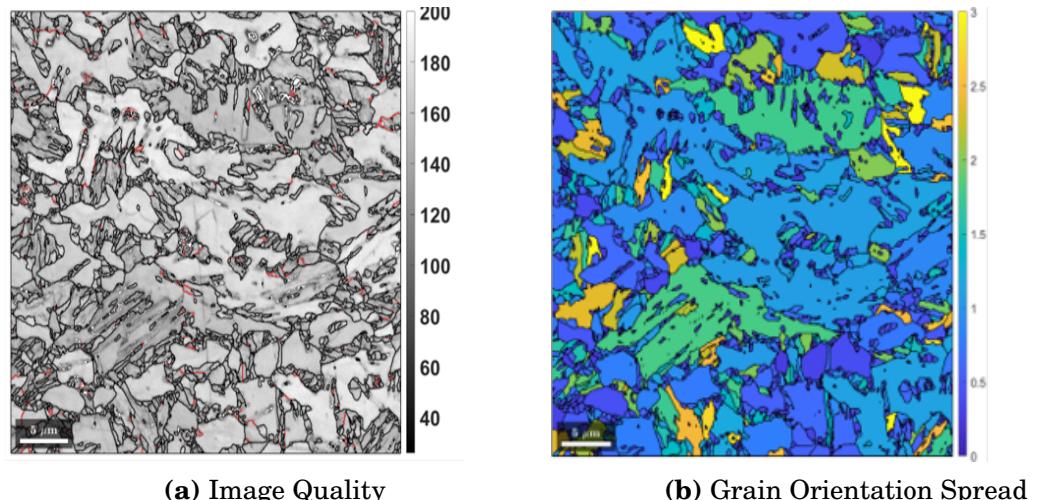


Figure 3.9. Phase separation using image preprocessing.

Reference values for segmenting the phase based on IQ and GOS were found with curve fitting. Grains that have lower IQ than the reference value were identified as fresh martensite. Grains with higher IQ, and lower GOS than reference value were distributed as ferrite. Austenite was

segmented due to its different crystal structure, thus remaining grains were selected as tempered martensite and bainite. EBSD files for each phases were exported during the process, and used to analyse distribution data.

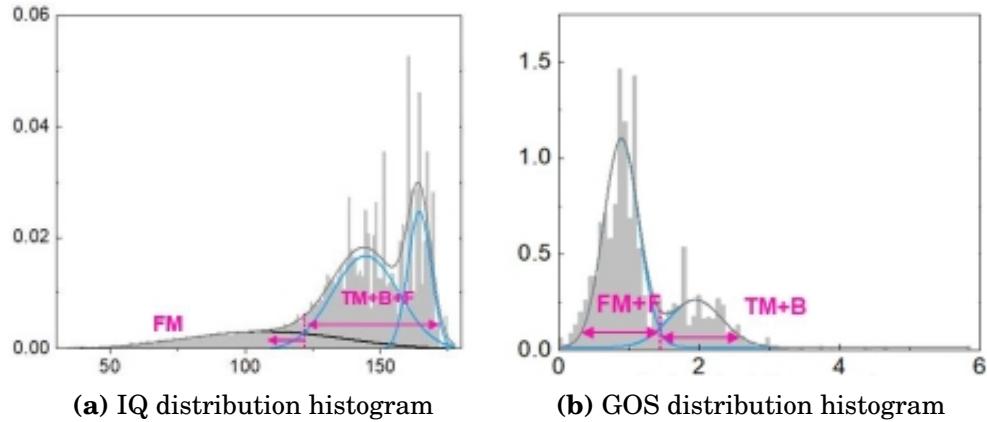


Figure 3.10. Bimodal Gaussian distribution of IQ and GOS for phase segmentation.

3.2.2 Multi-phase RVE generation

Using the EBSD files we exported earlier, microstructure analysis for each phase is completed. The result is presented in Table 3.2.

Table 3.2. Micro structure analysis of Q&P steel for generation

Phases	Phase Fraction	Average grain size, d	Grain size distribution, μ	Grain size distribution, σ	Grain shape factor, asp
Austenite	0.07	0.37	-1.1	0.47	1:0.3:0.21
Ferrite	0.31	3.8	1.1	0.69	1:0.5:0.3
Fresh Martensite	0.02	0.6	-0.64	0.5	1:0.37:0.29
Tempered Martensite/ Bainite	0.6	2.25	0.66	0.55	1:0.49:0.48

Major flow of pipelines is same with single-phase material, but we made four different phases representing austenite, fresh martensite, ferrite, and tempered martensite with bainite at the first step, *StatsGenerator*. Preset Statistic Models are set as rolled, which populates the statistic tabs with data that is designed to generate a rolled microstructure with an elongated grain shape factor we got. For shape distribution, α value and β value from the beta distribution of each phase are used. Due to a more complicated structure, generation time increases with bigger dimensions and resolution, smaller RVEs are generated for acceptable computing duration. An added

process for phase segmentation can be summarized as below:

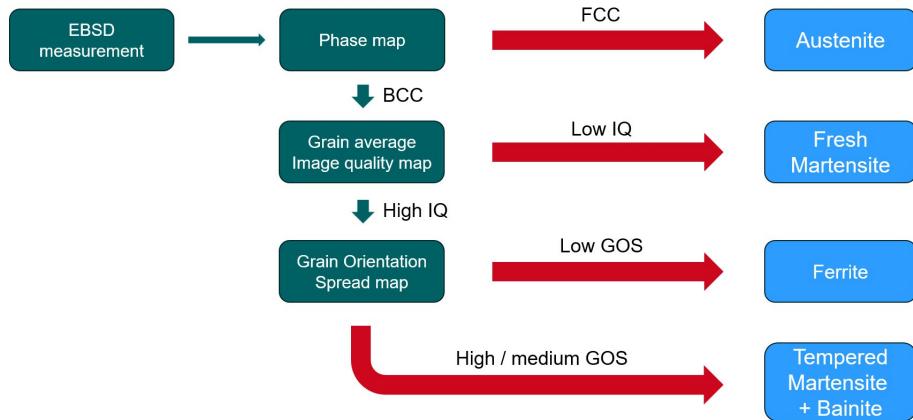


Figure 3.11. Overview of the extended phase segmentation process.

After initial running and if there are enough grains for every phase, pipelines are saved as a .json file and prepared for automatic generation under CSC. Here, CSC is an IT Center for Science Ltd. providing IT support and modeling, computing, and information services for academia, research institutes, and companies in Finland. We will use the supercomputing service, known as HPC (High Performance Computing) to run engineering simulations, such as DREAM3D, Matlab code, and Python script in our case.

3.3 Multi-phase RVE evaluation process

This section will describe the evaluation process for multi-phase material, or in our case, Q&P steel. The evaluation criteria is similar to the mono-phase, which are size, shape and texture. Because of multi-phase properties of the material, phase fraction also need to be considered. Furthermore, some extensions for the proposed fitting ellipsoids methods to multi-phase materials are also discussed.

Personal computer was used for evaluation with the following specification:

Operating system	Microsoft Window 11
Processor	Intel(R) Core(TM) i7-1355U, 1700 Mhz, 10 Core(s), 12 Logical Processor(s).
RAM	16.0 GB
System type	64-bit operating system, x64-based processor

3.3.1 Extension of slicing method to multi-phase RVEs

In order to process the multiphase RVEs, the phase numbers value from the CSV files and TXT files is utilized. Phase numbers is used for identification of each phase, including ferrite, austenite, fresh martensite and tempered martensite with bainite. As a result, different average values for size, shape and texture from different phases are collected.

In addition to previous evaluating criteria, phase fraction of each phase is needed to consider because of multiple phases. The volume of each grain is used in calculating the total volume of each phase, and then use to calculate the fraction volume of each phase. The proportional volume is then used for comparison with the experimental data. The average in deviation of each phase is use for evaluate the accuracy of the RVE.

3.3.2 Extension of fitting ellipsoids method to multi-phase RVEs

From the basis of evaluating single-phase RVEs by fitting ellipsoids, we can make an extension to further evaluate multi-phase materials, namely Q&P980 steel. The process is fairly straightforward as the only difference now is that we have to fit the ellipsoids to the grains in each phase separately, and from here we would obtain n different distribution for n -phase material. However, this extension requires us to consider two questions:

- How can we include several phases into the error function?
- How can we include the phase fraction, i.e. how much of each phase is in the material, in the error function?

We will first address the former question. Assuming that we want to evaluate our material with every phases of equal importance, that is the difference in one phase should not be treated differently from other phases.

