

Gaussian Process Regression in Atom Probe Tomography Data Reconstruction

Teemu Turpeinen

School of Science

Bachelor's thesis
Espoo 22.11.2024

Supervisor

Prof. Nuutti Hyvönen

Advisor

MSc Aslam Shaikh



Copyright © 2024 Teemu Turpeinen

The document can be stored and made available to the public on the open internet pages of Aalto University.
All other rights are reserved.

Author Teemu Turpeinen

Title Gaussian Process Regression in Atom Probe Tomography Data Reconstruction

Degree programme Bachelor's Programme in Science and Technology

Major Mathematics and System Sciences

Code of major SCI3029

Teacher in charge Prof. Nuutti Hyvönen

Advisor MSc Aslam Shaikh

Date 22.11.2024

Number of pages 27+5

Language English

Abstract

Atom probe tomography (APT) is an imaging technique that provides precise spatial and compositional mapping of materials. APT is based on evaporating specimens inside a strong electric field onto a detector. The data gathered on the detector is used to reconstruct the spatial and compositional structure of the specimen. Recent development of microelectronics and nanomaterials has produced an increase in the use of APT for applications, but complex field evaporation events and materials still pose a challenge for traditional reconstruction techniques. These techniques are based on assuming linear evaporation events and therefore do not take into account more complex evaporation events or grain structures.

In this thesis, an alternative reconstruction technique was considered in the form of Gaussian process regression (GPR). GPR is a supervised machine learning method that is based on Bayesian inference and thus well suited for tasks where there is inherent uncertainty. This property makes GPR suited for the uncertain evaporation events that occur in evaporating specimens onto a detector in APT. After a thorough theoretical overview of GPR, this method was applied in reconstructing crystallographic specimens with different grain structures. In order to form a complete understanding of the capabilities of GPR, two GPR models with different kernels were considered in the reconstruction of both specimens. These specimens were field evaporated using a simulation model that simulates conjectured field evaporation events. This enabled the direct comparison of the reconstruction with the original specimen structure.

It was found that GPR can be used to reconstruct the general composition of specimens with varying complexity, while the precise reconstruction of grain structures is still limited to uniform grain specimens. These results show that with further refinement of the method, GPR could be utilized more in APT data reconstruction.

Keywords Atom Probe Tomography, Crystallography, Gaussian Process, Kernel

Tekijä Teemu Turpeinen**Työn nimi** Gaussisia prosesseja hyödyntävä regressiomenetelmä APT-datan rekonstruktiossa**Koulutusohjelma** Teknistieteellinen kandidaattiohjelma**Pääaine** Matematiikka ja systeemitieteen**Pääaineen koodi** SCI3029**Vastuuopettaja** Prof. Nuutti Hyvönen**Työn ohjaaja** FM Aslam Shaikh**Päivämäärä** 22.11.2024**Sivumäärä** 27+5**Kieli** Englanti**Tiivistelmä**

APT (Atom Probe Tomography) on kuvantamismenetelmä, joka mahdollistaa tarkan materiaalien avaruudellisen ja koostumuksellisen karttoittamisen. APT perustuu näytteiden hajoittamiseen voimakkaassa sähkökentässä ilmaisimelle, jolle kertynyt data mahdollistaa näytteiden avaruudellisen ja koostumuksellisen rakenteen jälleenrakentamisen. Viimeikainen mikroelektronikan ja nanomateriaalien kehittyminen on kasvattanut APT:n käyttöä käytännön sovelluksissa, mutta monimutkaiset haihtumistapahtumat ja materiaalit aiheuttavat haasteita perinteisille rekonstruktio menetelmille. Nämä menetelmät perustuvat haihtumistapahtumiien lineaariseen projisointiin ja näin ollen ne eivät ota huomioon monimutkaisempia haihtumistapahtumia tai hilarakenteita.

Tässä opinnäytetyössä tutkitaan vaihtoehtoisena rekonstruktio menetelmänä regresiomalleja, jotka hyödyntyvät gaussisia prosesseja, eli GPR-malleja. GPR on ohjattu koneoppimismenetelmä, joka perustuu bayesilaiseen päättelyyn ja soveltuu näin tehtäviin, joissa esiintyy luontaista epävarmuutta. Tämä ominaisuus tekee GPR:stä hyvin soveltuvan epävarmoihin haihtumistapahtumiin, joita esiintyy APT:ssa näytteiden haihtuessa ilmaisimelle. Työssä hyödynnettiin GPR:ää eri kiderakenteisten näytteiden jälleenrakentamisessa. Molempien näytteiden jälleenrakentamisessa käytettiin kahta GPR-mallia eri kerneleillä, jotta kokonaiskuva GPR-mallien tehokkuudesta olisi mahdollisimman kattava. Tutkimuksen näytteet hajoitettiin simulaatiomallilla, joka simuloi spekuloituja haihtumistapahtumia. Tämä mahdollisti jälleenrakennetun ja alkuperäisen näytteen rakenteiden välittömän vertailun.

Tutkimuksen tuloksena havaittiin, että GPR-menetelmää voidaan käyttää eri kiderakenteisten näytteiden yleisen koostumuksen rekonstruktiossa, mutta tarkka hilarakenteiden rekonstruktio on vielä rajoittunut yksinkertaisiin hilarakenteisiin. Näiden tulosten perusteella GPR-menetelmää kehittämällä voitaisiin sitä hyödyntää enemmän APT-datan rekonstruktiossa.

Avainsanat APT, Kristallografia, Gaussinen prosessi, Kerneli

Contents

Abstract	3
Abstract (in Finnish)	4
Contents	5
1 Introduction	6
2 Theoretical background	7
2.1 Atom Probe Tomography	7
2.1.1 Imaging Setup	7
2.1.2 Data reconstruction	7
2.1.3 Post-processing	8
2.2 Gaussian Process Regression	10
2.2.1 Multivariate Gaussian distribution	10
2.2.2 Gaussian Processes	10
2.2.3 Kernels and hyperparameter estimation	12
3 Methodology	15
3.1 Forward model	15
3.2 Data and feature selection	15
3.3 Reconstruction models	17
4 Results	19
4.1 Uniform Data Set	19
4.1.1 Metrics and Predictions	19
4.1.2 Crystallography	21
4.2 Grain Boundary Data Set	22
4.2.1 Metrics and Predictions	22
4.2.2 Crystallography	23
5 Discussion and Conclusions	25
A Mathematical Background	28
A.1 Matrix identities	28
A.2 Multivariate Gaussian distribution	28
B Figures and Tables	30
B.1 Figures	30
B.2 Tables	30
C Source Code	32

1 Introduction

Atom probe tomography (APT) is an imaging technique that combines precise sub-nanometer imaging of materials with simultaneous compositional analysis [1]. APT provides a solution to one of the main challenges of material science: understanding the properties of materials at an atomic scale. This level of understanding has become increasingly essential in the development of microelectronics, and APT has seen important application in the development of semiconductors [2].

In APT, a specimen of interest is evaporated onto a detector by the method of field evaporation. The resulting data gathered on the detector is utilized in the spatial and chemical reconstruction of the specimen. Reconstructing the three-dimensional (3D) lattice of the evaporated specimen is a central part in determining the precision and robustness of APT. This has historically been done with geometrical algorithms inherited from field ion microscopy (FIM). These algorithms make a number of assumptions and perform poorly when challenged with complex materials and field evaporation events. Insights into field evaporation physics [3] have proposed that the process of field evaporation contains complex probabilistic phenomena. This has led to the development of simulation models that aim to simulate theorized evaporation events. Consequently, new methods of data reconstruction are required.

The aim of this thesis is to test and assess the performance of Gaussian process regression (GPR) in APT data reconstruction, while providing a thorough mathematical review of GPR. GPR provides a highly flexible and probabilistic method that can capture, in theory, the uncertainty inherent in field evaporation events. The performance of GPR is estimated by performing data reconstruction for two crystallographic specimens with different grain structures, where both specimens have been evaporated with a simulation model that incorporates theorized field evaporation events.

In order to form a comprehensive view of the potential of GPR in APT data reconstruction, two different reconstruction models are implemented with different kernel structures. This enables the comparison of reconstruction performance and guides future development of GPR models for APT data reconstruction.

The structure of the thesis is as follows. First, we provide a general overview of APT and a comprehensive review of GPR in section 2. In section 3, we give a detailed description of the data sets and the reconstruction models and frameworks used in the analysis. Finally, section 4 shows the reconstruction results and performance between data sets, while section 5 discusses future research and the feasibility of GPR as a new reconstruction technique.

2 Theoretical background

2.1 Atom Probe Tomography

In APT, the goal is to determine both the distribution and composition of atoms within a material, which in turn gives insights into the different properties exhibited by the material. This section will aim to give an overview of APT, from the general imaging setup to data reconstruction and common post-processing methods.

2.1.1 Imaging Setup

In APT, the specimen of interest is cooled to temperatures in the range of 20-80K and shaped into the form of a needle with a typical tip radius below 100 nm [1]. The specimen is then placed inside an ultra-high vacuum chamber in preparation for actual measurements. These preparation steps serve multiple purposes. The low temperature reduces thermal motion within the atomic lattice, which especially reduces the lateral motion of tip atoms. The specimen's shape concentrates the electric field, enabling the field evaporation of surface atoms. Finally, the ultra-high vacuum chamber prevents atomic interactions in the flight-paths of atoms.

In field evaporation, the specimen is exposed to a strong electric field of around 30-60 V/nm, which causes the breakdown of atomic bonds at the surface of the specimen [4]. This results in the ionization of the surface atoms, which leave the specimen and fly out onto a detector. For field evaporation to occur, the electric field is charged close to an evaporation threshold, after which the field evaporation events can happen. Then by pulsing the electric field over the threshold, atom-by-atom field evaporation is achieved. This precise method enables the measurement of flight-times for each positively charged ion and consequently facilitates the performance of mass spectrometry on the specimen. It is notable that this is often a thermally assisted process, where the voltage pulses are accompanied by laser pulses on the specimen.

The data gathered on the detector includes the hitting position coordinates and the hitting time. By combining detector information with information gathered from the electric field and voltage pulses, we end up with flight-times, masses, orders of arrival and other information regarding the current measurement settings. In figure 1, one can observe an illustration of a basic atom probe imaging set up.

2.1.2 Data reconstruction

A crucial part of APT is the task of forming a 3D reconstruction of the original specimen based on data gathered from the detector. Traditional reconstruction techniques have been purely based on geometry and stem from early methods used in FIM [4, p. 184]. These techniques include algorithms such as the zeroth order algorithm and the wide field of view algorithm that make basic assumptions which enable data reconstruction. These assumptions include nearly straight flight-paths of evaporated atoms and that the surface of the specimen remains spherical throughout the evaporation process.

