IPP-Report

MAX-PLANCK-INSTITUT
FÜR PLASMAPHYSIK

Nikolai Kanter

Assessment of 2D Gaussian Process Regression for spatio-temporal evolution of plasma parameters

IPP 2024-18
August 2024

BACHELOR THESIS

Assessment of 2D Gaussian Process Regression for spatio-temporal evolution of plasma parameters

Fakultät II – Zentrum für Astronomie und Astrophysik
Technische Universität Berlin
Prof. Dr. Robert Wolf
Prof. Dr. Wolf-Christian Müller

by
Nikolai Kanter
Matr.-Nr. 0409199

Berlin
May 2024

Eigenständigkeitserklärung

Hiermit versichere ich, dass ich die vorliegende Arbeit eigenständig ohne Hilfe Dritter und ausschließlich unter Verwendung der aufgeführten Quellen und Hilfsmittel angefertigt habe. Alle Stellen die den benutzten Quellen und Hilfsmitteln unverändert oder sinngemäß entnommen sind, habe ich als solche kenntlich gemacht.

Sofern generische KI-Tools verwendet wurden, habe ich Produktnamen, Hersteller, die jeweils verwendete Softwareversion und die jeweiligen Einsatzzwecke (z.B. sprachliche Überprüfung und Verbesserung der Texte, systematische Recherche) benannt. Ich verantworte die Auswahl, die Übernahme und sämtliche Ergebnisse des von mir verwendeten KI-generierten Outputs vollumfänglich selbst.

Die Satzung zur Sicherung guter wissenschaftlicher Praxis an der TU Berlin vom 30. Mai 2023, `https://www.static.tu.berlin/fileadmin/www/10002457/K3-AMBl/Amtsblatt_2023/Amtliches_Mitteilungsblatt_Nr._16_vom_30.05.2023.pdf`,
habe ich zur Kenntnis genommen.

Ich erkläre weiterhin, dass ich die Arbeit in gleicher oder ähnlicher Form noch keiner anderen Prüfungsbehörde vorgelegt habe.

Berlin, 8. Mai 2024

N. Kanter

Abstract

This thesis is concerned with the construction of a two-dimensional Gaussian Process (2D GP) for fitting spatial-temporal evolutions of plasma parameters. The 2D GP regression method is applied to artificial data and experimental data with strong temporal variations. A multiplication of two squared exponential (SE) kernels is used to allow for different length scales along each dimension. The spatial hyperparameters (i.e. parameters of the kernel, which determine length and vertical scale) are optimized by maximizing the marginal likelihood. The impact of noise and sample size is studied. The 2D GP performs better than joining multiple independent 1D GP reconstructions of each time-slice, due to the consideration of temporal correlations. For the optimization of the temporal length scale a new physics informed approach is developed which calculates a time dependent length scale $l_t(t)$ using the temporal derivative of the line integrated electron density $\partial_t \frac{1}{L} \int n_e \mathrm{d}L$. It is observed that strong variations in $l_t(t)$ lead to oscillations in the 2D GP reconstruction. Due to the oscillations, a fixed length scale equivalent to the temporal resolution of the measurements $l_t = \Delta t$ performes better than the time dependent length scale $l_t(t)$. The reconstruction of partial derivatives and boundary conditions, such as a vanishing gradient in the plasma center, is currently not included and will be addressed in future work.