Therefore, we can characterise this by using simple unweighted average (arithmetic mean): let H_d^i be the Hellinger distance from the RVE size distribution to the experimental distribution of phase i (as a convention, i will be called phase ID), then for a n -phase material the difference of the size distribution (equivalent diameters) can be evaluated as:

$$E_d = \frac{1}{n} \sum_{i=1}^n H_d^i = \frac{1}{n} \sum_{i=1}^n \sqrt{1 - \sqrt{\frac{2\sigma_{\text{exp},i}\sigma_i}{\sigma_{\text{exp},i}^2 + \sigma_i^2}} e^{-\frac{1}{4} \frac{(\mu_{\text{exp},i} - \mu_i)^2}{\sigma_{\text{exp},i}^2 + \sigma_i^2}}} \quad (3.15)$$

Similarly, for a n -phase material the difference of the shape distribution (grain aspect ratios) can be evaluated as:

$$E_a = \frac{1}{n} \sum_{i=1}^n H_a^i = \frac{1}{n} \sum_{i=1}^n \sqrt{1 - \frac{B\left(\frac{\alpha_{\text{exp},i} + \alpha_i}{2}, \frac{\beta_{\text{exp},i} + \beta_i}{2}\right)}{\sqrt{B(\alpha_{\text{exp},i}, \beta_{\text{exp},i})B(\alpha_i, \beta_i)}}} \quad (3.16)$$

Now we will move on to the second question: how can we include the phase fraction in the error function? Let f^{exp} be our normalized experimental phase fraction, and F be our normalized phase fraction obtained from the RVE data. Here, F is calculated by recording the phase ID each voxel and looping through every voxels in the RVE. Now, since both f^{exp} and F are distribution of the same size \mathbb{R}^n as both of them are normalized vectors of n phase fractions, we can calculate the difference between f^{exp} and F using *Hellinger distance for discrete distributions*:

$$E_f = H(f^{\text{exp}}, F) = \frac{1}{\sqrt{2}} \sqrt{\sum_{i=1}^n \left(\sqrt{f_i^{\text{exp}}} - \sqrt{F_i} \right)^2} \quad (3.17)$$

Since all values of H_d^i , H_a^i and $H(f^{\text{exp}}, F)$ are bounded between 0 and 1 by definition of Hellinger distance, it is guaranteed that E_d , E_a and E_f stays between 0 and 1. The final error value can thus be the arithmetic mean of our three errors as we want to evaluate all three features with equal weights:

$$E_{\text{multi}} = \frac{1}{3}(E_d + E_a + E_f) \quad (3.18)$$

The value of E_{multi} is bounded between 0 and 1, and in this model the best RVE would yield the smallest value of E_{multi} (having the best combination of fitted size, shape and phase distributions). For the time being, this method has not been extended to texture evaluation yet.

The code used in this manuscript is available at <https://github.com/dinhtrung21/fit-ellipsoid>.

4. Result

This part will cover the results of our work during the project. Section 4.1 will briefly explain the results we obtained while running the given pipeline on single-phase AISI439 stainless steel. Section 4.2 will go into the RVEs that we have generated using the extended pipeline to multi-phase materials, and finally section 4.3 will include the evaluation results of the RVEs from 4.2 using both slicing method and fitting ellipsoids method. Emphasis will be given on fitting ellipsoids method, and for both sections 4.2 and 4.3 some technical aspects of the algorithms will also be discussed.

4.1 Single-phase RVE generation and evaluation using current pipeline

First, we have the grains data collected using the AISI439 Stainless Steel EBSD analysis is shown in Table 4.1. Moreover, the texture file (ODF) of the material was generated and collected from DREAM3D.

Grain size distribution		Grain Shape Distribution	
μ	σ	α	β
3.5503	0.5996	3.1206	1.3618

Table 4.1. Grain data from AISI439

The RVEs was generated using different configuration consist of the resolution and dimension. The configuration was chosen such that there would be at least 1000 grain generated, and the reason for this can be illustrated in Figure 4.1: when the number of grains is too small, it would indicate that either the RVE has failed to capture small grains or it has captured a grain so big that it occupies most of the space inside that RVE and hence, lose the representativeness of the material. Another important note here is that the threshold of 1000 grains is chosen through trials and errors, and

a more quantitatively sensible choice of threshold would require further statistical analysis on the large scale.



Figure 4.1. Two RVEs with 319 grains and 1110 grains, respectively. It can be easily seen that the one on the left has failed to capture small grains.

The estimated number of grains inside RVEs of each configuration are presented in Table 4.2 below.

Table 4.2. 1. Estimated number of primary features (grains). Configurations with estimated grains smaller than 1000 are noted in italic.

Resolution	4	4.5	5	5.5	6
80 x 80 x 80	319	501	730	802	1207
100 x 100 x 100	685	851	1110	1813	2192
120 x 120 x 120	1082	1765	2132	2606	3487

These RVEs are then evaluated and compare to each other. The mean errors are presented in Table 4.3 below. According to the result, we find that the best configuration is dimension 100x100x100 and resolution 5.5.

Table 4.3. Error from the original file, %. The RVEs with numbers in italic does not satisfy the amount of grains

Resolution	4	4.5	5	5.5	6
80 x 80 x 80	1.16	1.68	3.15	1.50	4.28
100 x 100 x 100	4.26	3.43	4.27	2.72	3.40
120 x 120 x 120	3.36	2.95	4.04	2.88	2.87

The data was input to the pipeline in CSC which would utilize DREAM3D to generate the RVEs. In total 50 RVEs of this configuration was generated and evaluated using the slicing method for the project. The optimal RVE has the error of 1.906. The distribution of error in the different RVEs is shown in Figure 4.2.

From the data generated, we can also visualize our "best" RVE using ParaView, and the resultant visualization is shown in Figure 4.3. Notice that although this RVE has successfully evade the problems of small grains exclusion and over-sized grains inclusion, it is not guaranteed that this RVE will best represent our material due to the shortcomings addressed in part 3.1.2. This shortcoming can be mitigated later on in multi-phase material using fitting ellipsoids method discussed in part 3.1.3.

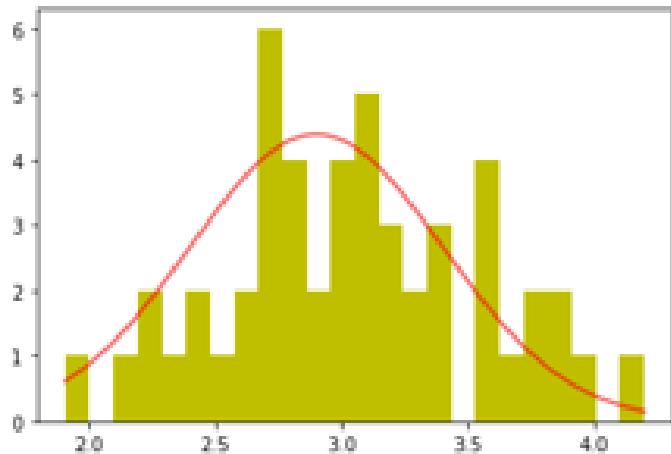


Figure 4.2. Error distribution of the generated RVEs. The error roughly centers at around 2.725770.

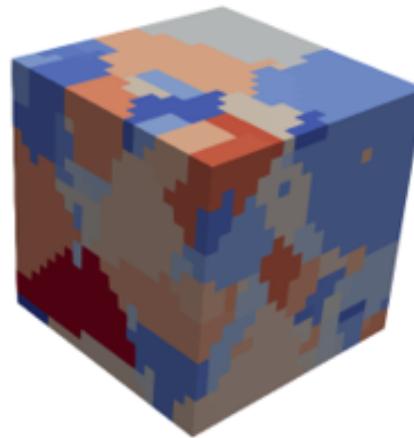


Figure 4.3. Corresponding best single-phase RVE from the above generation.

4.2 Multi-phase RVE generation

Here, we will discuss first the time and space complexity of our extended pipeline, and then the results of the generation will be presented. Generation testing environment is as below.

Operating system	Microsoft Window 10
Processor	Intel(R) Core(TM) i7-10750H CPU @ 2.60GHz, 6 Core(s)
RAM	32.0 GB
System type	64-bit operating system, x64-based processor

4.2.1 RVE generation output

Below figures are from one of the generated RVE through CSC system with dimension as 400x400x400 and resolution as 0.05x0.05x0.05. About 10000 grains are included in this RVE, from much smaller one to bigger one.

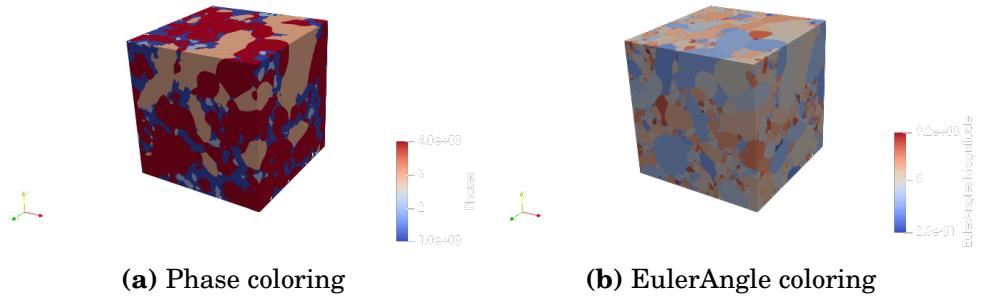


Figure 4.4. Visualization of the best RVE generated.

According to the data we have got from iterating generation and evaluation of single RVE with different dimensions, we found that dimension should be equal or higher than 150 to get meaningful result. Our variance of dimension set from 150 to 350, with interval of 50 at this point. Total 50 grains, 10 RVEs each for 5 different dimensions, were generated under CSC. Around 3 hours are required to generate, and around 7 GB of storage are needed.