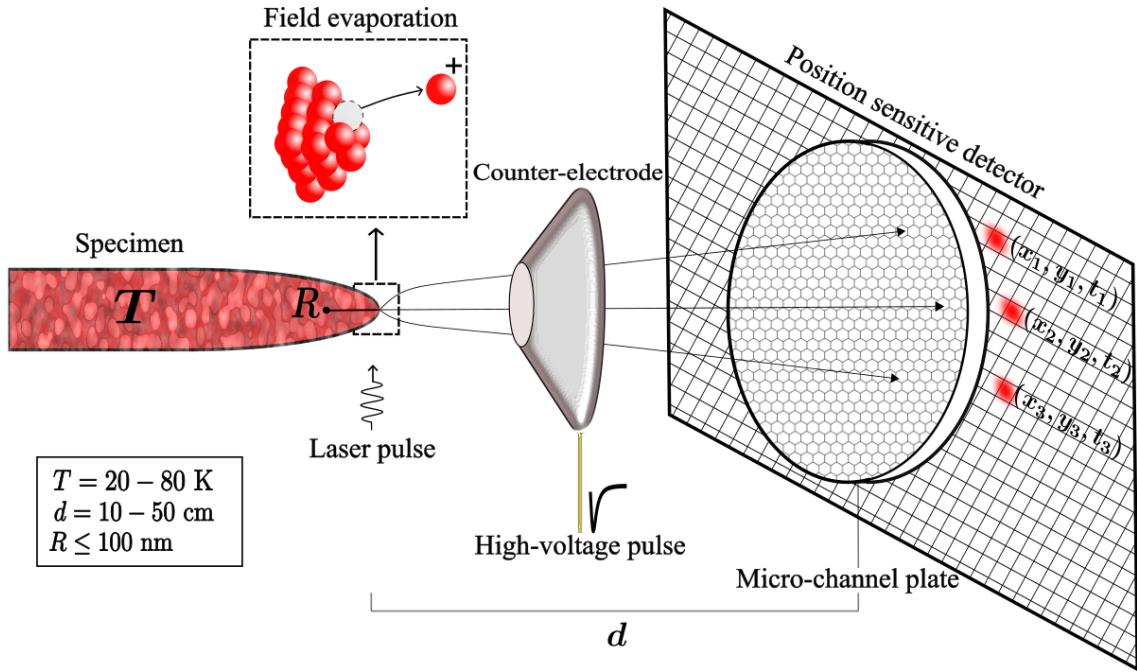


Figure 1: An illustration of the atom probe imaging set up with common temperature ranges T , flight distances d and a tip radius R . On the detector, we denote the hitting coordinates by x and y respectively and the hitting time by t .

Field evaporation events are complex, and understanding the interactions between surface atoms during field evaporation is an ongoing area of research (see e.g. [5]). For example, surface atoms with different evaporation fields have been observed to cause avalanching patterns on the detector [6], which greatly limit the accuracy of traditional reconstruction algorithms due to their restricting assumptions. Another theorized evaporation event involves the rolling-up motion of surface atoms to nearby positions before evaporating [3]. The complex nature of field evaporation has also led to extensive research in the construction of simulation models that would account for these more complex phenomena [7].

In order to account for the complex field evaporation physics, new forms of data reconstruction are needed. The recent advances in computational capacity and machine learning has led to different methods being utilized for APT data reconstruction [8]. Thus, by combining simulation models and machine learning based reconstruction, the aim is to eventually achieve a robust and precise method for the reconstruction of experimental APT data.

2.1.3 Post-processing

After 3D reconstruction, the obtained data is a point cloud with anywhere from a few hundred thousand to a few million atoms. In order to extract meaningful information about the microstructural imperfections or characteristics of a material, post-processing is needed. One such central microstructural feature evident in some crystalline materials is a grain boundary, which alters the properties of a material

greatly [1]. Grain boundaries are areas between two grains within a material, as depicted in figure 2.

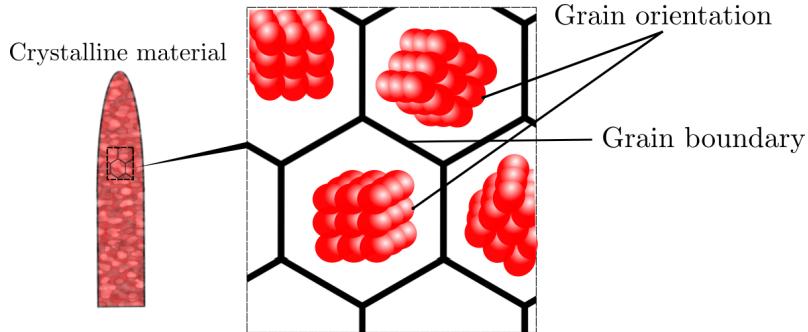


Figure 2: An illustration of how grain boundaries are formed within a 3D crystallographic structure.

Grain boundaries are detected by observing differently oriented atomic planes in the reconstructed point cloud. If the reconstructed material retains the crystallographic structure, these atomic planes can be detected by computing Fourier transforms for the reconstructed atomic lattice [1]. In this thesis, the Fourier transform method is used to determine lattice distancing. Other methods include calculating autocorrelation functions or spatial distribution maps for the reconstructed lattice [4, p. 196].

Other post-processing methods include the compositional analysis of a specimen by splitting the 3D point cloud into identical cubes called voxels. Then each voxel is analyzed separately for microstructural features or compositional variation [9]. This can give information about specific regions of the specimen but also the aggregate composition of the specimen.

These tools and methods enable a comprehensive analysis of the structure and chemical composition of a reconstructed specimen and subsequently give information about the reconstruction performance. This is made possible through the use of a simulation model, where we can directly compare the real specimen structure to that of the reconstruction.

2.2 Gaussian Process Regression

Traditional APT data reconstruction methods cannot capture the possible rolling-up motion and other phenomena evident in the field evaporation process. A more refined approach to reconstruction is achieved by considering the nature of evaporation events as probabilistic. In order to capture the uncertainty present in field evaporation events, we propose a Bayesian inference approach in the form of Gaussian process regression (GPR).

In this section, the theoretical background of GPR is presented by deriving the posterior predictive distribution and by discussing the choice of priors.

2.2.1 Multivariate Gaussian distribution

The necessary mathematical background to understand GPR is based on understanding multivariate Gaussian distributions. A multivariate Gaussian distribution extends the normal distribution for several normally distributed random variables. Consider an n -dimensional random vector $\mathbf{X} = [X_1, \dots, X_n]^T$ where each element $X_i \in \mathbf{X}$ is a normally distributed random variable $X_i \sim \mathcal{N}(\mu_i, \sigma_i^2)$ for $i \in \{1, \dots, n\}$. The probability density is defined as

$$p(\mathbf{X} = \mathbf{x}) = \frac{1}{(2\pi)^{\frac{n}{2}} |\Sigma|^{\frac{1}{2}}} \exp\left(-\frac{1}{2}[\mathbf{x} - \boldsymbol{\mu}]^T \Sigma^{-1} [\mathbf{x} - \boldsymbol{\mu}]\right),$$

where $\boldsymbol{\mu} \in \mathbb{R}^n$ is the mean vector and $\Sigma \in \mathbb{R}^{n \times n}$ the covariance matrix. This is commonly denoted as $\mathbf{X} \sim \mathcal{N}(\boldsymbol{\mu}, \Sigma)$. Now let us consider a random vector $\mathbf{x} = [\mathbf{x}_1 \ \mathbf{x}_2]^T$, which follows a multivariate Gaussian distribution, where $\mathbf{x}_1 \in \mathbb{R}^n$ and $\mathbf{x}_2 \in \mathbb{R}^m$. The conditional probability (see section A.2 for a detailed derivation) is given by

$$p(\mathbf{x}_2 | \mathbf{x}_1) \propto \exp\left(-\frac{1}{2}[\mathbf{x}_2 - \boldsymbol{\mu}_{2|1}]^T \Lambda^{-1} [\mathbf{x}_2 - \boldsymbol{\mu}_{2|1}]\right), \quad (1)$$

where $\Lambda = \Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{21}^T$ and $\boldsymbol{\mu}_{2|1} = \boldsymbol{\mu}_2 + \Sigma_{21}\Sigma_{11}^{-1}(\mathbf{x}_1 - \boldsymbol{\mu}_1)$.

2.2.2 Gaussian Processes

A stochastic process is a collection of random variables $\{Y(x) : x \in X\}$, where X is an indexing set. Consequently, a function $f: X \rightarrow Y$ is a stochastic process if the output space Y consists of random variables and the input space X acts as the indexing set. A Gaussian process (GP) is a stochastic process $f: X \rightarrow Y$ where every finite collection of the random variables Y follows a multivariate Gaussian distribution. Therefore, a GP can be interpreted as a distribution over functions, characterized entirely by a mean function $m: X \rightarrow \mathbb{R}$ and covariance function $k: X \times X \rightarrow \mathbb{R}$, which are defined as follows:

$$\begin{aligned} m(\mathbf{x}) &= \mathbb{E}[f(\mathbf{x})], \\ k(\mathbf{x}, \mathbf{x}') &= \mathbb{E}\left[\left(f(\mathbf{x}) - \mathbb{E}[f(\mathbf{x})]\right)\left(f(\mathbf{x}') - \mathbb{E}[f(\mathbf{x}')]\right)\right]. \end{aligned}$$

A Gaussian process is commonly denoted by $f(\mathbf{x}) \sim \mathcal{GP}(m(\mathbf{x}), k(\mathbf{x}, \mathbf{x}'))$, where the mean function represents the prior beliefs about the expected value of f for any \mathbf{x} , and the covariance function the covariance between function values $f(\mathbf{x})$ and $f(\mathbf{x}')$. The choice of these functions is discussed in section 2.2.3.

In GPR, one assumes that the observed data of inputs $X \in \mathbb{R}^{n \times m}$ and outputs $\mathbf{y} \in \mathbb{R}^n$ has been generated by an underlying Gaussian process $f: X \rightarrow Y$. The objective of GPR is to predict the function value $f(\mathbf{x}_*)$ at a new point \mathbf{x}_* by conditioning on the observed data $\{X, \mathbf{y}\}$. The connection between the Gaussian process function and measurements is of the form

$$y_i = f(\mathbf{x}_i) + \varepsilon,$$

where $\varepsilon \sim \mathcal{N}(0, \sigma^2)$ is independent and identically distributed Gaussian noise. Given that $f(\mathbf{x})$ and ε are assumed to be Gaussian distributed, their sum will also follow a normal distribution. Therefore, the prior on the outputs \mathbf{y} is normally distributed and characterized by [10, p. 16]

$$\begin{aligned} \mathbb{E}[\mathbf{y}] &= m(\mathbf{x}), \\ \text{cov}[\mathbf{y}] &= K(X, X) + \sigma^2 I, \end{aligned}$$

where $K_{ij}(X, X) = k(\mathbf{x}_i, \mathbf{x}_j)$ for $\mathbf{x}_i, \mathbf{x}_j \in X$. Consequently, the joint distribution for a test point $f(\mathbf{x}_*)$ and the observed outputs \mathbf{y} can be derived as

$$\begin{bmatrix} \mathbf{y} \\ f(\mathbf{x}_*) \end{bmatrix} \sim \mathcal{N} \left(\begin{bmatrix} m(X) \\ m(\mathbf{x}_*) \end{bmatrix}, \begin{bmatrix} K(X, X) + \sigma^2 I & K(X, X_*) \\ K(X_*, X) & K(X_*, X_*) \end{bmatrix} \right),$$

where X_* denotes the input matrix with the added test point \mathbf{x}_* and $m(X)$ the mean vector $[m(\mathbf{x}_1) \cdots m(\mathbf{x}_n)]^T$. This joint distribution is a multivariate Gaussian distribution, and the conditional distribution for the predictions $p(f(\mathbf{x}_*)|X, \mathbf{y}, \mathbf{x}_*)$ can be derived according to (1). Therefore, the mean and covariance of the predictive distribution are given by

$$\begin{aligned} \mathbb{E}[f(\mathbf{x}^*)|X, \mathbf{y}, \mathbf{x}_*] &= m(\mathbf{x}_*) + K(X_*, X)[K(X, X) + \sigma^2 I]^{-1}(\mathbf{y} - m(X)), \\ \text{Cov}[f(\mathbf{x}_*)] &= K(X_*, X_*) - K(X_*, X)[K(X, X) + \sigma^2 I]^{-1}K(X_*, X)^T. \end{aligned} \quad (2)$$

Without loss of generality it is often assumed that $\mathbb{E}[f(\mathbf{x})] = 0$, as data can be normalized before computation. If one makes this assumption, the expected value (2) is commonly presented in the form given in [10, p. 16]

$$\mathbb{E}[f(\mathbf{x}^*)|X, Y, \mathbf{x}_*] = K(X_*, X)[K(X, X) + \sigma^2 I]^{-1}\mathbf{y}.$$

This gives a full characterization of the posterior predictive distribution for a test point \mathbf{x}_* . An example of GPR is given in figure 3.