Abstract

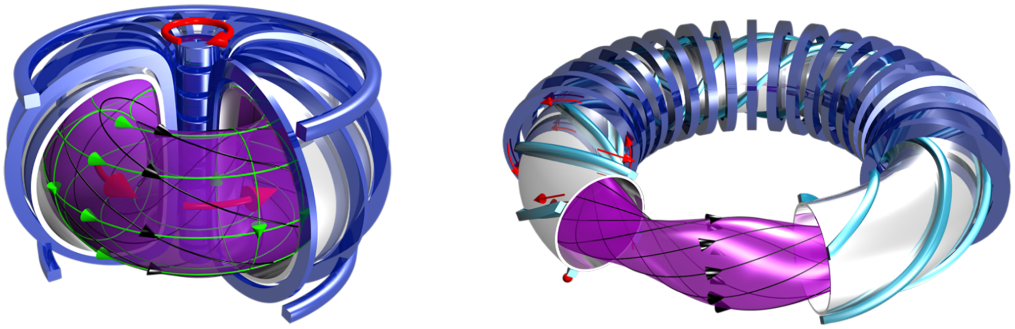
Diese Arbeit beschäftigt sich mit der Konstruktion eines zweidimensionalen Gauß Prozesses (2D GP) für die Regression räumlich-zeitlicher Entwicklungen von Plasmaparametern. Die 2D GP Regressionsmethode wird auf künstliche Daten und experimentelle Daten mit starken zeitlichen Änderungen angewendet. Ein Produkt von zwei quadratisch exponentiellen Kernel wurde verwendet, um verschiedene Korrelationslängen für jede Dimension zu ermöglichen. Die räumlichen Hyperparameter (d.h. Parameter des Kernels, welche die horizontale und vertikale Skala bestimmen) wurden durch Maximierung der marginal likelihood optimiert. Der Einfluss von Rauschen und Anzahl der Daten wird untersucht. Der 2D GP erzielt bessere Ergebnisse als die Verbindung von mehreren unabhängigen 1D Rekonstruktionen eines jeden Zeitpunktes, da zeitliche Korrelationen beachtet werden. Für die Optimierung der zeitlichen Längenskala wurde ein neuer physikalischer Ansatz entwickelt, der die zeitabhängige Längenskala $l_t(t)$ mithilfe der zeitlichen Ableitung der linienintegrierten Dichte $\partial_t \frac{1}{L} \int n_e \mathrm{d}L$ berechnet. Es wird beobachtet, dass starke Variationen in $l_t(t)$ zu Oszillationen in der 2D GP Rekonstruktion führen. Aufgrund der Oszillationen erzielt die konstante Längenskala, gleich der zeitlichen Auflösung der Messung $l_t = \Delta t$, bessere Ergebnisse als die zeitabhängige Längenskala $l_t(t)$. Die Rekonstruktion von partiellen Ableitungen und Randbedingungen, wie ein verschwindener Gradient im Plasmazentrum, ist aktuell nicht eingearbeitet und wird in der Zukunft adressiert.

Contents

1 Introduction

To meet the increasing demand for energy, it is advantageous to have multiple energy resources that can replace fossil fuels [1]. In addition to green fuels such as solar and wind energy, another option for energy production is the fusion of nuclei, the suns process of generating energy [2]. Fusion is the process by which two light nuclei combine to form a heavier nucleus. The reaction with the highest reactivity $\langle \sigma v \rangle$ is the fusion of deuterium (D = ${}^{2}_{1}\mathrm{H}$) and tritium (T = ${}^{3}_{1}\mathrm{H}$) [3].

In order to fuse nuclei, they must possess high kinetic energies to overcome the repulsive Coulomb force. One method to realize fusion is by heating the fuel to a plasma state, creating an ionized gas with free ions and electrons. Temperatures of $\sim 13\,\mathrm{keV}$ have to be achieved for the maximum reaction rate $R \sim \langle \sigma v \rangle / T^2$, which is equivalent to ~ 150 million Kelvin [3]. To prevent energy loss of the plasma, contact with any other matter has to be avoided. Additionally, a reduction in transport is necessary to achieve a high ion temperature T_i with a reasonable heating power, which is done by confining the plasma in a helical magnetic field. The tokamak and stellarator are the two main construction types for fusion reactors with magnetic confinement, with the schematic structures shown in Fig. 1. In order to maintain the plasma particles in a confined space, the magnetic field lines are arranged in a toroidal configuration. This utilizes the property of plasma particles to follow magnetic field lines. The curvature of the magnetic field lines would lead to a seperation of charges (∇B-drift). As a result an electrical field would be created, which would cause the plasma particles to drift outwards ($E \times B$-drift). This can be prevented by combining a poloidal and toroidal magnetic field, resulting in a helical magnetic field.