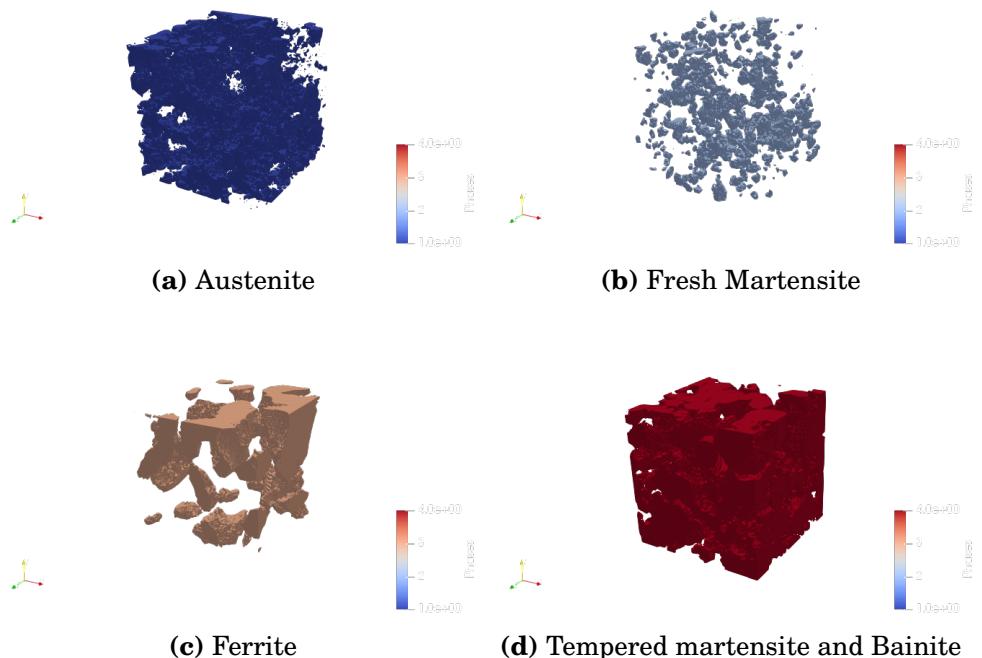


Figure 4.5. Phase maps: coloring based on phase

When Euler angles were used to coloring reference, we found that very small grains are also included in the RVE.

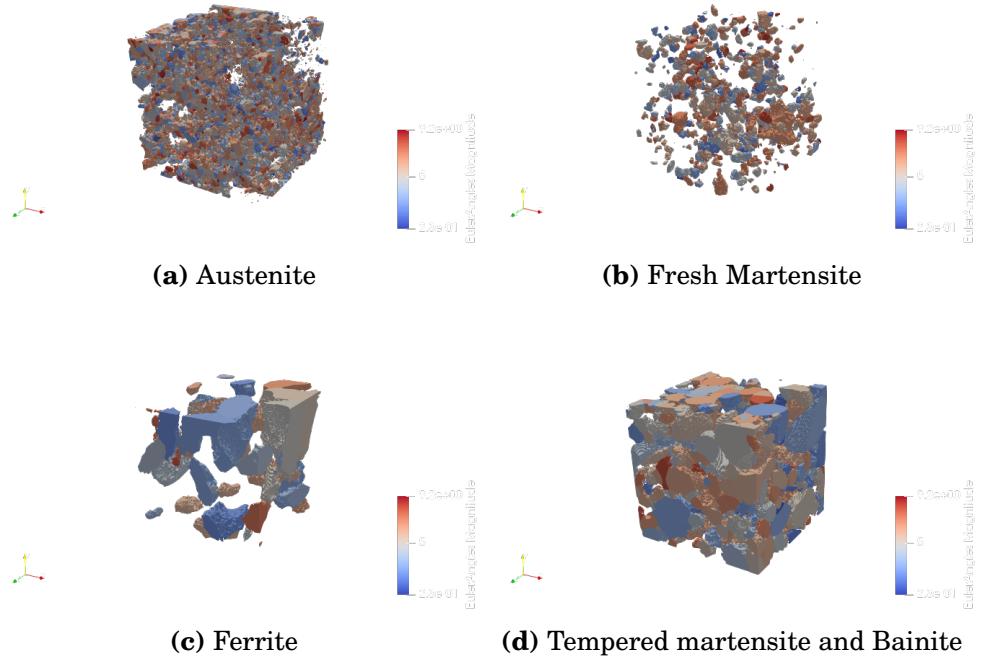


Figure 4.6. Phase maps: coloring based on Euler angle

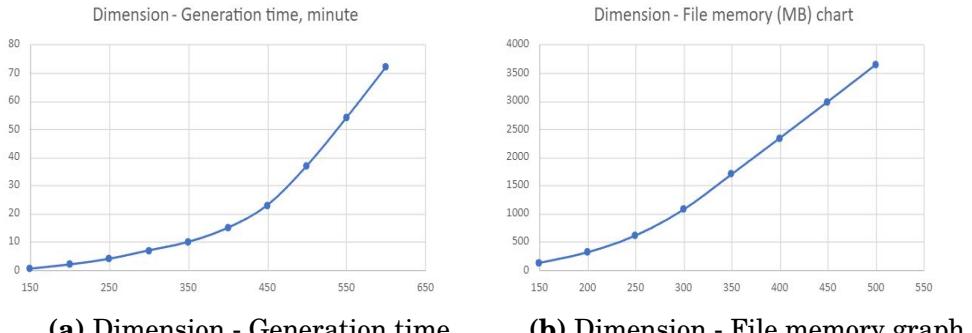
4.2.2 Time and space complexity

To handle the storage and duration for generation of multiple RVEs, time consumed and file size of each RVE with different dimension were recorded to find applicable dimension. Above figures present the relationships. It showed that the time required for higher dimension grows in $O(n^3)$, where n is dimension. These relationships prove the need for CSC usage and implementation.

Additionally, the fact that evaluation process took longer than generation process was also taken into account when deciding the dimensions. At this point, size of files were considered as well, since we needed to download all of required files in our evaluating environment.

4.3 Multi-phase RVE evaluation

Fifty multi-phase RVEs were evaluated using the two methods mentioned above. Because texture was unable to implement in the generation, it was not included in evaluation. Hence, only size, shape and phase fraction was

**Figure 4.7.** Time and space complexity

considered. The determining error is the average of these three aspect error.

4.3.1 RVE evaluation using slicing method

This method was not usable for the generated RVEs due to the format in the CSV file become unstable, making this method inefficient to tackle the problem. As the result, further modification of the script is needed, or the generation pipeline need to be studied in order to resolve the problem.

```

3159,0.88074073,0.080251523,0.14774171,0.43058303,2.50130,2.1503427,2.7207000,8,2,1,0.055250003
3160,0.12557346,-0.2141951,0.95069081,0.13170134,0.66663456,0.50183046,2.7477205,15,2,1,0.25387502
3161,-0.49310568,-0.3752574,-0.73736554,0.26892525,1.8715857,1.336604,0.57056755,5,2,1,0.19100001
3162,0.0091244038,-0.025160961,-0.99925554,0.027786056,1.0408727,1.466427,2.3764503,6,2,0,0.030750001
3163,0,0,0,1,4.694931,1.7508942,1.754315,23,2,0,1.064375
3164,-0.92339349,-0.32580695,0.081591383,0.18584163,6.2086849,2.7328253,5.5302858,17,2,1,0.65850002
3165,0.2850416,-0.57655299,-0.15714072,0.74942958,2.2366314,1.3972507,4.4599261,16,2,1,0.46112502
Feature_ID,NumNeighbors,NeighborList
1,5,48,633,1671,2792,2828
2,1,2799
3,8,133,797,1424,2780,2798,2804,2808,2911
4,6,16,479,847,2314,2463,2803
5,13,354,374,382,860,1308,1601,1630,2376,2549,2876,2893,3008,3032
6,9,51,2083,2123,2799,3003
7,9,615,1341,1387,2068,2147,2610,2819,2925,3058
8,12,19,102,146,546,1381,1910,2454,2494,2562,2696,2835,3080

```

Figure 4.8. The abrupt change in the CSV file stop the pipeline from working

4.3.2 RVE evaluation using fitting ellipsoid method

The evaluation process identified the RVE that has the lowest derivation from experimental data is an RVE with dimension 300x300x300 and resolution 0.05, with the error is 0.43127. The error from each RVE is provide in Table 4.9 and the comparison between the RVE and the experimental data is provide in the Figure 4.13. There are two things that worth noting here. The first one this that the generated RVE indeed follows the lognormal distribution for equivalent diameter and beta distribution for aspect ratio, indicating that Dream3D algorithms indeed follows the type

of distribution given. The second thing to note here is that although we can see better fitting for size distribution whereas the fitting is off for shape distribution, it will be later shown that the estimated errors are roughly similar in both cases. In other words, the beta fitting works well despite of the misleading graphs.