It is notable that this derivation only considered a single output variable, while APT reconstruction requires three-dimensional coordinates as output. Multi-output regression with GPR is an ongoing area of research, but there are a few notable approaches, e.g., extending the kernel for multi-output regression [11] or using multiple independent Gaussian processes for each output variable. The approach used in this thesis is discussed in section 3.3.

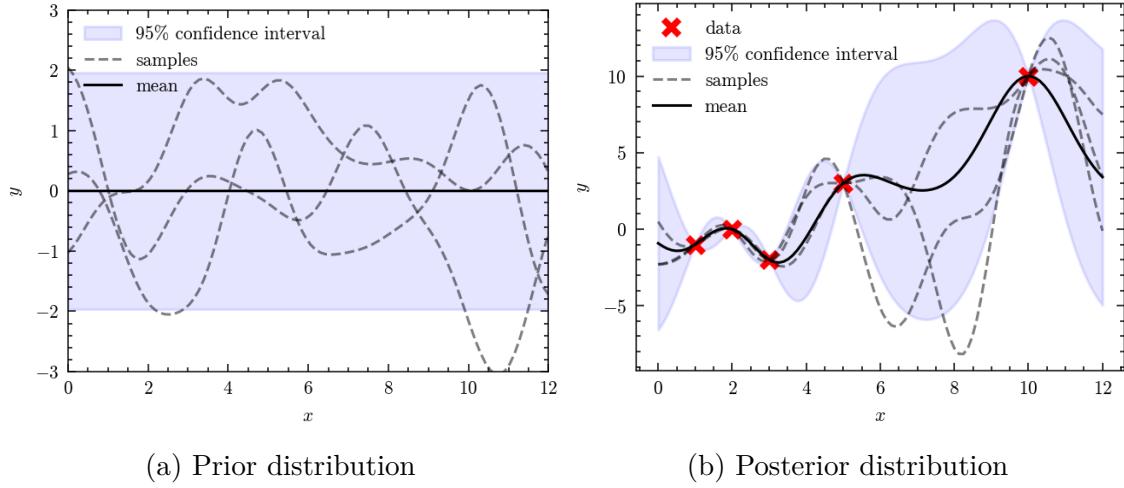


Figure 3: (a) An illustration of a prior distribution of functions for GPR with 3 drawn samples. (b) An illustration of the posterior after observing 5 data points with 3 drawn samples.

2.2.3 Kernels and hyperparameter estimation

As stated previously, the mean and covariance function dictate the properties assigned to the regression function f and consequently represent the prior beliefs about the properties of f . These properties can be things such as smoothness and periodicity, and they are mainly encoded in the covariance function, often referred to as the kernel. Typically, the mean function is assumed to be constant, and we often set $m(\mathbf{x}) = 0$ by normalizing output data prior to training. Thus, our primary focus will be on a few possible kernel choices.

Some of the most used kernels in machine learning are the isotropic and anisotropic forms of the radial basis function (RBF), also referred to as the squared exponential kernel. The anisotropic form is defined as

$$k_{\text{RBF}}(\mathbf{x}, \mathbf{x}') = s^2 \exp\left(-\frac{1}{2}(\mathbf{x} - \mathbf{x}')^T \Phi^{-2}(\mathbf{x} - \mathbf{x}')\right), \quad (3)$$

where s is a scale parameter and the diagonal of Φ contains length scale parameters l for each input dimension. The isotropic form is retained when the matrix Φ is reduced to a scalar length scale l .

The RBF kernel is infinitely differentiable and guarantees very smooth behaviour. However, this behaviour can be seen as unrealistic for many physical processes, and thus it is useful to also consider kernels such as the Matérn class of kernels [10, p. 83]. The Matérn class of kernels is given by

$$k_{\text{Matérn}}(\mathbf{x}, \mathbf{x}') = s^2 \frac{2^{1-\nu}}{\Gamma(\nu)} (\sqrt{2\nu} d)^\nu K_\nu(\sqrt{2\nu} d), \quad (4)$$

where s is a scale parameter, ν a smoothness parameter, K_ν a modified Bessel function of the second kind and d is defined as $d = (\mathbf{x} - \mathbf{x}')^T \Phi^{-2}(\mathbf{x} - \mathbf{x}')$, where again

the length scales are arrayed across the diagonal of Φ . Depending on the smoothness parameter ν , the Matérn kernel can be used to model non-smooth behaviour or extremely smooth behaviour, as the covariance function reduces to the RBF as $\nu \rightarrow \infty$ [10, p. 85].

The parameters within a kernel are referred to as hyperparameters. In table 1 one can observe a few kernels and their respective hyperparameters.

Table 1: Examples of kernels used in GPR and their respective hyperparameters.

Kernel name	Kernel function $k(\mathbf{x}, \mathbf{x}')$	Hyperparameters
RBF	$s^2 \exp(-\frac{1}{2}(\mathbf{x} - \mathbf{x}')^T \Phi^{-2}(\mathbf{x} - \mathbf{x}'))$	s, l_1, l_2, \dots
Matérn	$s^2 \frac{2^{1-\nu}}{\Gamma(\nu)} (\sqrt{2\nu}d)^\nu K_\nu(\sqrt{2\nu}d)$	s, ν, l_1, l_2, \dots
Polynomial	$(\mathbf{x}^T \mathbf{x}' + s^2)^p$	s, p
Constant	s^2	s

After choosing a kernel and mean function, one must still determine values for the possible hyperparameters $\boldsymbol{\theta}$ and the noise variance σ^2 . This is generally done by maximizing the log marginal likelihood function. The log marginal likelihood is given by

$$\log p(\mathbf{y}|X, \boldsymbol{\theta}) = -\frac{1}{2}(\mathbf{y} - m(X))^T [K + \sigma^2 I]^{-1}(\mathbf{y} - m(X)) - \frac{1}{2} \log |K + \sigma^2 I| - \frac{n}{2} \log 2\pi,$$

where $K = K(X, X)$. For a detailed derivation, see [10, p. 27-29]. Again, if one makes the assumption $\mathbb{E}[f(\mathbf{x})] = 0$, the log marginal likelihood takes the more commonly presented form given in [10, p. 19]

$$\log p(\mathbf{y}|X, \boldsymbol{\theta}) = -\frac{1}{2}\mathbf{y}^T [K + \sigma^2 I]^{-1}\mathbf{y} - \frac{1}{2} \log |K + \sigma^2 I| - \frac{n}{2} \log 2\pi. \quad (5)$$

Maximizing (5) with respect to the hyperparameters $\boldsymbol{\theta}$ is equivalent to estimating the hyperparameters for which the observed data is most likely. The noise parameter σ is commonly included as a hyperparameter. Thus, the optimal parameters $\boldsymbol{\theta}^*$ are given by solving the optimization problem

$$\boldsymbol{\theta}^* = \arg \max_{\boldsymbol{\theta} \in \Theta} \log p(\mathbf{y}|X, \boldsymbol{\theta}),$$

where Θ denotes the viable space of values for the hyperparameters $\boldsymbol{\theta}$. This optimization problem is most commonly solved through some variation of gradient descent, where we consider the partial derivatives of (5) with respect to each hyperparameter. Achieving a global maximum is not guaranteed, and the log marginal likelihood can have multiple local maxima [10, p. 115]. Thus, optimizing the hyperparameters is a crucial part of determining the final GPR model, and it represents the learning step of the GPR method.

In the context of APT, the data sets are large, and thus it is useful to address some computational costs of GPR. GPR can be split into two major phases, the

training phase and the prediction phase. In both cases, one has to operate with the inverse matrix $[K + \sigma^2]^{-1}$. This inverse matrix is not explicitly computed due to computational demand, but methods such as Cholesky factorization are used to compute the terms $\mathbf{y}^T [K + \sigma^2 I]^{-1} \mathbf{y}$ and $[K + \sigma^2 I]^{-1} \mathbf{y}$ seen in the training and prediction phase respectively. However, the computational demand is still large and scales as $\mathcal{O}(n^3)$, where n is the number of training data points. A number of approximative methods have been developed to tackle this problem (see e.g. [10] chapter 8). The approach used in this thesis is discussed in section 3.3.

3 Methodology

This section provides a general description of the forward model and a detailed explanation of the data sets and the reconstruction models used in the analysis. All reconstruction models were built using GPyTorch architecture and executed with GPU assisted computation and other computational resources provided by the Aalto Science-IT project. The source code can be accessed through appendix C.

3.1 Forward model

The APT simulation model used to generate the data sets is based on the Robin-Rolland model (RRM) presented by Roland et al. [12] with added complexity in ion trajectories. Unlike earlier implementations, which modelled the specimen surface using a fine rigid mesh, the RRM directly computes a charge distribution across the specimen surface [12]. The charge distribution consists of partially ionized surface atoms, where the surface atoms with the highest charge are most likely to evaporate. After an atom is selected to evaporate according to evaporation probabilities, it is removed from the specimen surface, and the charge distribution is then recalculated locally around the area where the atom was removed. The ion trajectory is then computed according to a Newton motion equation [12], completing one evaporation cycle of the simulation.

The expanded forward model extends the RRM. Namely, the expanded model considers the possible rolling-up motions of surface atoms before evaporation. When a surface atom experiences rolling-up motion, it moves up to a nearby lattice position that enables easier evaporation prior to the actual evaporation event [3].

3.2 Data and feature selection

There are two specimens of interest with different grain structures. The first specimen has a uniform grain structure, while the second specimen has a grain boundary. The first specimen consists of a face-centered cubic (fcc) grain oriented along the [001] direction, while the grain boundary in the second specimen separates two fcc grains oriented along the [593] and [217] directions, as specified by Miller indices. The chemical compositions of both specimens are uniform, and the analysis will focus on structural features. The specimens and their detector maps can be observed in figures 4 and 5 respectively.