(a) (b)

Figure 1: (a) depicts a schematic structure of tokamak. (b) displays a schematic structure of a stellarator. The magnetic coils are shown in blue with the plasma vessel in grey. The red arrows show the direction of the currents, the green arrows show the poloidal and toroidal parts of the magnetic field, resulting in the magnetic field lines in black. The violet surface shows a flux surface of the plasma. The images were provided by [4].

In order to gain better understanding of fusion in magnetic confinement and building potential fusion reactors various experiments are conducted worldwide. Two ongoing stellarator experiments are the *Wendelstein 7-X* (W7-X) of the *Max-Planck-Institute for Plasmaphysics* (IPP) in Greifswald (Germany) and the *Large Helical Device* (LHD) of the *National Institute for Fusion Science* (NIFS) in Toki (Japan). More precisely, LHD is a heliotron with a major radius of $R = 3.9\,\mathrm{m}$ and an average plasma radius of $a = 0.6\,\mathrm{m}$ which utilizes a pair of intertwined helical toroidal coils as shown in Fig. 2. It uses additional vertical field coils as well as ten pairs of local island divertor [5]. Due to its many years of succesfull operation, LHD provides a large

repository for experimental data relating to spatio-temporal evolutions of plasma parameters, which will be utilized later in this thesis.

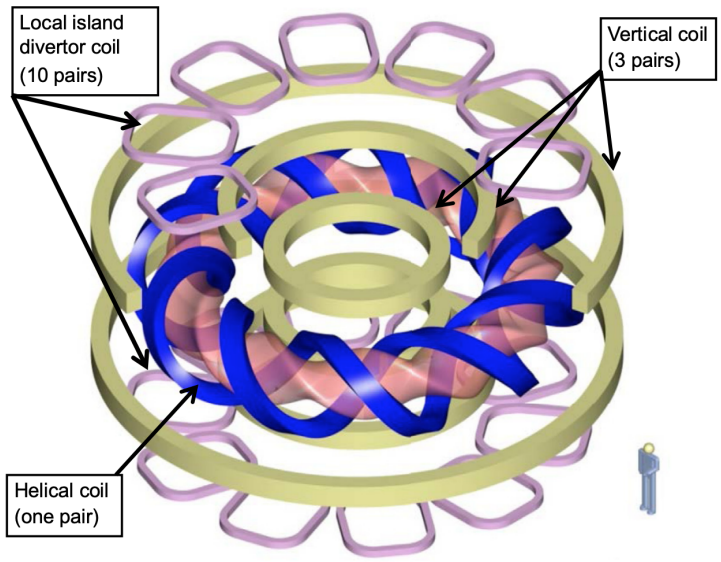


Figure 2: Schematic depiction of the coil system of LHD. The pair of helical coils are shown in blue and three pairs of circular poloidal coils are illustrated in light yellow. There are also 10 pairs of divertor coils. The plasma flux surface is shown in pink. Courtesy of [5].