Dimension	Error, Δ										
150x150x150	0.48904	0.47646	0.48665	0.46857	0.47039	0.43150	0.50036	*	*	*	*
200x200x200	*	0.55788	0.49533	*	0.43439	*	0.47665	0.43943	0.46423	0.51270	
250x250x250	0.46958	0.49067	0.48096	0.46429	0.49183	0.49122	0.46613	0.49795	0.47923	0.50585	
300x300x300	0.45302	0.46086	0.45308	0.45803	0.45708	0.45436	0.48020	0.44052	0.45238	0.43127	
350x350x350	0.45437	0.46723	0.43151	0.47661	0.43691	0.44938	0.43648	0.43656	0.45879	0.44756	

Figure 4.9. Error calculate using fitting ellipsoid method. Red error means the best RVE and * means the error doesn't converge

Using the average error can be misleading due to the distribution between grains from different phases. Therefore, the distribution for the best RVE from each remaining dimension is provided in the Appendix. We are also interested in the error distribution for each features. The error for dimension 300 and 350 RVEs are summarised in Figure 4.10. The error is consistent at about 0.28, relatively low for both size and shape distribution. It is hence proven that both size and shape distribution of the RVEs are well-evaluated.

However, when it comes to phase fraction, the errors are abnormally high. The reason behind the high value of error for phase fraction is because in the experimental data, some phase is infinitesimal, therefore a small error in generation would scale up largely. This indicate a need for a new formulation for fraction error.

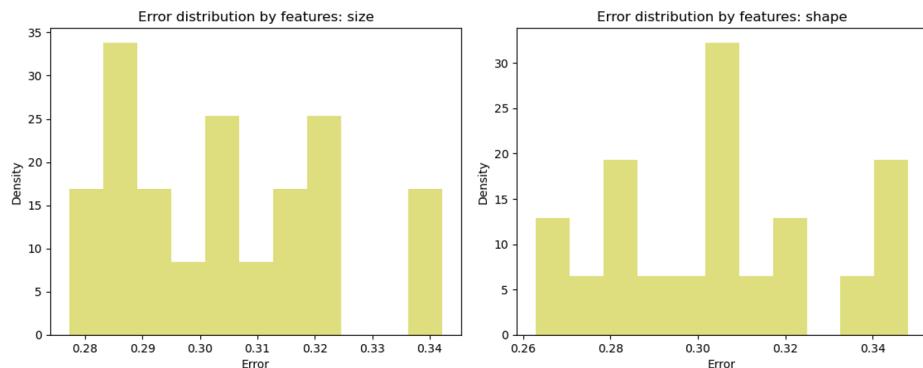


Figure 4.10. Error distribution of size and shape of the generated multi-phase RVEs

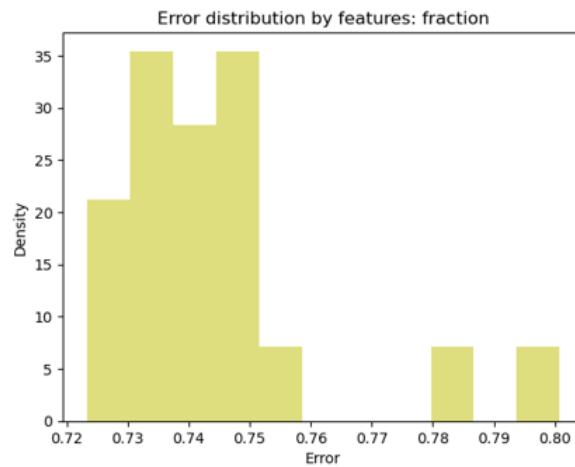


Figure 4.11. Fraction error distribution of the generated multi-phase RVEs

4.3.3 Time complexity

Similar to generation, it takes a tremendous amount of time to run the evaluation. As the file become larger, the processing time scale mirror to the amount of data. However, the time for each RVE evaluation in the same dimension can also varies, as it is based on the number of grains within the RVEs. The processing time for each dimension is provide in Figure 4.12.

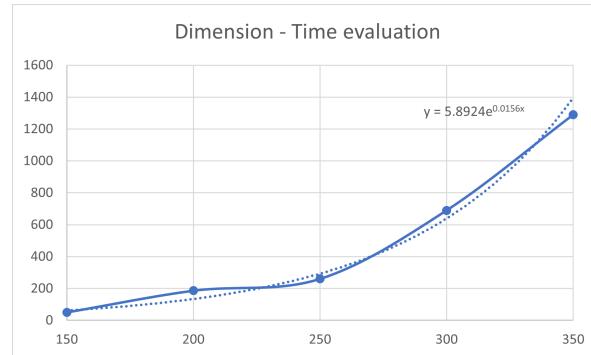


Figure 4.12. Dimension - time relationship in evaluation

The dimension-time relationship can be fit using the exponential relationship, suggesting higher dimension or higher number of grains would take immense computational power and time.

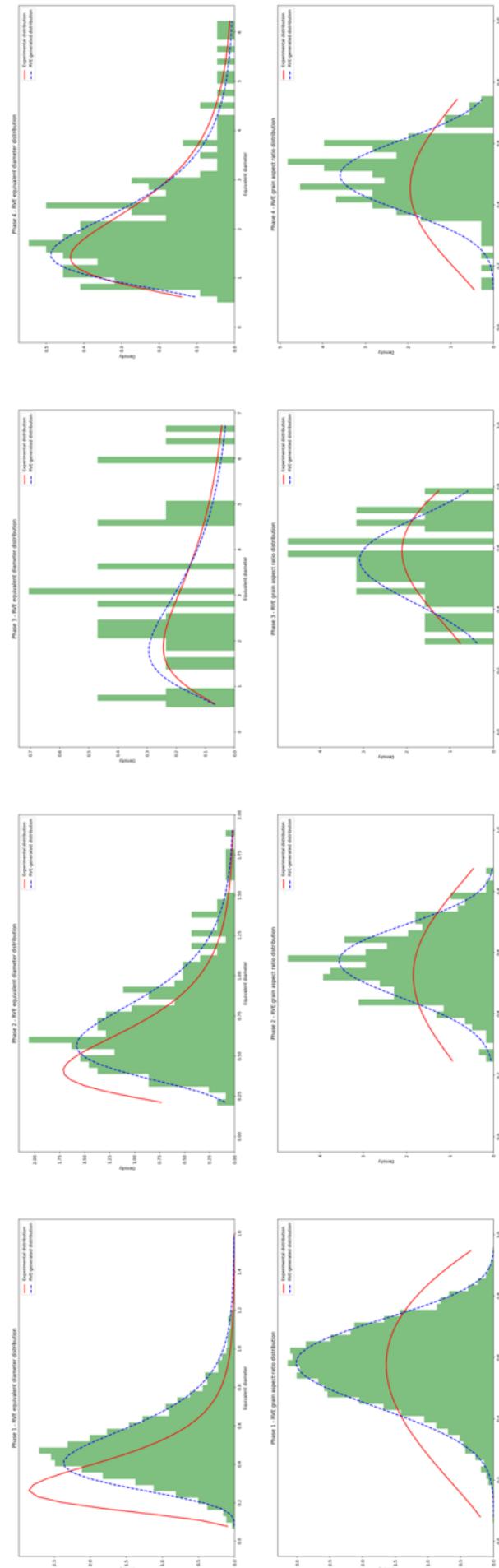


Figure 4.13. Comparison between generated and experimental data for different phases. Dimension 300^3 . The upper row includes size distribution, and the lower row includes shape distribution. The phases are in the order: austenite, fresh martensite, ferrite, tempered martensite and bainite.

5. Discussion

In this part, we will discuss about the results, the applicability of the methods given, and possible next steps in the research of RVEs. For RVE generation, scalability and feature inclusion is the main problem to be discussed, and alternatives such as deep learning will be mentioned; and for RVE evaluation, some current shortcomings of the fitting ellipsoids method will be taken into account, and we will propose several potential resolution to the shortcomings as well as ways that we can further utilize this method.

5.1 Multi-phase RVE generation discussion

Overall, the extended multi-phase RVE generation pipeline has addressed the most basic problem of this project: to include several phases into generation and produce RVEs of good quality based on the results of the evaluation. Three of the most important features are included in this pipeline, that is equivalent diameters, grain aspect ratios, and phase fraction. However, this pipeline also have some disadvantages:

- **Failure of other features inclusion:** this is probably the biggest drawback of this methods, as well as the main issue that our project hasn't been able to address properly. Characterizing materials require more information, such as orientation, misorientation, ODF, MDF, and tilt angles. In the case of orientation, the failure to transition from single-phase to multi-phase is due to the complexity of separately analyze the orientation of different phases inside the material. The reason behind this failure to include features mainly stems from the fact that currently EBSD analysis is mainly carried

on MTEX, and MTEX output is not fairly straightforward to be given as Dream3D inputs. Therefore, the most prominent solution yet is to separately analyze the feature to get the inputs Dream3D requires.

- **Phase segmentation:** phase segmentation is an important step of the pipeline, and the main issue with the current segmentation method is its high dependency on "human judgment". Most of the parameters are set by manually evaluating the quality of the images, and this is prone to subjecting to bias compared to other more quantitative methods. Another main problem is the failure to segment tempered martensite and bainite, leading to this project working with 4 phases instead of 5 phases as the nature of the material.
- **Autonomy:** currently there is no unified pipeline that can take in the EBSD scan of a material and automatically giving the RVE output. Many subprocesses are required for this pipeline, including phase segmentation, EBSD analysis, pipeline generation, and pushing the pipeline on CSC.
- **Randomness:** the way that this pipeline picks out the best RVE from several generated RVEs is essentially randomly generate a large number of RVEs and evaluate for the best one. Hence, the choice of best RVE depends largely on the randomness of the generation process rather than the optimization method.
- **Scalability:** as discussed in the results, the running time of the generation scaled in $O(n^3)$ for dimension n even with the help of supercomputing platforms. This is an inherent problem of the whole concepts, so for the time being, mitigation is a more possible resolution than a complete solution.