The resulting data consists of specimen coordinates, detector coordinates, orders of arrival, electric field strengths and other properties recorded during measurements. The features selected for reconstruction were the detector coordinates and orders of arrival, which enable the identification of each atom while keeping the dimensionality of the GPR model at a reasonable level. The reasoning behind these feature choices is twofold. It is safe to assume that atoms that have similar detector hit coordinates came from a relatively identical region of the specimen. Additionally, the order of arrival assumes a correlation between recently evaporated atoms, which can be seen as a good indicator for how many layers of the specimen have evaporated.

First, the analysis was carried out for the uniform data set of size 200k atoms and secondly for the more complex grain boundary data set of size 500k atoms. Due to the computational demands of GPR, less traditional training/test splits were used. A training/test split of 10%/90% was used for the uniform specimen and a 6%/94% split for the grain boundary specimen, as these training sizes were near the maximum computational capacity of a single GPU.

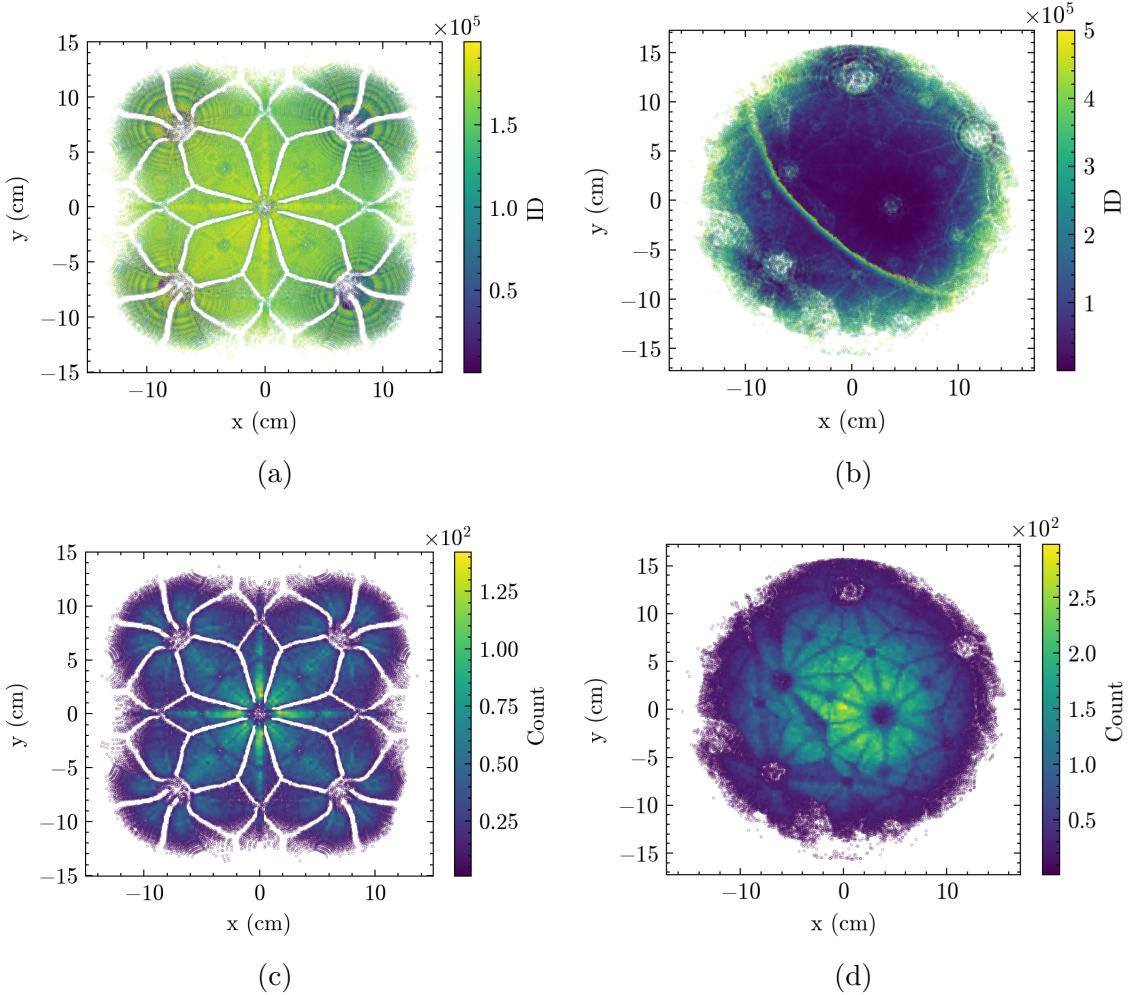


Figure 4: Detector maps for the uniform grain specimen (left) and the grain boundary specimen (right). (a-b) Detector maps with colour scaling by order of arrival (ID). (b-d) Detector hit maps with colour scaling by number of hits for a 2D histogram with 100 bins.

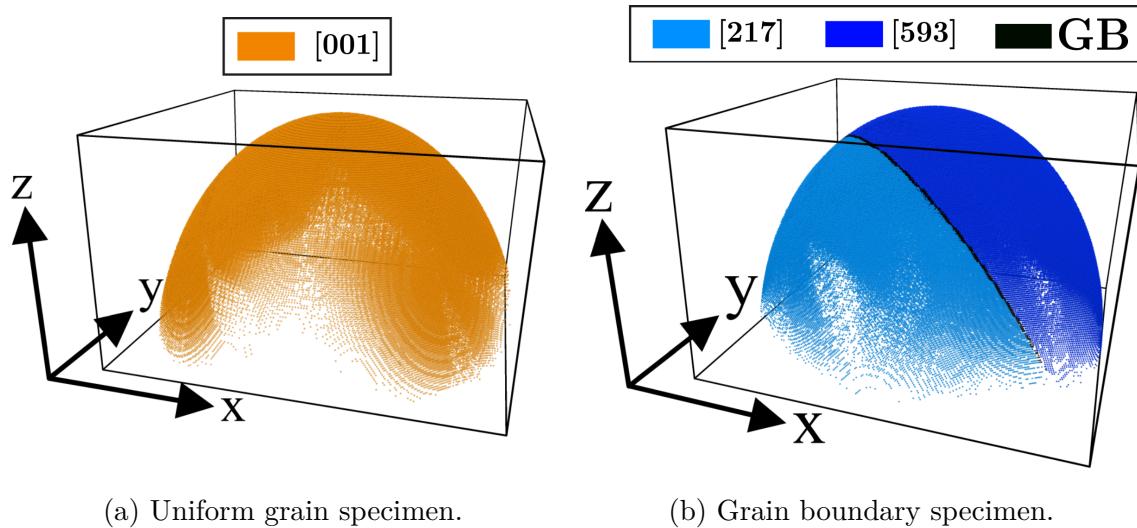


Figure 5: Original specimens with their respective grain orientations.

3.3 Reconstruction models

All reconstruction models were implemented in Python using the GPyTorch library [13], which provides a modular and fast framework for building GPR models. The increase in performance provided by GPyTorch comes from Blackbox Matrix-Matrix multiplication, which was shown to be much faster than standard Cholesky decomposition in computing inverse matrices with GPU acceleration [13]. Blackbox Matrix-Matrix multiplication is based on the conjugate gradient method [13], which is an efficient method for solving large sparse systems of linear equations. The models implemented in this thesis utilize exact inference with GPU acceleration in order to adapt to large APT data sets.

All models in this thesis utilized a constant mean function $m(\mathbf{x}) = \boldsymbol{\mu}$, where $\boldsymbol{\mu}$ is a learned hyperparameter. Initially, the models used the RBF kernel (3) due to its well-known properties and versatility, but since the reconstruction of a crystallographic lattice implies a discrete output space, the smoothness property can be seen as redundant. Therefore, the Matérn kernel (4) with $\nu = 2.5$ was also considered as an alternative, which, while analogous, does not guarantee smoothness. The base GPR models used in this thesis can thus be summarized by

$$\begin{aligned} f_{\text{RBF}}(\mathbf{x}) &\sim \mathcal{GP}\left(\boldsymbol{\mu}, k_{\text{RBF}}(\mathbf{x}, \mathbf{x}')\right), \\ f_{\text{Matérn}}(\mathbf{x}) &\sim \mathcal{GP}\left(\boldsymbol{\mu}, k_{\text{Matérn}}(\mathbf{x}, \mathbf{x}')\right). \end{aligned}$$

APT data reconstruction is a multi-output regression problem with three-dimensional output. The multi-output regression approach used in this thesis considers the outputs as correlated and is based on the approach presented by Bonilla et al. [11], which is readily implemented in GPyTorch. This approach extends the kernel by considering the covariance between output variables in addition to the base kernels.

The hyperparameters to be learned are the same for both models. The models learn three different length scales $\{l_1, l_2, l_3\}$ and mean parameters $\{\mu_1, \mu_2, \mu_3\}$,

which correspond to the three different input features and target values respectively. They also include the noise parameter σ as a hyperparameter. The optimal values for each model and data set can be seen in appendix B.

After defining the GPR model, the hyperparameters were optimized by the method of maximum log likelihood estimation as outlined in section 2.2.3. The optimization was done with the adaptive moment estimation algorithm (Adam) for its computational efficiency and small memory requirements. Adam is a form of stochastic gradient descent that is suited for large non-convex optimization problems, making it an appropriate choice for large APT data sets.

The prediction accuracy was assessed with several statistical measures in addition to visual analysis and the post-processing methods outlined in section 2.1.3. In order to estimate the proportion of explained variance by the model for the target variables, the coefficient of determination R^2 was used, which is defined as

$$R^2 = 1 - \frac{SS_{\text{residual}}}{SS_{\text{total}}},$$

where SS_{residual} denotes the residual sum of squares $\sum_{i=1}^n e_i^2$ and SS_{total} the total sum of squares $\sum_{i=1}^n (y_i - \bar{y})^2$ for n observations. Due to the nature of GPR adapting nearly exactly to training data, this statistic can be seen as biased. Thus, we also considered absolute metrics such as root mean squared error (RMSE) and mean absolute error (MAE) defined as follows

$$\begin{aligned} \text{MAE} &= \frac{1}{n} \sum_{i=1}^n |e_i|, \\ \text{RMSE} &= \sqrt{\frac{1}{n} \sum_{i=1}^n e_i^2}, \end{aligned}$$

where $e_i = y_i - \hat{y}_i$ denotes the residuals between true values y_i and estimates \hat{y}_i .

4 Results

This section presents the reconstruction results for the uniform grain data set and the grain boundary data set. The performance between reconstruction models is actively discussed, while lastly the performance between data sets is discussed in section 5.

In both cases, the data produced by the reconstruction models was cleaned by discarding infeasible values, that is, values outside the physical dimensions of the specimen. The proportion of infeasible values was negligible, amounting to less than one hundredth of a percent for the grain boundary data set and none for the uniform grain data set.

4.1 Uniform Data Set

4.1.1 Metrics and Predictions

The MAE and R^2 metrics of the reconstruction model for both the RBF and Matérn kernel can be seen in table 2, while the RMSE values can be observed in appendix B.2. The statistics show that the predictions are good on average, with only slight deviations in the reconstruction. The worst relative performance is seen in predicting the z -coordinate, while the worst absolute distances from true values is seen with the lateral coordinates.

Table 2: Performance metrics for the uniform specimen reconstruction.