With the aim of reaching high temperatures of $\sim 13\,\mathrm{keV}$ the plasma must be heated, which requires a large amount of power. The triple product $nT\tau_E$ of particle density n, temperature T and energy confinement time τ_E describes the fusion performance [6]. The Lawson criterion $nT\tau_E > 0.5\,\mathrm{MJs/m^3}$ [7] determines a minimum value for the triple product, where the power losses are compensated by the power of the alpha particles resulting from the fusion reaction. τ_E is the characteristic loss time of the energy and depends on energy transport in the plasma. Power is lost through conduction and convection, as well as radiative power losses in the form of bremsstrahlung [8]. The plasma density and temperature can be externally controlled by adjusting the heating power and gas influx, whereas τ_E cannot be directly controlled. The desired temperature for DT-fusion is $T \sim 13\,\mathrm{keV}$ with a plasma density of $n \sim 10^{-20}\,\mathrm{m^{-3}}$. To fulfill the mentioned Lawson criterion, with the determined plasma density and temperature, τ_E must reach 3 s at minimum [7]. This implies that the energy needs to be confined for more than 3 s in order to produce a power output equal to the input. Recent studies show that one way of increasing the confinement time is the injection of hydrogen pellets into the plasma core [9]. It is also a method for refueling the plasma, which is needed for a fusion power plant. However, excessive pellet fueling may lead to termination of the plasma [10]. This shows the significance of improving confinement and extending the energy confinement time, while avoiding plasma termination.

Due to the dependence of τ_E on transport mechanics, it is crucial to develop a thorough understanding of the transport mechanics and turbulence in the plasma. The temperature and density gradients control instabilities of the plasma [11]. Considering this aspect, the temperature gradient ∇T is of interest as it influences the mechanism of particle transport, whereas the density gradient controls diffusion. The diffusion coefficients are significant parameters, describing the

rate at which plasma particles diffuse outwards and arise in the particle flux $\mathbf{\Gamma}$

$$\mathbf{\Gamma} = -D\nabla n \tag{1.1}$$

with diffusivity D. The temporal evolution of particle density n is described by the continuity equation

$$\frac{\partial n}{\partial t} = -\nabla \cdot \mathbf{\Gamma} + S$$

where $S = S(\boldsymbol{x}, t)$ represents the source through ionisation and recombination. The source can also describe the fueling of the plasma through pellet injection. Using cylindrical coordinates and substituting $\mathbf{\Gamma}$ with Eq. (1.1) we obtain the inhomogenous partial differential equation

$$\frac{\partial n}{\partial t} = \frac{1}{r}\frac{\partial}{\partial r}\left(rD\frac{\partial n}{\partial r}\right) + S. \tag{1.2}$$

It is evident that the spatial and temporal derivatives of the particle density are needed in order to calculate the diffusion coefficients. The diffusivity D is often calculated numerically due to its complexity [12]. It has to be noted that a number of differential equations can be derived for different scenarios, taking into account a variety of different effects. E.g. a similar differential equation can be derived for the heat flux $\boldsymbol{q} = -n\chi\nabla T$ with the thermal conductivity coefficient χ [7].

Experimental data for plasma parameters are often noisy, making it necessary to fit the data. Another reason for fitting the data is the measurement at distinct locations and times, eventhough in Eq. (1.2) the partial derivatives for all r and t are needed. A regression method is the Gaussian Process (GP) [13]. In contrast to parametric regression, a GP is a non-parametric probabilistic regression method. Thus, it is not necessary to specify a parametric function that describes the temperature and density profiles. The GP is widely used in plasma physics for fitting noisy plasma profiles of fusion experiments at a certain toroidal position for specific times resulting in a one-dimensional fit (1D GP) [14, 15]. Furthermore, it can be used for reconstructing partial derivatives, which could be used for the calculation of diffusion coefficients instead of using numerical methods. It also enables the reconstruction of higher derivatives, such as the second derivatives, which are needed for the description of turbulence [16]. The probabilistic nature of the GP enables a simple way to estimate uncertainties [17]. This also applies to the uncertainty estimation of partial derivatives. Examplary applications of a 2D GP is the reconstruction of magnetic fields $\boldsymbol{H}(x, y, z)$ for two-dimensional spatial training data [18] and the regression of experimental edge plasma evolution [19].