From the drawbacks, we can easily see that one main reason is our dependency on Dream3D generation algorithm. Therefore, scientists have already looked for alternative generating methods, and one prominent candidate on this race is machine learning models. Specifically, in the field of deep learning, models such as Generative Adversarial Network (GAN) [21] have shown great potential in generating RVEs of high quality with less computational time due to low dependency on material statistics. However, this can also be the source of problems, since GAN models work on image processing principles rather than that of materials, therefore suitable evaluation is also needed to couple with these models.

5.2 Multi-phase RVE evaluation discussion

One of the main output of this report is the fitting ellipsoids evaluation scheme. This method has addressed the main issues with the current slicing method and propose a reliable mean to evaluate 3D aspect ratio of the grains. However, this method also have some problems that need to be resolved:

- **Beta fitting convergence:** although beta fitting is used to analyze the aspect ratio, this is an issue more dependent on the size characteristic of the material. Specifically, when the grain size is too small then that grain would occupy a single voxel in the RVE, leading to the algorithm fitting into a perfect sphere with aspect ratio equal to 1. This can also happen to a perfectly spherical grain (although rare), and when this happen, SciPy fails to fit a beta distribution to data inclusive of 0 and 1. Removing these grains can mitigate the problem, but the better solution is to choose the resolution of the RVE suitably, so that the smaller grains can be captured in the right shape instead of collapsing into single voxels.
- **Error function for phase fraction:** the choice of discrete Hellinger distance for phase fraction has shown to be unsuitable in the result part. Other simple alternatives include Euclidian distance (Eq. 5.1) and dot product (Eq. 5.2) for normalized phase fraction vectors.

$$(p, q) = \sqrt{\sum_{i=1}^n (p_i - q_i)^2} = \sqrt{(p_1 - q_1)^2 + \cdots + (p_n - q_n)^2} \quad (5.1)$$

$$\mathbf{a} \cdot \mathbf{b} = \sum_{i=1}^n a_i b_i = a_1 b_1 + a_2 b_2 + \cdots + a_n b_n \quad (5.2)$$

However, the problem with the first method is that we can only analyze if the phase fractions are very similar (or the same), and it would fail to characterize the relationship between the Euclidian distance and the distribution difference, i.e. will the error be larger when the distributions are more different? Similarly, dot product can characterize low difference (small angles), but does the two distribution really be the most different when the phase fraction vectors are orthonormal (dot product equal to 0)? This is a question that would require substantially more studies.

- **Degree of errors:** this issue is more of a evaluation design for

researchers' intuition. Although from the graph we can see that an error of approximately 0.3 corresponds to a relatively well-fitted distribution, it can easily be mistaken as 30% error (which is not a good number for an evaluation scheme). This can be mitigated easily by using the squared Hellinger distance, which basically is the Hellinger distance without the square root.

Despite the mentioned shortcomings, the fitting ellipsoids method has also proven itself to be a powerful and potential tool for evaluation. The biggest potential of this method lies in the diagonalization of the dispersion matrix of the grains:

$$\Sigma = R^T L R$$

Here, R is the rotational matrix of the grain, and with proper decomposition we can break down R into three improper Euler angles (demonstrated in the Appendix). The Euler angles then can be utilized to evaluate many other features, namely orientation and tilt angles. Finally, since this process is conducted in Python and not MATLAB (no dependency on MTEX), it can be easily merged in with the generation part of the pipeline for combined parallel evaluation on CSC system, improving both the running time and the quantities of RVEs evaluated.

6. Self-evaluation

This part will be a self-evaluation by each member of our group on the project workload, the new things we learnt and our reflection on this project as a whole journey.

6.1 Ben Nguyen

Working on the project was quite an experience for me. I have had working on the thesis during the summer with a topic that related to material science, yet the topic of the thesis is completely different. The topic this time was more related to the computational aspect, which is fine, considering the course is the bachelor project. However, I find the options for the topic was quite narrow as it was mostly based on material science. I understand that it is because the responsible fall on the Material Science department, yet I would like to see more diverse topic, coming from the other courses as well, i.e. solid mechanics or fluid mechanics.

Our topic was quite interesting for me, as its background was introduce in the MSE course in previous year. The topic was interesting and important, and I find that it was understandable on the objective of the project. I find that the potential exploration on this topic was high, and initially there was a lot of different way to complete it. Yet setting the direction vaguely wasn't ideal, as it was overwhelming on which direction we could go. The resource was good in my opinion, and having the files set up wasn't too much of a difficult task. The learning process was good and allowed us to understand what were the aspect needed for evaluation; running the code and understand the code was more difficult, as some part isn't clear and hence modifying them ran to a few problem that was later resolved. All and

all, I think the workload for the course is approximately 10 credits for what we have delivered. Had the initial plan carried out, solving misorientation or implement a new phase segmentation, the course would worth much more than 10 credits.

For my personal thought on the course, I think it was quite difficult. I was working mainly on reading about the steel and doing the evaluation task using the slicing method and the fitting ellipsoid method. I greatly enjoy the fact that we have an advisor, Rongfei, to provide us with instruction on what to do in a enthusiastic manner. Moreover, the course is project based with the concrete frame for timeline, making it flexible in time that needed. I'm happy that I was able to contributed to the project, and I owed it to my teammates, Trung and Byeongmin, for helping me and providing explanation when I needed. The course was able to provide me insight about the study path for my upcoming Master program. It was a difficult journey and I'm glad to see it has ended nicely.

6.2 Byeongmin Oh

The project was an impressive journey of three months. I remember the first day of project course where the topics are introduced. For me, who was used to just taking set classes and doing assignments, a project without a set answer felt like walking through a fog. In that sense, I think I chose the topic very well. The concepts used in MSE, which was one of the classes I studied the most over the past two years, were expanded, and the programs used were ones I had used before, so it wasn't difficult to adapt. Of course, it was not without difficulties. I did not have as much knowledge in this field as the other members, and there was also a simple hardware problem. However, as time passed, I became more and more immersed in the topic, and as I solved each problem one by one, I gained the motivation to move forward, and got a result at the end.

In terms of content, I think there are more things to be disappointed about. As someone who was mainly responsible for creating the RVE, the most disappointing thing is that I was unable to apply the texture and misorientation of Q&P steel. If I conduct research on a similar topic in the future, I hope to achieve better results by including the parts that were not completed this time. Outside of the topic, it seems that my overall

understanding of materials as well as areas related to simple projects has expanded.

Navigating through this course, I found it quite challenging. What added immense value to the experience was having an enthusiastic advisor, Rongfei, who provided clear instructions. The project-based nature of the course, coupled with a well-defined timeline, allowed for flexibility in managing time effectively. I take pride in my contribution to the project and owe a debt of gratitude to my teammates, Trung and Ben, for their teamwork, support, and explanations.

6.3 Trung Nguyen

Beside from my bachelor's thesis, I would say this project is the highlight of this academic year of mine. It is not everyday that I get to experience participating in a 10-ECTS course and working with a team on a topic that is actually emerging and relevant both in academia and research industries. During the project, I get to learn many new concepts and tools, bond over with my classmates, and most important of all I get to know better the career path that I want to pursue in the future.

Regarding the project topic on RVE generation and evaluation, it is especially relevant to me as at that point I was also doing some similar research on RVE construction at work. Working on the project indeed gave me more insight on RVEs and its application, as well as the material science industry as a whole. The experience was therefore beneficial to both my studies and my work outside academia. Two of the things that really stick with me after the project are my experience working with CSC, and finding a whole new method for evaluating RVEs' 3D data. Regarding the first task, it was not an easy responsibility learning how to run code on a supercomputing system from square one, and it took me nearly two weeks just to be comfortable navigating about the pipeline. Nevertheless, it was a worthwhile experience having my first touch with HPC and shell scripts. On the finding a new evaluation method, it was the fitting ellipsoids method discussed during the project, and honestly creating a whole new GitHub repository from scratch is, so far, my proudest programming achievement up-to-date. Organizing a repository and version control has always been useful skills that I want to learn during my studies, and the

fact that I managed to be proficient in both of them in a timespan of 4 weeks is the best proof that after the project I am, although still having many things to learn, a much more prolific student and researcher than I was at the beginning of the course.

Overall, I appreciate the journey I took working on this project. I want to send my gratitude to professor Junhe for guiding us through such an important milestone of our degree, to Rongfei for her unwavering support both academically and mentally in the past 3 months, and finally I want to say thank you to Ben and Aaron for being the best teammates I can have during my bachelor studies. This has been a challenging period, and I am grateful that I went to this period with the best people that I can possibly ask for.