Models	Relative metrics (%)						Absolute metrics (nm)		
	R_x^2	R_y^2	R_z^2	$\frac{\text{MAE}_x}{\text{Range}}$	$\frac{\text{MAE}_y}{\text{Range}}$	$\frac{\text{MAE}_z}{\text{Range}}$	MAE_x	MAE_y	MAE_z
RBF	99.97	99.97	99.94	0.265	0.269	0.344	0.394	0.340	0.219
Matérn	99.97	99.97	99.95	0.240	0.244	0.299	0.356	0.362	0.190

In figure 6, the predictions against test values are shown for the Matérn kernel, while the same figure for the RBF can be seen in appendix B.1. The Matérn kernel outperforms the RBF in relative metrics, but both kernels exhibit similar behavior. The prediction performance is relatively uniform, with the best performance exhibited on the very tip of the specimen, as evidenced by the tapering in the plot for z -coordinates.

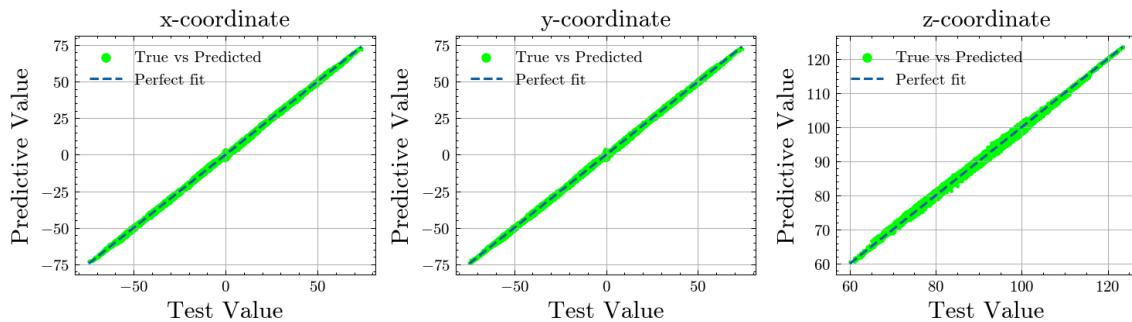


Figure 6: Test values vs. predictions for the Matérn kernel.

Both kernels capture the general form and shape of the specimen in the reconstruction. However, there are clear sparse radial patterns in the reconstruction, as seen in figure 7. These sparse patterns are further highlighted in figure 8, where the original specimen is visualized with a color gradient corresponding to Euclidean error. The sparse areas correspond with the greatest sources of error in the reconstruction.

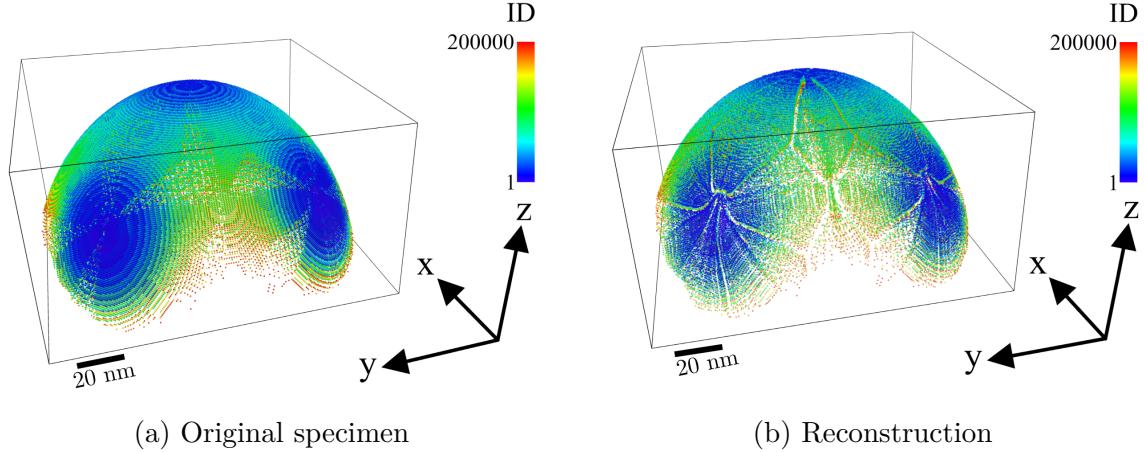


Figure 7: (a-b) Specimen plots with color gradients denoting the order of arrival on the detector.

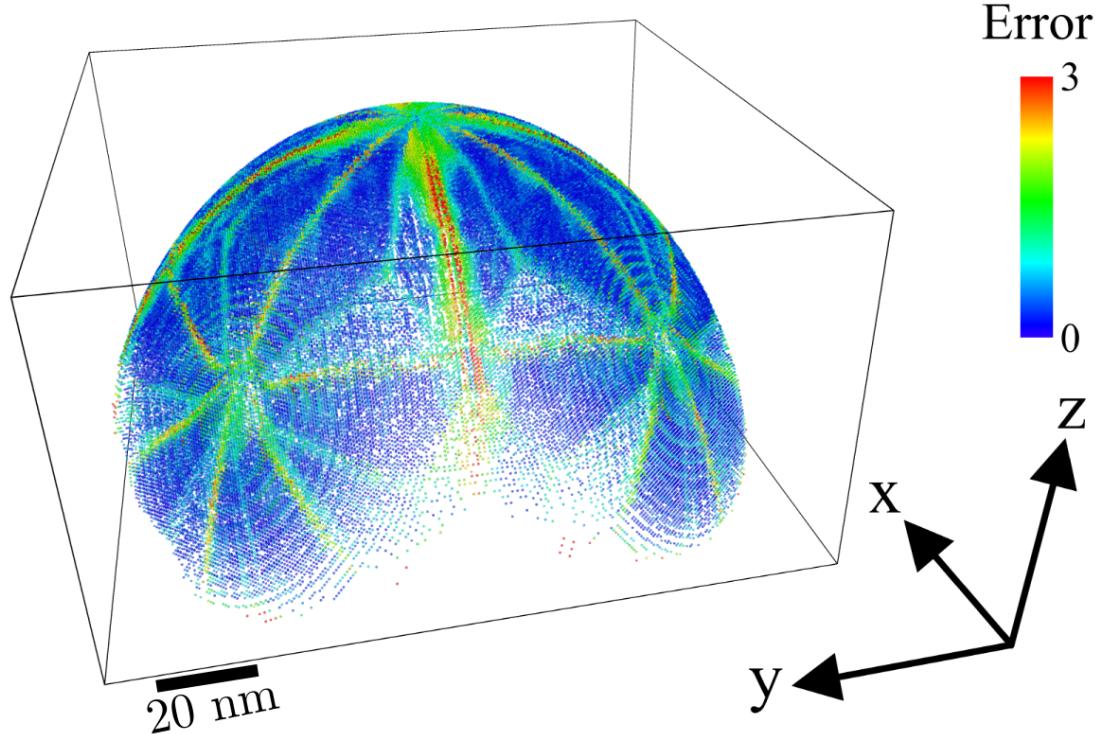


Figure 8: Original specimen with a color gradient corresponding to Euclidean distance between the original positions and reconstructed positions.

4.1.2 Crystallography

In order to explore the crystallographic structure of the reconstruction, a voxel was inspected around the tip of the specimen. The [001] grain direction is clearly observable in the reconstruction, as seen by comparing figures 9a and 9b. Furthermore, the lattice distancing of the reconstruction that was obtained from the Fourier transforms taken along the [001] direction agrees with the original lattice distancing. This is evidenced by comparing the distance between meaningful peaks in figure 9d and comparing it to the distance between peaks in figure 9c.

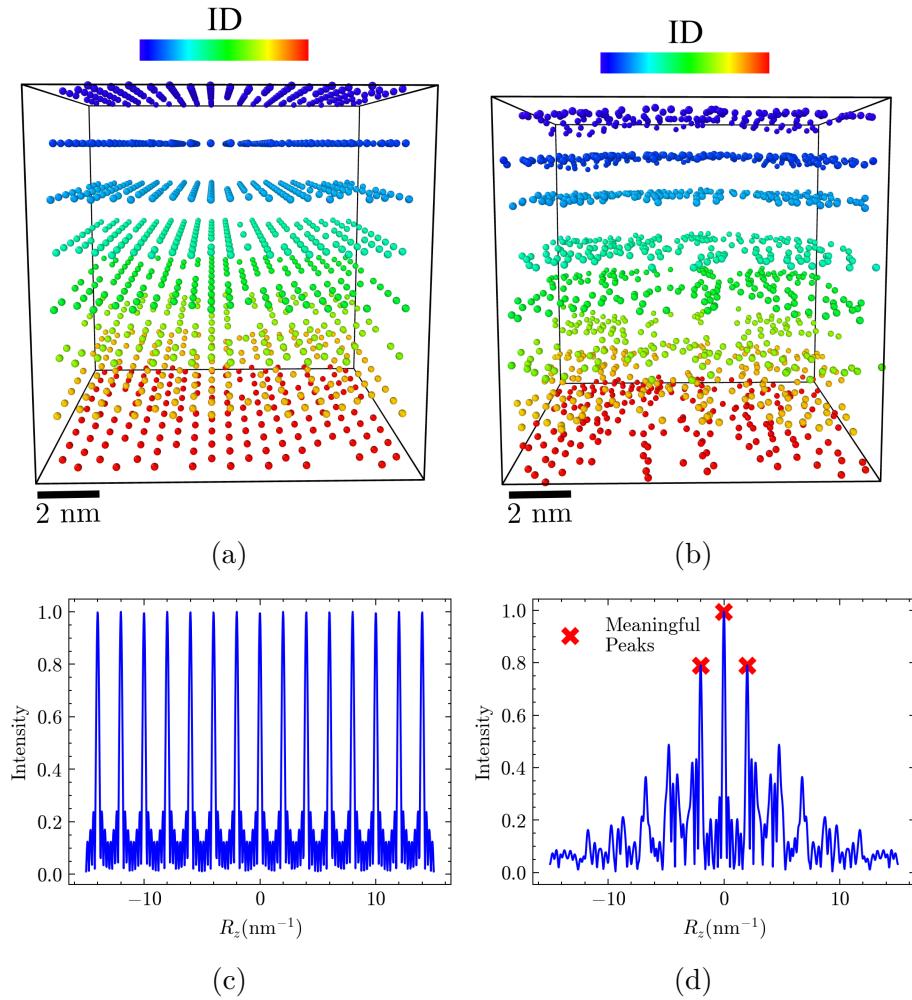


Figure 9: (a-b) Voxels taken around the tip of the specimen for the original lattice and the reconstruction. (c-d) The corresponding Fourier transforms for the original specimen and the reconstruction, respectively.

4.2 Grain Boundary Data Set

4.2.1 Metrics and Predictions

The R^2 and MAE statistics for the data set with the added complexity of a grain boundary can be seen in table 3 and the RMSE values in appendix B. The accuracy of predictions clearly decreases when compared with the uniform data set. As in the uniform case, the worst relative performance is seen with the z -coordinate and the worst absolute performance with lateral coordinates. In this case, the Matérn kernel outperforms the RBF slightly in nearly every metric.

Table 3: Performance metrics for the grain boundary data set reconstruction.