The aim of this work is to develop a method to fit the spatial-temporal evolution of plasma parameters, such as the temperature $T(r, t)$ and particle density $n(r, t)$, with a two-dimensional GP fit (2D GP) and to work out advantages and disadvantages. The overall goal of reconstructing spatial-temporal plasma parameters is to deliver input for studies of transport dynamics and instabilities. It is assumed that by adding a temporal dimension the temporal correlation between data points will be taken into account. This may help to improve regressions of termination processes, which can be faster than the temporal resolution of the measurement. Additionally, taking the temporal dimension into account enables the calculation of partial derivatives w.r.t. time. The spatial and temporal partial derivatives are needed in order to determine diffusion coefficients, as well as for the classification of transport mechanics.

2 Theoretical Background

2.1 Gaussian Processes for regression

The GP is a non-parametric probabilistic method used for regression and is defined as follows:

Definition: *A Gaussian Process is a collection of random variables, any finite number of which have a joint Gaussian distribution.* [20]

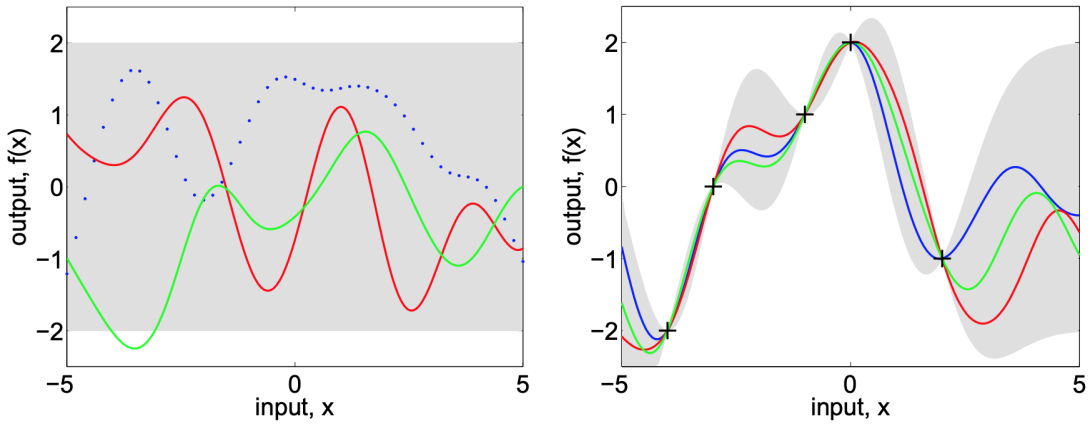
It is based on Bayes' theorem

$$\overbrace{p(\boldsymbol{\theta}|\boldsymbol{y},\boldsymbol{x})}^{\text{posterior}} = \frac{\overbrace{p(\boldsymbol{y}|\boldsymbol{x},\boldsymbol{\theta})}^{\text{likelihood}} \times \overbrace{p(\boldsymbol{\theta}|\boldsymbol{x})}^{\text{prior}}}{\underbrace{p(\boldsymbol{y}|\boldsymbol{x})}_{\text{evidence}}}$$

where $p(\boldsymbol{y}|\boldsymbol{x})$ (evidence) denotes the marginal likelihood

$$p(\boldsymbol{y}|\boldsymbol{x},\boldsymbol{\theta}) = \int p(\boldsymbol{y}|\boldsymbol{x},\boldsymbol{\theta})p(\boldsymbol{\theta}|\boldsymbol{x})\mathrm{d}\boldsymbol{\theta}.$$

The matrix $\boldsymbol{x}$ represents the independent variable and $\boldsymbol{y}$ the values of the dependent variable. The observations (experimental data) $\mathcal{D} = (\boldsymbol{x}, \boldsymbol{y})$ are referred to as training data and will be used for conditioning. The vector $\boldsymbol{\theta}$ contains the so called hyperparameters, which will be eludicated later in this section. The prior shown as the grey area in Fig. 3a can be considered as the knowledge of the process without regarding any data. It defines the range of possible outputs and is restrained when considering data (conditioning). The prior conditioned on the data is called posterior and is shown in Fig. 3b. The marginal likelihood serves the purpose of a normalizing constant [20].