Reference

- [1] X. Chen, C. Niu, C. Lian, and J. Lin, “The evaluation of formability of the 3rd generation advanced high strength steels qp980 based on digital image correlation method,” *Procedia Engineering*, vol. 207, pp. 556–561, 2017. International Conference on the Technology of Plasticity, ICTP 2017, 17-22 September 2017, Cambridge, United Kingdom.
- [2] M. Zhou, Y. Li, Q. Hu, X. Li, and J. Chen, “Investigations on edge quality and its effect on tensile property and fracture patterns of qp980,” *Journal of Manufacturing Processes*, vol. 37, pp. 509–518, 2019.
- [3] S. Bargmann, B. Klusemann, J. Markmann, J. E. Schnabel, K. Schneider, C. Soyarslan, and J. Wilmers, “Generation of 3d representative volume elements for heterogeneous materials: A review,” *Progress in Materials Science*, vol. 96, pp. 322–384, 2018.
- [4] M. Gljušić, M. Franulović, D. Lanc, and A. Žerovnik, “Representative volume element for microscale analysis of additively manufactured composites,” *Additive Manufacturing*, vol. 56, p. 102902, 2022.
- [5] P. Ferretti, G. M. Santi, C. Leon-Cardenas, E. Fusari, G. Donnici, and L. Frizziero, “Representative volume element (rve) analysis for mechanical characterization of fused deposition modeled components,” *Polymers*, vol. 13, no. 20, 2021.
- [6] D. Morin, L. E. B. Dæhli, J. Faleskog, and O. S. Hopperstad, “A numerical study on the effect of porosity distribution on ductile failure using size-dependent finite element-based representative volume elements,” *European Journal of Mechanics - A/Solids*, vol. 101, p. 105051, 2023.
- [7] X. Hu, X. Sun, L. Hector, and Y. Ren, “Individual phase constitutive properties of a trip-assisted qp980 steel from a combined synchrotron x-ray diffraction and crystal plasticity approach,” *Acta Materialia*, vol. 132, pp. 230–244, 2017.
- [8] I. Gitman, H. Askes, and L. Sluys, “Representative volume: Existence and size determination,” *Engineering Fracture Mechanics*, vol. 74, no. 16, pp. 2518–2534, 2007.
- [9] A. J. Wilkinson and T. B. Britton, “Strains, planes, and ebsd in materials science,” *Materials Today*, vol. 15, no. 9, pp. 366–376, 2012.

- [10] Outokumpu, “Moda 439/4510.” <https://www.outokumpu.com/en/products/product-ranges/moda/moda-439-4510>.
- [11] Metalcor, “1.4510 (aisi439) data sheet.” <https://www.metalcor.de/en/datenblatt/26/>.
- [12] Z. Li, Y. Chang, J. Rong, J. Min, and J. Lian, “Edge fracture of the first and third-generation high-strength steels: Dp1000 and qp1000,” *IOP Conference Series: Materials Science and Engineering*, vol. 1284, p. 012021, 06 2023.
- [13] L. Wang and J. G. Speer, “Quenching and partitioning steel heat treatment,” *Metallography, Microstructure, and Analysis*, 2013.
- [14] “Quenching and partitioning martensite—a novel steel heat treatment,” *Materials Science and Engineering: A*, vol. 438-440, pp. 25–34, 2006. Proceedings of the International Conference on Martensitic Transformations.
- [15] W. Callister and D. Rethwisch, *Materials Science and Engineering: An Introduction, 8th Edition*. Wiley, 2009.
- [16] M. A. Groeber and M. A. Jackson, “Dream.3d: A digital representation environment for the analysis of microstructure in 3d,” *Integrating Materials and Manufacturing Innovation*, vol. 3, p. 56–72, Apr. 2014.
- [17] M. GROEBER, “A framework for automated analysis and simulation of 3d polycrystalline microstructures. part 2: Synthetic structure generation,” *Acta Materialia*, vol. 56, p. 1274–1287, Apr. 2008.
- [18] F. Azhari, W. Davids, H. Chen, S. P. Ringer, C. Wallbrink, Z. Sterjovski, B. R. Crawford, D. Agius, C. H. Wang, and G. Schaffer, “A comparison of statistically equivalent and realistic microstructural representative volume elements for crystal plasticity models,” *Integrating Materials and Manufacturing Innovation*, vol. 11, pp. 214–229, Mar. 2022.
- [19] K. F. Mulchrone and K. R. Choudhury, “Fitting an ellipse to an arbitrary shape: implications for strain analysis,” *Journal of Structural Geology*, vol. 26, no. 1, pp. 143–153, 2004.
- [20] C. Chatfield and A. J. Collins, *Introduction to Multivariate Analysis*. Routledge, Feb. 2018.
- [21] F. Pütz, M. Henrich, N. Fehlemann, A. Roth, and S. Münstermann, “Generating input data for microstructure modelling: A deep learning approach using generative adversarial networks,” *Materials*, vol. 13, p. 4236, Sept. 2020.
- [22] G. G. Slabaugh, “Computing euler angles from a rotation matrix,” Aug. 1999.

A. Fitting an ellipsoid to an arbitrary three-dimensional shape

A.1 General equation of an ellipsoid in three dimensions

From basic geometry, the equation of the locus of a point $\mathbf{p}' = (x', y', z')$ on the boundary of an ellipsoid, with centre $\mu' = (\mu'_x, \mu'_y, \mu'_z)$, and l_1, l_2, l_3 being the length of the semi-axes, respectively, and major axis parallel to the x' -axis, is given by:

$$\frac{(x' - \mu'_x)^2}{l_1^2} + \frac{(y' - \mu'_y)^2}{l_2^2} + \frac{(z' - \mu'_z)^2}{l_3^2} = 1 \quad (1.1)$$

In general, the major axis orientation can be represented by 3D elemental rotation, that is rotation about x-, y-, or z-axis using the right-hand rule. Specifically:

- Rotation by an angle θ about the x-axis using the right-hand rule:

$$R_x(\theta) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos(\theta) & -\sin(\theta) \\ 0 & \sin(\theta) & \cos(\theta) \end{bmatrix}$$

- Rotation by an angle θ about the y-axis using the right-hand rule:

$$R_y(\theta) = \begin{bmatrix} \cos(\theta) & 0 & \sin(\theta) \\ 0 & 1 & 0 \\ -\sin(\theta) & 0 & \cos(\theta) \end{bmatrix}$$

- Rotation by an angle θ about the z-axis using the right-hand rule:

$$R_z(\theta) = \begin{bmatrix} \cos(\theta) & -\sin(\theta) & 0 \\ \sin(\theta) & \cos(\theta) & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

In our contexts, we are interested in *Euler angles*, and using matrix multiplication we can represent an extrinsic rotation whose (improper) Euler angles are α , β , and γ :

$$\begin{aligned} R &= R_z(\gamma)R_y(\beta)R_x(\alpha) = \begin{bmatrix} \cos \gamma & -\sin \gamma & 0 \\ \sin \gamma & \cos \gamma & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \cos \beta & 0 & \sin \beta \\ 0 & 1 & 0 \\ -\sin \beta & 0 & \cos \beta \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos \alpha & -\sin \alpha \\ 0 & \sin \alpha & \cos \alpha \end{bmatrix} \\ &= \begin{bmatrix} \cos \beta \cos \gamma & \sin \alpha \sin \beta \cos \gamma - \cos \alpha \sin \gamma & \cos \alpha \sin \beta \cos \gamma + \sin \alpha \sin \gamma \\ \cos \beta \sin \gamma & \sin \alpha \sin \beta \sin \gamma + \cos \alpha \cos \gamma & \cos \alpha \sin \beta \sin \gamma - \sin \alpha \cos \gamma \\ -\sin \beta & \sin \alpha \cos \beta & \cos \alpha \cos \beta \end{bmatrix} \end{aligned}$$

Therefore, from the reference coordinate $\mathbf{p}' = (x', y', z')$, the rotation produces new variables $\mathbf{p} = (x, y, z)$ as:

$$\mathbf{p} = R\mathbf{p}' \quad (1.2)$$

Now, notice that the equation (1.1) can be written in this format as follows:

$$(\mathbf{p}' - \mu')^T L^{-1} (\mathbf{p}' - \mu') = \begin{bmatrix} x' - \mu'_x & y' - \mu'_y & z' - \mu'_z \end{bmatrix} \begin{bmatrix} l_1^2 & 0 & 0 \\ 0 & l_2^2 & 0 \\ 0 & 0 & l_3^2 \end{bmatrix}^{-1} \begin{bmatrix} x' - \mu'_x \\ y' - \mu'_y \\ z' - \mu'_z \end{bmatrix} = 1$$