Models	Relative metrics (%)						Absolute metrics (nm)		
	R_x^2	R_y^2	R_z^2	$\frac{\text{MAE}_x}{\text{Range}}$	$\frac{\text{MAE}_y}{\text{Range}}$	$\frac{\text{MAE}_z}{\text{Range}}$	MAE $_x$	MAE $_y$	MAE $_z$
RBF	99.84	99.92	99.81	0.424	0.415	0.577	0.674	0.657	0.449
Matérn	99.88	99.92	99.82	0.411	0.403	0.551	0.653	0.637	0.429

In figure 10, the predictions against test values are shown for the Matérn kernel, while the corresponding figure for the RBF can be viewed in appendix B. It is evident that the prediction accuracy is worst around the edges of the specimen, and the model produces more outliers when compared with the uniform data set. However, the portion of predictions with an error greater than 3 nm in any coordinate direction is relatively low, at approximately 1% of predictions. It is notable that the characteristics of the x and y coordinate prediction plots are distinct. This asymmetry is either caused by the added complexity or the reconstruction model.

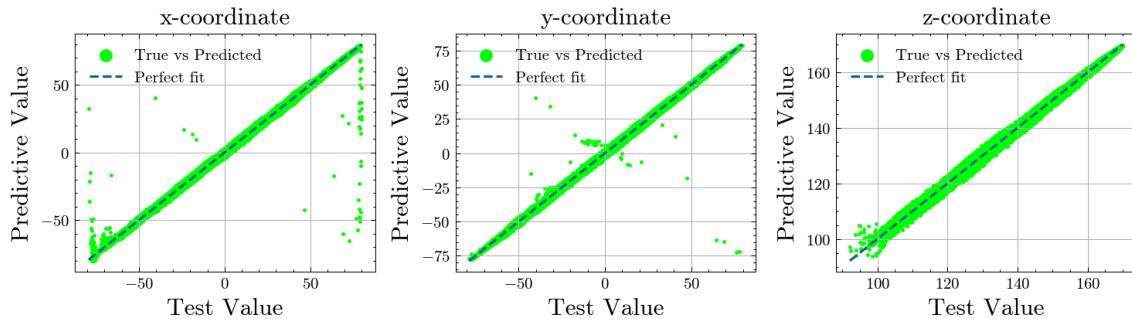


Figure 10: Test values vs. predictions for the Matérn kernel.

The general composition and spacing in the reconstruction matches that of the original specimen, as seen in figure 11. The grain boundary takes the form of a sparse area between both fcc grains in the reconstruction. Other sparse regions akin to the ones observed in the uniform case are not immediately discernible. However, we do observe radial error patterns when visualizing the original specimen with a color gradient corresponding to Euclidean error, as seen in figure 12. The grain

boundary is clearly discernible as an area of high error, while the errors form clear radial patterns.

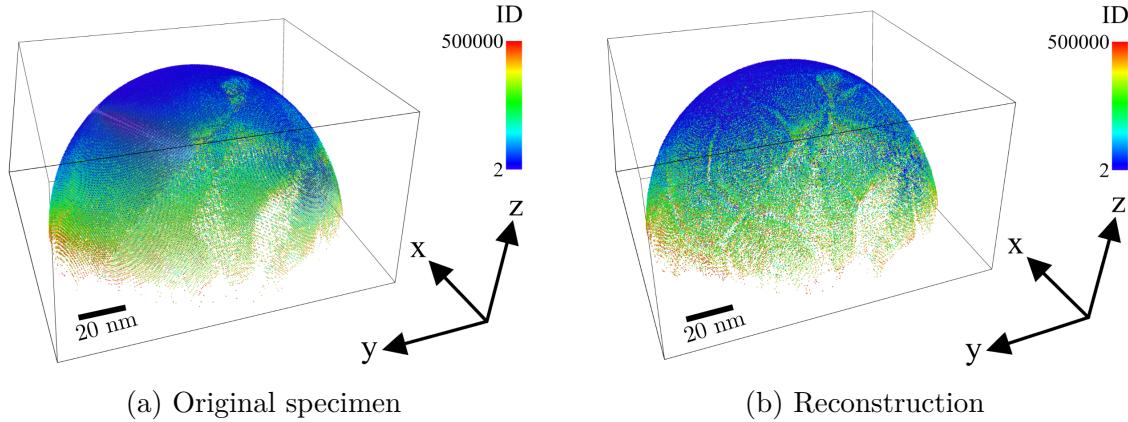


Figure 11: Specimen plots with color gradients denoting the order of arrival on the detector.

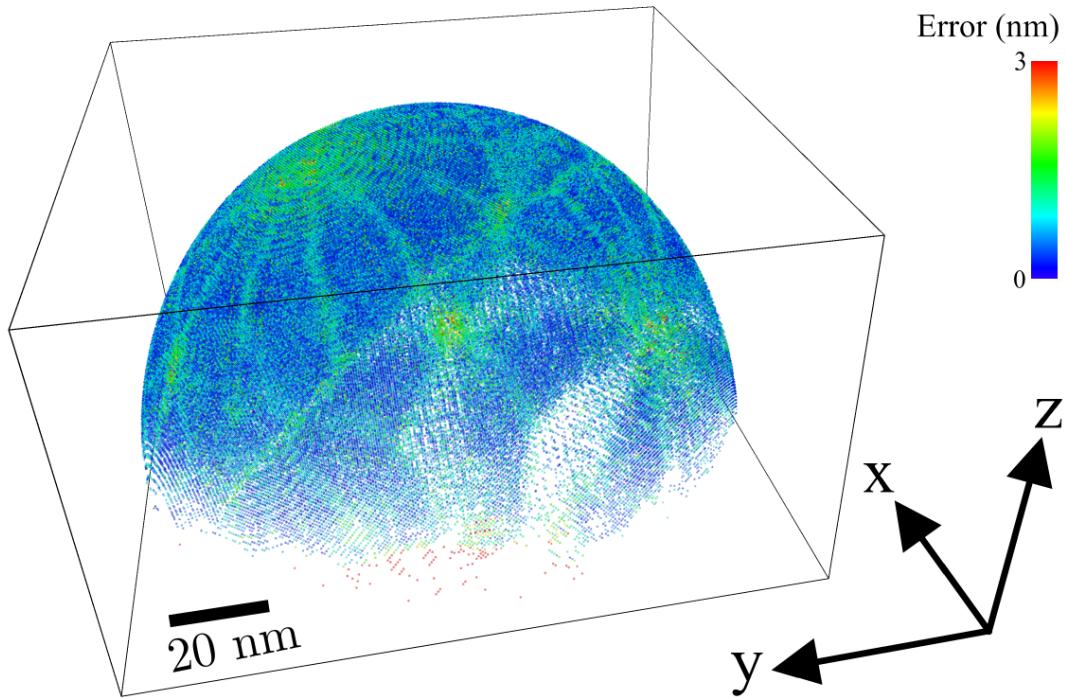


Figure 12: Original specimen with a color gradient corresponding to Euclidean distance between the original positions and reconstructed positions.

4.2.2 Crystallography

In order to explore the crystallographic structure of the reconstruction, two voxels were inspected in both the [217] and [593] grains. In both grains, the reconstruction

struggles with capturing the atomic planes that form the grains, as seen in figure 13. As a result, the lattice distancing is not recovered in the reconstruction. A visual inspection does reveal that there is some periodic structure observed in the [593] grain, but it remains too small for clear distinguishing of atomic planes.

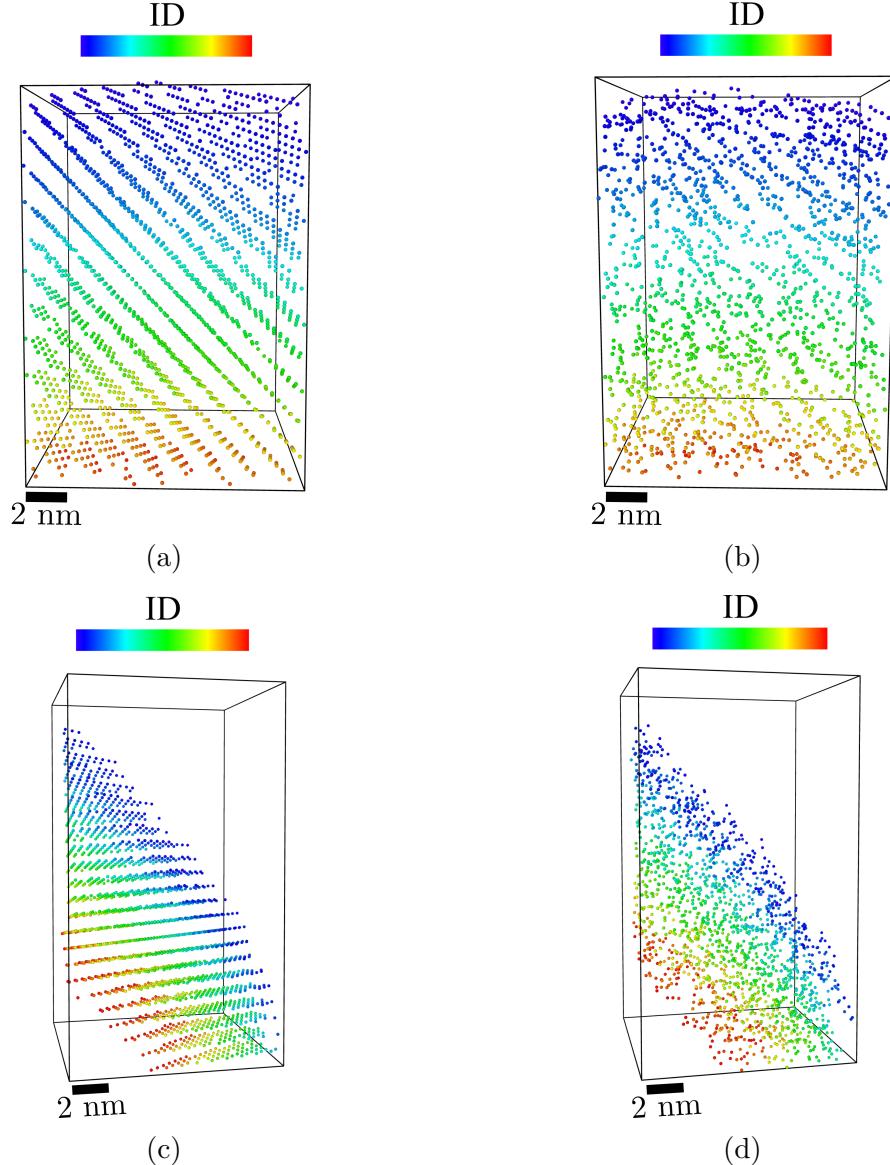


Figure 13: Voxels taken from the original specimen (left) and the reconstruction (right). (a-b) Voxels taken from the [593] grain. (c-d) Voxels taken from the [217] grain.

5 Discussion and Conclusions

The aim of this thesis was to assess and explore the use of GPR in APT data reconstruction, while giving a theoretical overview of GPR. The results show that GPR can be used to obtain the general composition of an APT specimen with both a simple and multigrain structure. GPR also shows promise in the reconstruction of crystallographic features and was shown to be able to accurately reconstruct a uniform grain structure. Precise crystallographic reconstruction for more complex grain structures requires further development of the method.