(a) (b)

Figure 3: (a) depicts three arbitrary functions (samples) from the prior. The functions can be obtained by joining a large number of output values $\boldsymbol{y}$. (b) shows three functions from the posterior. These functions are prior functions conditioned on five noisy data points. The shaded area depicts the 95% confidence region, respectively. The images are taken from [20].

When considering the values of the training data $\boldsymbol{y} = f(\boldsymbol{x}) + \boldsymbol{\varepsilon}$ with normally distributed noise $\boldsymbol{\varepsilon} \sim \mathcal{N}(0, \varepsilon_n^2 \boldsymbol{I})$ at the position $\boldsymbol{x}$ as the random variables, the GP can be viewed as a distribution over functions. The resulting distribution

$$f(\boldsymbol{x}) \sim \mathcal{GP}(m(\boldsymbol{x}), k(\boldsymbol{x}, \boldsymbol{x'}))$$

is therefore fully described by its mean function $m(\boldsymbol{x})$ and covariance function $k(\boldsymbol{x}, \boldsymbol{x'})$. The new input values $\boldsymbol{x^*}$ and the predictions $f^*(\boldsymbol{x^*})$ are referred to as test data. In general, the training data and test data can be n-dimensional. In this work, a GP fit with one-dimensional training and test data (e.g. space x) is referred to as a 1D GP, whereas a GP fit with two-dimensional training and test data (e.g. space x and time t) is referred to as a 2D GP. For making predictions $f^*(\boldsymbol{x^*})$ the prior is conditioned on the training data as shown in Fig. 3b. Consequently, the joint distribution of $\boldsymbol{y}$ and $\boldsymbol{y^*} = f^*(\boldsymbol{x^*})$ is

$$\begin{bmatrix} \boldsymbol{y} \\ \boldsymbol{y^*} \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} m(\boldsymbol{x}) \\ m(\boldsymbol{x^*}) \end{bmatrix}, \begin{bmatrix} \boldsymbol{K_{xx}} + \varepsilon_n^2 \boldsymbol{I} & \boldsymbol{K_{xx^*}} \\ \boldsymbol{K_{x^*x}} & \boldsymbol{K_{x^*x^*}} \end{bmatrix} \right)$$

where $\boldsymbol{K}(\boldsymbol{x}, \boldsymbol{x}) = \boldsymbol{K_{xx}}$ denotes the covariance matrix between training data and $\boldsymbol{K}(\boldsymbol{x}, \boldsymbol{x^*}) = \boldsymbol{K_{xx^*}}$ with $(\boldsymbol{K_{xx^*}})^{\mathrm{T}} = \boldsymbol{K_{x^*x}}$ is the covariance matrix between the combination of training and test data. The GP reconstruction is then computed by the mean of the posterior

$$\boldsymbol{y^*} = \boldsymbol{K_{x^*x}}(\boldsymbol{K_{xx}} + \varepsilon_n^2 \boldsymbol{I})^{-1} \boldsymbol{y} \tag{2.1}$$

with the posterior covariance matrix

$$\mathrm{cov}(\boldsymbol{y^*}) = \boldsymbol{K_{x^*x^*}} - \boldsymbol{K_{x^*x}}(\boldsymbol{K_{xx}} + \varepsilon_n^2 \boldsymbol{I})^{-1} \boldsymbol{K_{xx^*}}.$$

The predictions uncertainty (variance) can be optained from the diagonal elements of the covariance matrix

$$\Delta \boldsymbol{y^*} = \mathrm{diag}(\mathrm{cov}(\boldsymbol{y^*})). \tag{2.2}$$