The choice of axis length matrix L will be shown shortly after. Now, the change of variables can be introduced using the matrix form of the rotation transformation in Eq. (2) $\mathbf{p}' - \mu' = R^{-1}(\mathbf{p} - \mu)$, yielding:

$$(\mathbf{p} - \mu)^T (R^{-1})^T L^{-1} R^{-1} (\mathbf{p} - \mu) = 1$$

Here, $\mu = (\mu_x, \mu_y, \mu_z) = R\mu'$ are the rotated coordinates of the ellipsoid centre. Let $\Sigma = R^T L R$, we get the general equation of an ellipsoid in three dimensions (in matrix notation) as:

$$(\mathbf{p} - \mu)^T \Sigma^{-1} (\mathbf{p} - \mu) = 1 \quad (1.3)$$

A.2 Fitting an ellipsoid to a 3D region identified by points

In this section a method for fitting an ellipsoid E to a given region R containing n points $\mathbf{p}_i = (x_i, y_i, z_i)$, $i = 1, 2, \dots, n$ is derived. We define our error function to be a variant of the squared Euclidean distance:

$$\mathbf{d}(\mathbf{p}, E) = \sum_{i=1}^n (\mathbf{p}_i - \mu)^T \Sigma^{-1} (\mathbf{p}_i - \mu) + \frac{n}{2} \log |\Sigma| \quad (1.4)$$

We can see that letting $\Sigma = \mathbf{I}$ would give us the well-known least-square error function. Geometrically, $\mathbf{d}(\mathbf{p}, E)$ can be viewed as the Euclidean distance computed on variables transformed in such a way that the ellipsoid would appear as a sphere in the transformed space. We can also see that the solution to minimising $\mathbf{d}(\mathbf{p}, E)$ is the same as the estimation of the mean vector μ and dispersion matrix Σ of a trivariate Gaussian distribution, given a sample of n data points. The solution is given as:

- Mean:

$$\bar{\mu} = [\mu_x \quad \mu_y \quad \mu_z]^T = \left[\frac{1}{n} \sum_{i=1}^n x_i \quad \frac{1}{n} \sum_{i=1}^n y_i \quad \frac{1}{n} \sum_{i=1}^n z_i \right]^T \quad (1.5)$$

- Dispersion matrix:

$$\bar{\Sigma} = \frac{1}{n} \sum_{i=1}^n (\mathbf{p}_i - \bar{\mu})(\mathbf{p}_i - \bar{\mu})^T \quad (1.6)$$

which can be further expanded into:

$$\begin{aligned} & \begin{bmatrix} \frac{1}{n} \sum_{i=1}^n (x_i - \mu_x)^2 & \frac{1}{n} \sum_{i=1}^n (x_i - \mu_x)(y_i - \mu_y) & \frac{1}{n} \sum_{i=1}^n (x_i - \mu_x)(z_i - \mu_z) \\ \frac{1}{n} \sum_{i=1}^n (y_i - \mu_y)(x_i - \mu_x) & \frac{1}{n} \sum_{i=1}^n (y_i - \mu_y)^2 & \frac{1}{n} \sum_{i=1}^n (y_i - \mu_y)(z_i - \mu_z) \\ \frac{1}{n} \sum_{i=1}^n (z_i - \mu_z)(x_i - \mu_x) & \frac{1}{n} \sum_{i=1}^n (z_i - \mu_z)(y_i - \mu_y) & \frac{1}{n} \sum_{i=1}^n (z_i - \mu_z)^2 \end{bmatrix} \\ &= \begin{bmatrix} \text{Var}(\mathbf{x}) & \text{Cov}(\mathbf{x}, \mathbf{y}) & \text{Cov}(\mathbf{x}, \mathbf{z}) \\ \text{Cov}(\mathbf{y}, \mathbf{x}) & \text{Var}(\mathbf{y}) & \text{Cov}(\mathbf{y}, \mathbf{z}) \\ \text{Cov}(\mathbf{z}, \mathbf{x}) & \text{Cov}(\mathbf{z}, \mathbf{y}) & \text{Var}(\mathbf{z}) \end{bmatrix} \quad (1.7) \end{aligned}$$

This result implies that the best fitting ellipsoid (in the sense of minimising distance $\mathbf{d}(\mathbf{p}, E)$) has a center vector $\mu = \bar{\mu}$ and dispersion matrix $\Sigma = \bar{\Sigma}$. At this stage, it remains to identify the axis length parameters l_1, l_2, l_3 , and Euler angles α, β, γ that correspond to this closest fit. This can be

done by making the substitution $\Sigma = R^T L R$ and get:

$$R^T L R = \bar{\Sigma} \quad (1.8)$$

Now, we can see that the rotational matrix R is orthonormal, therefore $R^T = R^{-1}$. This leads to Eq. (8) be the diagonalization of dispersion matrix $\bar{\Sigma}$, therefore the semi-major axis length of our ellipsoid will be $\sqrt{\lambda_{\max}(\bar{\Sigma})}$, and the semi-minor axis length will be $\sqrt{\lambda_{\min}(\bar{\Sigma})}$. However, this is the ellipsoid that minimizes the error function given above with correct aspect ratios, and to get an ellipsoid that best "covers" our grain, we want to scale it with equivalent volume: $\text{vol}(E) = n$. Notice that:

$$\text{vol}(E) = \frac{4}{3}\pi \prod_{i=1}^3 l_i = \frac{4}{3}\pi \prod_{i=1}^3 \sqrt{\lambda_i(\bar{\Sigma})} = \frac{4}{3}\pi \prod_{i=1}^3 \sqrt{L_{ii}} = \frac{4}{3}\pi \sqrt{\det(L)}$$

The last equation is from the fact that L is a diagonal matrix. Therefore, the correct scaling for our 1D axes is $\left[\frac{n}{\text{vol}(E)}\right]^{\frac{1}{3}}$, and our semi-major and semi-minor axes are:

$$\max(l_i) = \left[\frac{3n}{4\pi\sqrt{\det(L)}} \right]^{\frac{1}{3}} \sqrt{\max(L_{ii})} \quad (1.9)$$

$$\min(l_i) = \left[\frac{3n}{4\pi\sqrt{\det(L)}} \right]^{\frac{1}{3}} \sqrt{\min(L_{ii})} \quad (1.10)$$

Finally, the Euler angles α, β, γ can be extracted from R as following [22]:

$$\alpha = \text{atan2}(R_{32}, R_{33}) \quad (1.11)$$

$$\beta = \text{atan2}(-R_{31}, \sqrt{R_{32}^2 + R_{33}^2}) \quad (1.12)$$

$$\gamma = \text{atan2}(R_{21}, R_{11}) \quad (1.13)$$

where `atan2` is the 2-argument arctangent.

B. Multi-phase evaluation using fitting ellipsoids method: full results

This part includes the result of the multi-phase RVE generation. It should be noted here that:

- The left column includes size distribution, and the right column includes shape distribution. The phases are in the order: austenite, fresh martensite, ferrite, tempered martensite and bainite.
- Only the result of the best RVE for each configuration is included.
- For the data of shape distribution, a filter has been applied so that only grains with aspect ratios strictly smaller than 1.0 are included in the analysis.

The data of this evaluation is available upon request. For a small demo of this method you can find instructions at <https://github.com/dinhtrung21/fit-ellipsoid>.