These results show that with further study, GPR could be utilized more in APT data reconstruction. Here, GPR was used to reconstruct the whole specimen tip from detector data. This approach could be refined by considering the traditional reconstruction algorithms and modelling a portion of the physically motivated algorithm as a GP.

The results also show unexpected systematic error patterns that possess a similarity to sparse patterns formed on the detector. These patterns reflect the constraints of the reconstruction models, where the connection between the detector map and the 3D reconstruction is perhaps too direct. The results also show that GPR models with more complex kernels may yield better results. However, future development of the method would benefit from kernels and function structures that are motivated by physics. The spatial periodicity of crystallographic structures is a property that could be enforced with the right kernel structure.

In addition to model complexity, one of the main challenges of GPR is scalability. Here it was shown that reconstruction can be done for data sets with a few hundred thousand atoms but scaling up to typical APT data sets with millions' of atoms remains an obstacle. This could be solved through the use of multiple GPUs or by using approximative methods such as sparse GPs.

References

- [1] B. Gault, A. Chiaramonti, O. Cojocaru-Mirédin, *et al.*, “Atom probe tomography,” en, *Nature Reviews Methods Primers*, vol. 1, no. 1, p. 51, Jul. 2021, ISSN: 2662-8449. DOI: [10.1038/s43586-021-00047-w](https://doi.org/10.1038/s43586-021-00047-w).
- [2] A. D. Giddings, S. Koelling, Y. Shimizu, *et al.*, “Industrial application of atom probe tomography to semiconductor devices,” *Scripta Materialia*, vol. 148, pp. 82–90, Apr. 2018, ISSN: 1359-6462. DOI: [10.1016/j.scriptamat.2017.09.004](https://doi.org/10.1016/j.scriptamat.2017.09.004).
- [3] F. Vurpillot, S. Parviainen, F. Djurabekova, D. Zanuttini, and B. Gervais, “Simulation tools for atom probe tomography: A path for diagnosis and treatment of image degradation,” *Materials Characterization*, vol. 146, pp. 336–346, Dec. 2018, ISSN: 1044-5803. DOI: [10.1016/j.matchar.2018.04.024](https://doi.org/10.1016/j.matchar.2018.04.024).
- [4] W. Lefebvre, F. Vurpillot, and X. Sauvage, Eds., *Atom Probe Tomography: Put Theory Into Practice*, English, 1st edition. London: Academic Press, Jun. 2016, ISBN: 978-0-12-804647-0.
- [5] L. Rousseau, A. Normand, F. F. Morgado, *et al.*, “Introducing field evaporation energy loss spectroscopy,” en, *Communications Physics*, vol. 6, no. 1, pp. 1–8, May 2023, Publisher: Nature Publishing Group, ISSN: 2399-3650. DOI: [10.1038/s42005-023-01203-2](https://doi.org/10.1038/s42005-023-01203-2).
- [6] F. Vurpillot, A. Bostel, and D. Blavette, “A new approach to the interpretation of atom probe field-ion microscopy images,” *Ultramicroscopy*, 46th International Field Emission Symposium, vol. 89, no. 1, pp. 137–144, Oct. 2001, ISSN: 0304-3991. DOI: [10.1016/S0304-3991\(01\)00097-3](https://doi.org/10.1016/S0304-3991(01)00097-3).
- [7] F. Vurpillot and C. Oberdorfer, “Modeling Atom Probe Tomography: A review,” *Ultramicroscopy*, 1st International Conference on Atom Probe Tomography & Microscopy, vol. 159, pp. 202–216, Dec. 2015, ISSN: 0304-3991. DOI: [10.1016/j.ultramic.2014.12.013](https://doi.org/10.1016/j.ultramic.2014.12.013).
- [8] Y. Li, Y. Wei, Z. Wang, *et al.*, “Quantitative three-dimensional imaging of chemical short-range order via machine learning enhanced atom probe tomography,” en, *Nature Communications*, vol. 14, no. 1, p. 7410, Nov. 2023, Publisher: Nature Publishing Group, ISSN: 2041-1723. DOI: [10.1038/s41467-023-43314-y](https://doi.org/10.1038/s41467-023-43314-y).
- [9] B. Gault, “A Brief Overview of Atom Probe Tomography Research,” *Applied Microscopy*, vol. 46, pp. 117–126, Sep. 2016. DOI: [10.9729/AM.2016.46.3.117](https://doi.org/10.9729/AM.2016.46.3.117).
- [10] C. E. Rasmussen and C. K. I. Williams, *Gaussian processes for machine learning* (Adaptive computation and machine learning), en. Cambridge, Mass: MIT Press, 2006, OCLC: ocm61285753, ISBN: 978-0-262-18253-9.
- [11] E. Bonilla, K. Chai, and C. Williams, “Multi-task Gaussian Process Prediction.,” Jan. 2007.

- [12] N. Rolland, F. Vurpillot, S. Duguay, and D. Blavette, “A Meshless Algorithm to Model Field Evaporation in Atom Probe Tomography,” eng, *Microscopy and Microanalysis: The Official Journal of Microscopy Society of America, Microbeam Analysis Society, Microscopical Society of Canada*, vol. 21, no. 6, pp. 1649–1656, Dec. 2015, ISSN: 1435-8115. DOI: [10.1017/S1431927615015184](https://doi.org/10.1017/S1431927615015184).
- [13] J. R. Gardner, G. Pleiss, D. Bindel, K. Q. Weinberger, and A. G. Wilson, *GPyTorch: Blackbox Matrix-Matrix Gaussian Process Inference with GPU Acceleration*, arXiv:1809.11165 [cs, stat], Jun. 2021. DOI: [10.48550/arXiv.1809.11165](https://doi.org/10.48550/arXiv.1809.11165).
- [14] T.-T. Lu and S.-H. Shiou, “Inverses of 2×2 block matrices,” *Computers & Mathematics with Applications*, vol. 43, no. 1, pp. 119–129, Jan. 2002, ISSN: 0898-1221. DOI: [10.1016/S0898-1221\(01\)00278-4](https://doi.org/10.1016/S0898-1221(01)00278-4).
- [15] T. Beckers, *An Introduction to Gaussian Process Models*, arXiv:2102.05497 [cs, eess], Feb. 2021. DOI: [10.48550/arXiv.2102.05497](https://doi.org/10.48550/arXiv.2102.05497).

A Mathematical Background

A.1 Matrix identities

In order to derive the form of a multivariate Gaussian distribution, one must derive a couple of auxiliary matrix identities.

Let us consider a matrix $\Sigma \in \mathbb{R}^{n \times n}$ that is invertible such that we can partition the matrix and its inverse as

$$\Sigma = \begin{bmatrix} A & B \\ C & D \end{bmatrix}, \quad \Sigma^{-1} = \begin{bmatrix} E & F \\ G & H \end{bmatrix},$$

where $A, E \in \mathbb{R}^{n_1 \times n_1}$ and $D, H \in \mathbb{R}^{n_2 \times n_2}$ are assumed to be square matrices with $n = n_1 + n_2$. If we assume that matrices A and D and the corresponding Schur complements $D - CA^{-1}B$ and $A - BD^{-1}C$ are invertible, one can derive well known expressions for the inverse of Σ^{-1} as [14]

$$\Sigma^{-1} = \begin{bmatrix} (A - BD^{-1}C)^{-1} & -(A - BD^{-1}C)^{-1}BD^{-1} \\ -D^{-1}C(A - BD^{-1}C)^{-1} & D^{-1} + D^{-1}C(A - BD^{-1}C)^{-1}BD^{-1} \end{bmatrix}$$

and

$$\Sigma^{-1} = \begin{bmatrix} A^{-1} + A^{-1}B(D - CA^{-1}B)^{-1}CA^{-1} & -A^{-1}B(D - CA^{-1}B)^{-1} \\ -(D - CA^{-1}B)^{-1}CA^{-1} & (D - CA^{-1}B)^{-1} \end{bmatrix}. \quad (\text{A1})$$

Both formulas can be derived by simply considering the identities $\Sigma^{-1}\Sigma = I$ and $\Sigma\Sigma^{-1} = I$ and rearranging terms to form expressions for matrices E, F, G and H .

A.2 Multivariate Gaussian distribution

This derivation is done for the benefit of the reader and is heavily inspired by a derivation implemented by Beckers [15].

Let $\mathbf{X} = [X_1, \dots, X_n]^T$ be a vector of normally distributed random variables with a mean vector $\boldsymbol{\mu} \in \mathbb{R}^n$ and covariance matrix $\Sigma \in \mathbb{R}^{n \times n}$ for some $n \in \mathbb{N}$. The multivariate normal distribution is defined as

$$p(\mathbf{X} = \mathbf{x}) = \frac{1}{(2\pi)^{\frac{n}{2}} |\Sigma|^{\frac{1}{2}}} \exp(-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu})^T \Sigma^{-1}(\mathbf{x} - \boldsymbol{\mu})), \quad (\text{A2})$$

where $\Sigma = \mathbb{E}[(\mathbf{X} - \mathbb{E}[\mathbf{X}])(\mathbf{X} - \mathbb{E}[\mathbf{X}])^T]$ and $\boldsymbol{\mu} = [\mathbb{E}[X_1], \dots, \mathbb{E}[X_n]]^T$.

Let $\mathbf{X} = [\mathbf{X}_1 \ \mathbf{X}_2]^T \in \mathbb{R}^n$ follow a multivariate Gaussian distribution with $\mathbf{X}_1 \in \mathbb{R}^{n_1}$ and $\mathbf{X}_2 \in \mathbb{R}^{n_2}$, where $n = n_1 + n_2$. Now the joint distribution of \mathbf{X}_1 and \mathbf{X}_2 can be expressed as

$$\begin{bmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{bmatrix}, \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}\right),$$

where $\boldsymbol{\mu}_1 \in \mathbb{R}^{n_1}$, $\boldsymbol{\mu}_2 \in \mathbb{R}^{n_2}$ and $\Sigma_{11} \in \mathbb{R}^{n_1 \times n_1}$, $\Sigma_{22} \in \mathbb{R}^{n_2 \times n_2}$ are the mean vectors and covariance matrices for \mathbf{X}_1 and \mathbf{X}_2 respectively. Additionally, note that $\Sigma_{12} = \Sigma_{21}^T$

and $\Sigma_{21} = \Sigma_{12}^T$ due to the symmetry of covariance. Our aim is to derive the conditional probability

$$p(\mathbf{x}_2|\mathbf{x}_1) = \frac{p(\mathbf{x}_1, \mathbf{x}_2)}{p(\mathbf{x}_1)},$$

where $p(\mathbf{x}_1, \mathbf{x}_2)$ is written according to (A2) as

$$p(\mathbf{x}_1, \mathbf{x}_2) = p(\mathbf{X} = \mathbf{x}) = \frac{1}{(2\pi)^{(n_1+n_2)/2} |\Sigma|^{\frac{1}{2}}} \exp \left(-\frac{1}{2} (\mathbf{x} - \boldsymbol{\mu})^T \Sigma^{-1} (\mathbf{x} - \boldsymbol{\mu}) \right).$$

In order to simplify the derivation we will only consider terms dependent on \mathbf{x}_2 , that is

$$p(\mathbf{x}_2|\mathbf{x}_1) \propto p(\mathbf{x}_2, \mathbf{x}_1) \propto \exp \left(-\frac{1}{2} (\mathbf{x} - \boldsymbol{\mu})^T \Sigma^{-1} (\mathbf{x} - \boldsymbol{\mu}) \right) \quad (\text{A3})$$