Eq. (2.1) and Eq. (2.2) describe the regression result: For any $\boldsymbol{x^*}$ (particularly for x-values aside the $\boldsymbol{x}$ of the training data), $\boldsymbol{y^*}$ and $\Delta \boldsymbol{y^*}$ are the most likely prediction and its uncertainty. The values $\boldsymbol{K}_{ij}$ of the covariance matrix are calculated by the covariance function $k(\boldsymbol{x}, \boldsymbol{x'})$ which is referred to as a kernel and determines the correlation between a pair of variables. That is, if the inputs x and x' are similar, it is assumed that the outputs $f(x)$ and $f(x')$ are also similar [21]. There are a lot of different kernels, which need to be specified according to the desired model. A typical standard kernel is the squared exponential (SE) kernel

$$k_{\mathrm{SE}}(x, x') = \sigma^2 \exp\left(-\frac{(x - x')^2}{2l^2} \right). \tag{2.3}$$

The parameters σ and l are called hyperparameters which need to be specified for the GP. The hyperparameters are not like the parameters in parametric regression as they only impact the kernel and thereby determine properties of the underlying function. While σ defines the vertical scale of the underlying function, l sets the characteristic length scale of the underlying function. Therefore, the hyperparameters have an influence on the rigidity of the fit. The SE kernel is infinitely differentiable, leading to smooth GP reconstructions. It is this property that makes Gaussian Processes a convenient tool for regression.

Hyperparameters can either be fixed or optimized. The standard method for optimizing hyperparameters is by maximizing the log marginal likelihood

$$\log p(\boldsymbol{y}|\boldsymbol{x}, \boldsymbol{\theta}) = -\frac{1}{2} \boldsymbol{y}^T (\boldsymbol{K_{xx}} + \varepsilon_n^2 \boldsymbol{I})^{-1} \boldsymbol{y} - \frac{1}{2} \log |\boldsymbol{K_{xx}} + \varepsilon_n^2 \boldsymbol{I}| - \frac{N}{2} \log(2\pi)$$

where the first term gives the goodness of the fit, the second term is a penalty term for too complex models and the last term functions as a normalizing constant [22]. For finding the maximum of $\log p(\boldsymbol{y}|\boldsymbol{x}, \boldsymbol{\theta})$ the partial derivatives $\partial k/\partial l$ and $\partial k/\partial \sigma$ for each hyperparameter are needed.

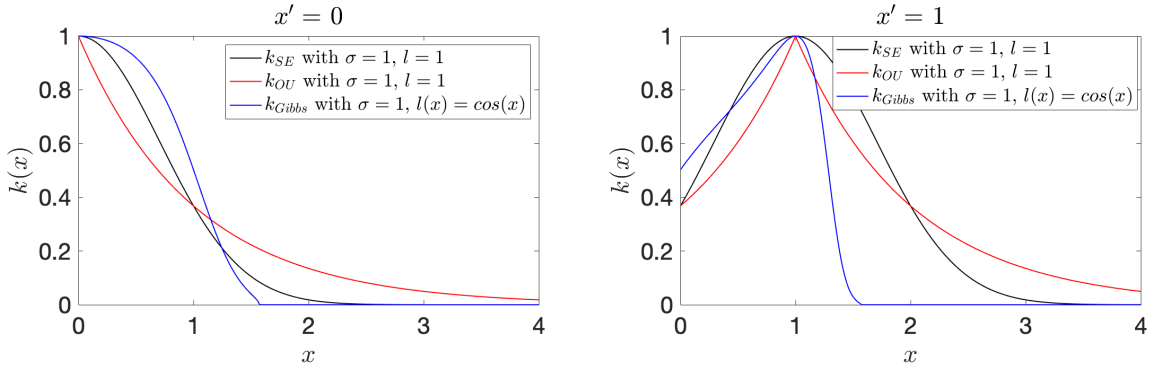
For some applications the priors obtained from the SE kernel might be too smooth, for which an Ornstein-Uhlenbeck (OU) kernel