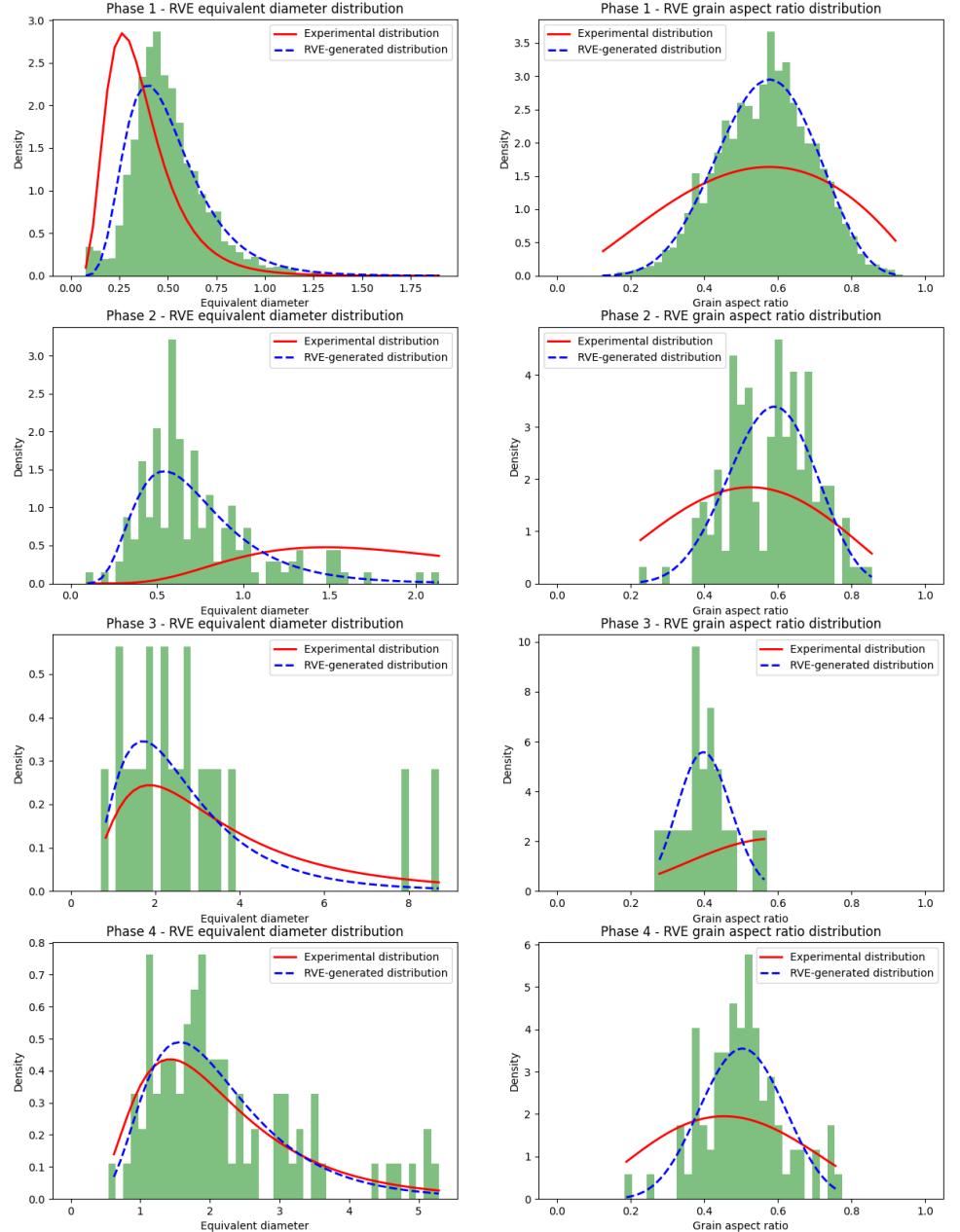


Figure 2.1. Comparison between generated and experimental data for different phases.
Dimension 200^3

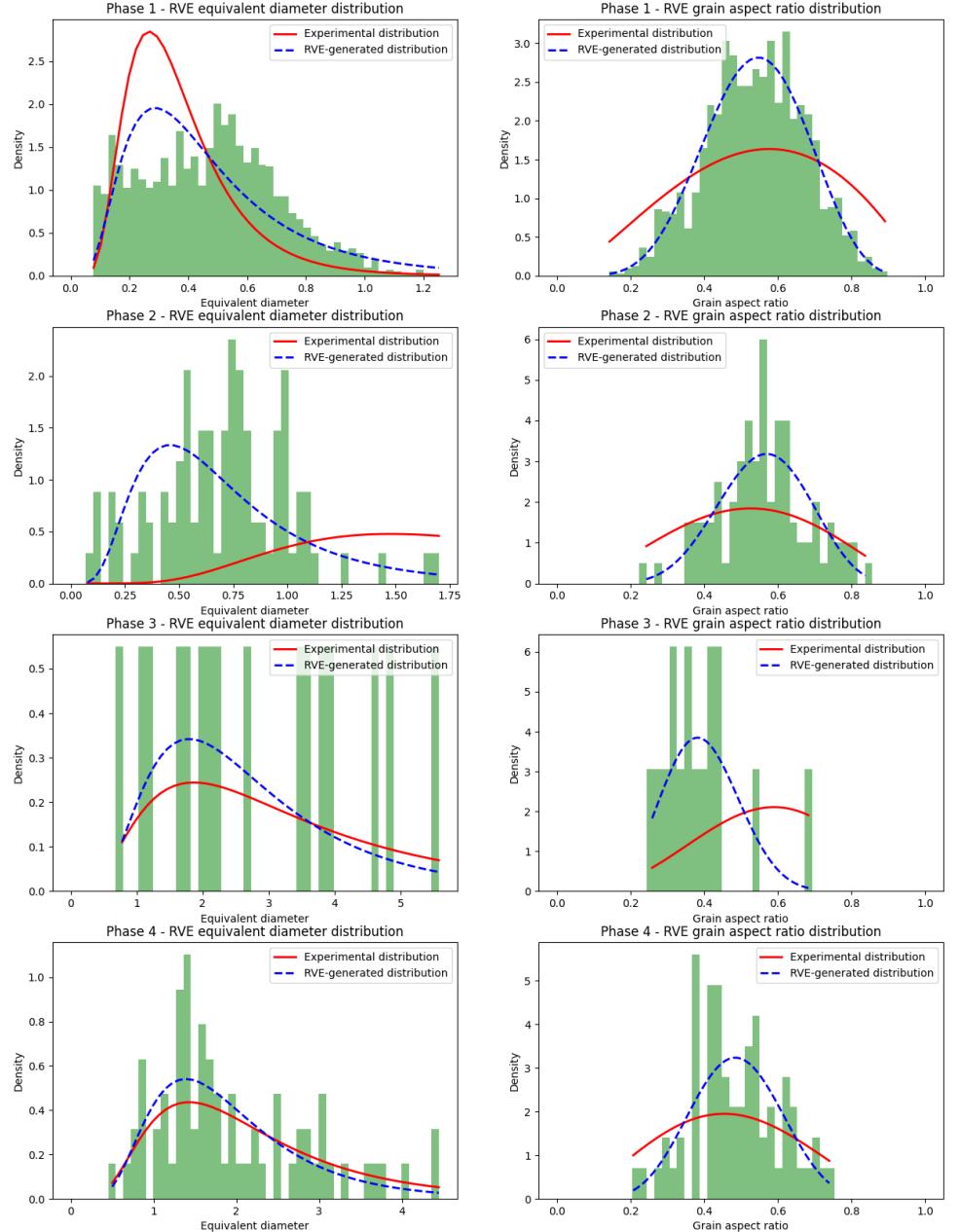


Figure 2.2. Comparison between generated and experimental data for different phases.
Dimension 250^3

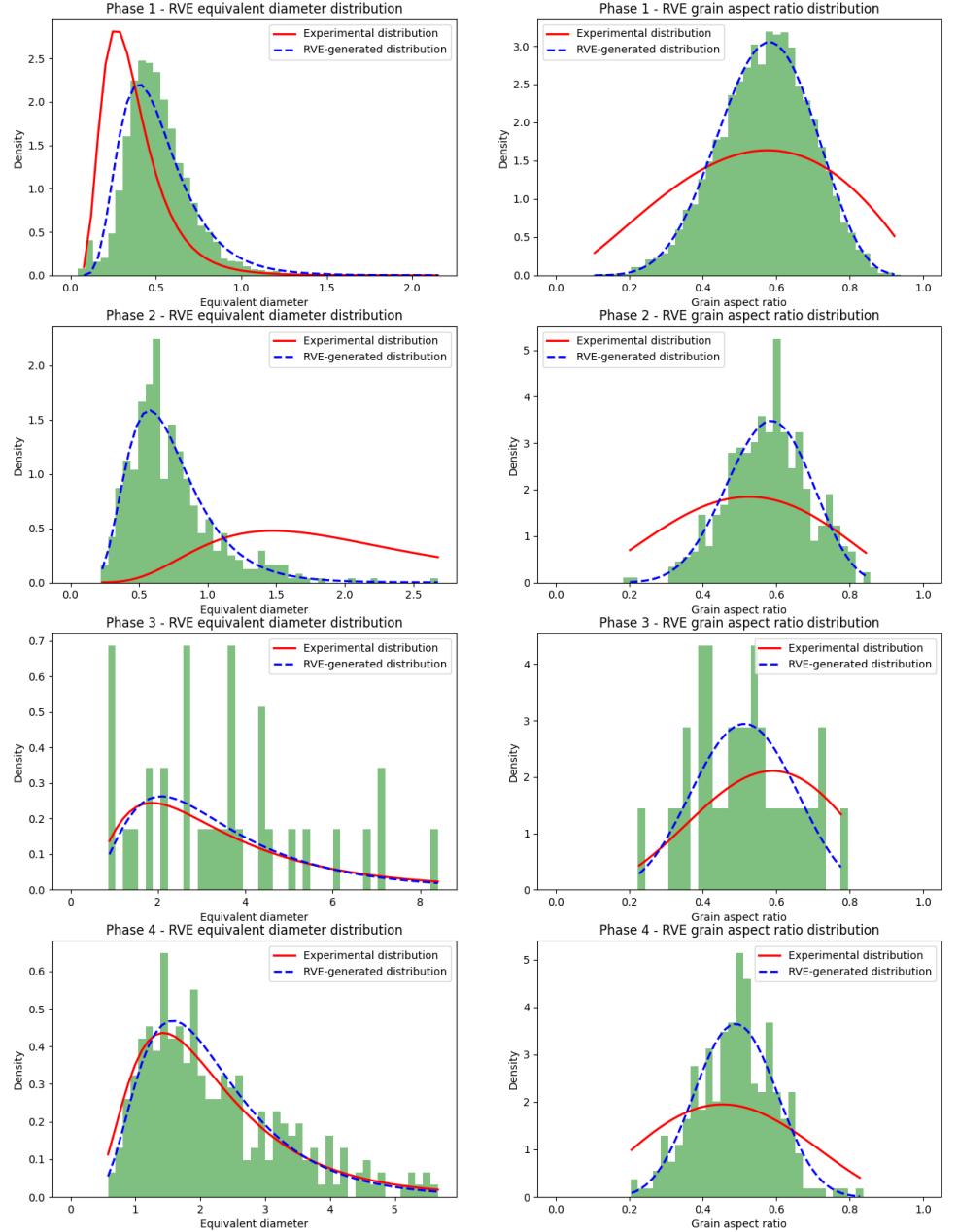


Figure 2.3. Comparison between generated and experimental data for different phases.
Dimension 350^3