Now the conditional probability can be derived by simply considering the exponential portion of (A3). The exponent can be simplified by considering the matrix identity (A1). For clarity, we use the notation $\Lambda = \Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{21}^T$.

$$\begin{aligned} p(\mathbf{x}_2|\mathbf{x}_1) &\propto \exp \left(-\frac{1}{2} \begin{bmatrix} \mathbf{x}_1 - \boldsymbol{\mu}_1 \\ \mathbf{x}_2 - \boldsymbol{\mu}_2 \end{bmatrix}^T \begin{bmatrix} \Sigma_{11} & \Sigma_{21}^T \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{x}_1 - \boldsymbol{\mu}_1 \\ \mathbf{x}_2 - \boldsymbol{\mu}_2 \end{bmatrix} \right) \\ &\propto \exp \left(-\frac{1}{2} \begin{bmatrix} \mathbf{x}_1 - \boldsymbol{\mu}_1 \\ \mathbf{x}_2 - \boldsymbol{\mu}_2 \end{bmatrix}^T \begin{bmatrix} \Sigma_{11}^{-1} + \Sigma_{11}^{-1}\Sigma_{21}^T\Lambda^{-1}\Sigma_{21}\Sigma_{11}^{-1} & -\Sigma_{11}\Sigma_{21}^T\Lambda^{-1} \\ -\Lambda^{-1}\Sigma_{21}\Sigma_{11}^{-1} & \Lambda^{-1} \end{bmatrix} \begin{bmatrix} \mathbf{x}_1 - \boldsymbol{\mu}_1 \\ \mathbf{x}_2 - \boldsymbol{\mu}_2 \end{bmatrix} \right) \\ &\propto \exp \left(-\frac{1}{2} [\mathbf{x}_1 - \boldsymbol{\mu}_1]^T (\Sigma_{11}^{-1} + \Sigma_{11}^{-1}\Sigma_{21}^T\Lambda^{-1}\Sigma_{21}\Sigma_{11}^{-1}) [\mathbf{x}_1 - \boldsymbol{\mu}_1] \right. \\ &\quad + \frac{1}{2} [\mathbf{x}_2 - \boldsymbol{\mu}_2]^T \Lambda^{-1} \Sigma_{21} \Sigma_{11}^{-1} [\mathbf{x}_1 - \boldsymbol{\mu}_1] \\ &\quad \left. + \frac{1}{2} [\mathbf{x}_1 - \boldsymbol{\mu}_1]^T \Sigma_{11} \Sigma_{21}^T \Lambda^{-1} [\mathbf{x}_2 - \boldsymbol{\mu}_2] - \frac{1}{2} [\mathbf{x}_2 - \boldsymbol{\mu}_2]^T \Lambda^{-1} [\mathbf{x}_2 - \boldsymbol{\mu}_2] \right) \\ &\propto \exp \left(-\frac{1}{2} [\mathbf{x}_1 - \boldsymbol{\mu}_1]^T \Sigma_{11}^{-1} [\mathbf{x}_1 - \boldsymbol{\mu}_1] \right) \\ &\cdot \exp \left(-\frac{1}{2} \left([\mathbf{x}_1 - \boldsymbol{\mu}_1]^T \Sigma_{11}^{-1} \Sigma_{21}^T \Lambda^{-1} \Sigma_{21} \Sigma_{11}^{-1} [\mathbf{x}_1 - \boldsymbol{\mu}_1] - 2[\mathbf{x}_2 - \boldsymbol{\mu}_2]^T \Lambda^{-1} \Sigma_{21} \Sigma_{11}^{-1} [\mathbf{x}_1 - \boldsymbol{\mu}_1] \right. \right. \\ &\quad \left. \left. + [\mathbf{x}_2 - \boldsymbol{\mu}_2]^T \Lambda^{-1} [\mathbf{x}_2 - \boldsymbol{\mu}_2] \right) \right) \\ &\propto \exp \left(-\frac{1}{2} \left([\mathbf{x}_1 - \boldsymbol{\mu}_1]^T \Sigma_{11}^{-1} \Sigma_{21}^T \Lambda^{-1} + [\mathbf{x}_2 - \boldsymbol{\mu}_2]^T \Lambda^{-1} \right) \left(\Sigma_{21} \Sigma_{11}^{-1} [\mathbf{x}_1 - \boldsymbol{\mu}_1] + [\mathbf{x}_2 - \boldsymbol{\mu}_2] \right) \right) \\ &\propto \exp \left(-\frac{1}{2} \left(\left[\mathbf{x}_2 - \underbrace{\boldsymbol{\mu}_2 + \Sigma_{21} \Sigma_{11}^{-1} [\mathbf{x}_1 - \boldsymbol{\mu}_1]}_{=\boldsymbol{\mu}_{2|1}} \right]^T \Lambda^{-1} \left[\mathbf{x}_2 - \underbrace{\boldsymbol{\mu}_2 + \Sigma_{21} \Sigma_{11}^{-1} [\mathbf{x}_1 - \boldsymbol{\mu}_1]}_{=\boldsymbol{\mu}_{2|1}} \right] \right) \right) \end{aligned}$$

Now, we see that the conditional probability takes the form of a normal distribution.

$$p(\mathbf{x}_2|\mathbf{x}_1) \propto \exp\left(-\frac{1}{2}\left([\mathbf{x}_2 - \boldsymbol{\mu}_{2|1}]^T \Lambda^{-1} [\mathbf{x}_2 - \boldsymbol{\mu}_{2|1}]\right)\right),$$

where the mean and covariance of the distribution are defined by

$$\boldsymbol{\mu}_{2|1} = \boldsymbol{\mu}_2 + \Sigma_{21}\Sigma_{11}^{-1}(\mathbf{x}_1 - \boldsymbol{\mu}_1), \quad (\text{A4})$$

$$\Lambda = \Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{21}^T. \quad (\text{A5})$$

B Figures and Tables

B.1 Figures

This appendix contains the prediction vs. test value figures for the RBF kernel and both the uniform grain data set [B1](#) and the grain boundary data set [B2](#).

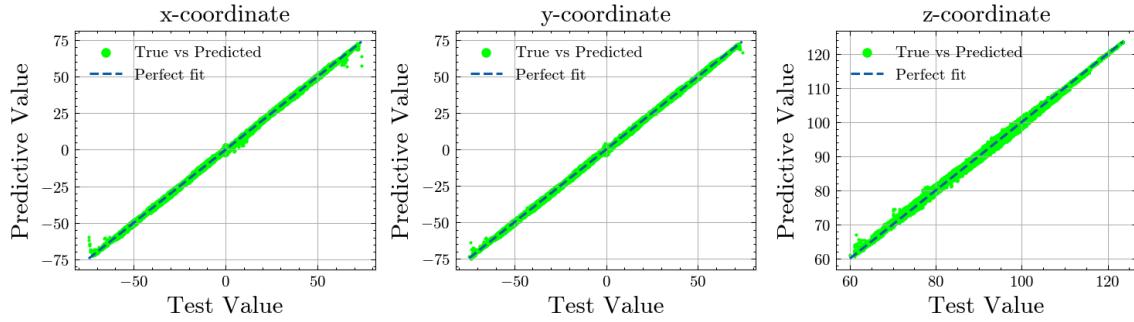


Figure B1: Test values vs. predictions for the RBF kernel and the uniform grain data set.

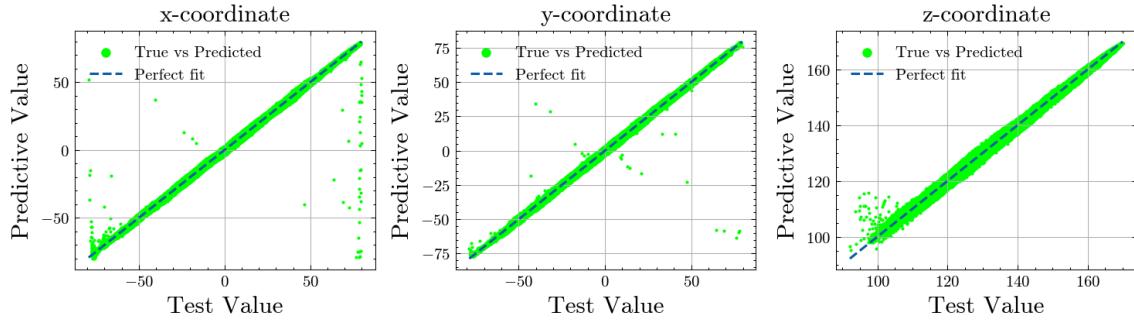


Figure B2: Test values vs. predictions for the RBF kernel and the grain boundary data set.

B.2 Tables

In this appendix, the root mean squared error (RMSE) metrics and the optimal hyperparameters are presented for both the uniform grain data set (tables [B1](#) and

[B2](#)) and the grain boundary data set tables ([B3](#) and [B4](#)). Notably, we include the raw hyperparameters learned from normalized input data and the scaled versions that might offer more insight.

Table B1: RMSE metrics for the uniform grain data set.

Models	Relative metrics (%)			Absolute metrics (nm)		
	RMSE _x Range	RMSE _y Range	RMSE _z Range	RMSE _x	RMSE _y	RMSE _z
RBF	0.423	0.423	0.578	0.628	0.627	0.366
Matérn	0.398	0.402	0.531	0.591	0.597	0.337

Table B2: Optimal hyperparameters for the uniform grain data set.

Models	Raw hyperparameters						
	l_1	l_2	l_3	μ_1	μ_2	μ_3	σ
RBF	0.245	0.243	9.890	-0.011	0.008	-0.697	0.0001
Matérn	0.723	0.710	11.0	-0.0079	0.0014	-1.549	0.0001
Models	Scaled hyperparameters						
	l_1	l_2	l_3	μ_1	μ_2	μ_3	σ
RBF	1.488	1.477	5.707×10^5	0.556	-0.294	90.745	0.008
Matérn	4.395	4.314	6.351×10^5	0.692	-0.524	78.143	0.008

Table B3: RMSE metrics for the grain boundary data set.

Models	Relative metrics (%)			Absolute metrics (nm)		
	RMSE _x Range	RMSE _y Range	RMSE _z Range	RMSE _x	RMSE _y	RMSE _z
RBF	0.912	0.647	0.820	1.451	1.023	0.639
Matérn	0.796	0.650	0.795	0.886	0.714	0.337

Table B4: Optimal hyperparameters for the grain boundary data set.

Models	Raw hyperparameters						
	l_1	l_2	l_3	μ_1	μ_2	μ_3	σ
RBF	0.349	0.310	10.452	0.108	-0.026	-0.667	0.0003
Matérn	0.900	0.842	11.441	0.536	0.109	-1.805	0.0003
Models	Scaled hyperparameters						
	l_1	l_2	l_3	μ_1	μ_2	μ_3	σ
RBF	1.868	1.722	1.510×10^6	9.290	3.609	136.418	0.024
Matérn	4.824	4.676	1.653×10^6	24.789	8.375	119.662	0.024

C Source Code

The source code for this project and other related research can be accessed by Aalto personnel through this [link](#).