$$k_{\mathrm{OU}}(x, x') = \sigma \exp\left(-\frac{|x - x'|}{l}\right)$$

could be used instead [20]. Gibbs proposed the non-stationary kernel

$$k_{\mathrm{Gibbs}}(\boldsymbol{x}, \boldsymbol{x'}) = \sigma \prod_{d=1}^{D} \left(\frac{2 l_d(\boldsymbol{x}) l_d(\boldsymbol{x'})}{l_d^2(\boldsymbol{x}) + l_d^2(\boldsymbol{x'})}\right)^{1/2} \exp\left(-\sum_{d=1}^{D} \frac{(x_d - x'_d)^2}{l_d^2(\boldsymbol{x}) + l_d^2(\boldsymbol{x'})}\right) \tag{2.4}$$

for D dimensions d and varying length scales $l_d(\boldsymbol{x})$ respectively [23]. The prefactor is to assure the positive definite property of a kernel. Note that for $D = 1$ and a constant $l(\boldsymbol{x}) = l(\boldsymbol{x'})$ the SE kernel in Eq. (2.3) is obtained. A non-stationary Gibbs kernel for fitting plasma profiles is used in [24], where the length scale is large in the plasma center and shorter at the plasma borders. The three presented kernels are compared in Fig. 4 for two different values of x', i.e. the kernels illustrate how strong each value of x correlates to x'. The Gibbs kernel in Fig. 4 uses the cosine function $l(x) = cos(x)$ for varying length scales. This function was selected without any specific reasoning. It can be seen that the varying length scale has a large effect on the Gibbs kernel and looks different for different values of x' as well as for different functions of $l(x)$. Whereas the SE and OU kernels are symmetric and are shifted according to the value of x'.



(a) (b)

Figure 4: Comparison of the SE, OU and Gibbs kernels evaluated at the points (a) $x' = 0$ and (b) $x' = 1$.

Furthermore, applying a linear operator $\hat{L}$ on a GP yields another GP [20]. For a zero mean GP, it can be written

$$\hat{L} f(\boldsymbol{x}) \sim \mathcal{GP}(0, \hat{L} k(\boldsymbol{x}, \boldsymbol{x'}) \hat{L}')$$

where $\hat{L}'$ acts from the right site w.r.t. x' [25]. Therefore, application of a linear operator can be expressed by applying the linear operator to the kernel. As the derivation is a linear operator this can be exploited to reconstruct partial derivatives along with the GP [20]. Additionally, the inclusion of linear operators enables the inclusion of boundary conditions. In plasma physics it is used to ensure a vanishing gradient in the center of plasma profile fits [14]. The application of linear operators, such as derivatives, allows for the construction of kernels that are solutions of (stochastic) partial differential equations [25, 26].

3 Methods

The goal of this work is to develop a 2D GP for the regression of the spatio-temporal evolution of plasma parameters. To test and evaluate this method, artificial data were created. Artificial data is particularly useful when constructing a new physics informed temporal hyperparameter to verify calculations. It is beneficial because it ensures that the calculations are proceeding as intended. Here, the artificial data is a Gaussian

$$f(x,t) = \frac{A(t)}{\sigma} \cdot e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \tag{3.1}$$

with the time dependent amplitude $A(t)$, standard deviation $\sigma = 0.3$ and mean $\mu = 0$. The choice of the artificial data being a Gaussian is not influenced by the Gaussian form of the SE kernel function in Eq. (2.3). It was chosen to roughly represent the radial temperature profile of experimental measurements. The artificial data consists of the radial dimension x, which represents the effective radius r_{eff} of the plasma and a temporal dimension t. It was created for an equidistant space interval of $x \in [-0.8, 0.8]$ with N_x points and an equidistant time interval of $t \in [0, 3]$ with N_t points to reflect both the typical effective radius of stellarators like W7-X and LHD, and a typical plasma discharge duration, respectively. Thereby, the time dependent amplitude is divided in three sections

$$A(t) = \begin{cases} 0.25t + 0.5 & , 0 < t < 2 \\ -2t + 1 & , 2 < t < 2.5 \\ 0 & , 2.5 < t < 3. \end{cases}$$

The linear increase of amplitude in the first two seconds with an abrupt decline to zero simulates the increase of electron or ion temperature of a plasma with an abrupt termination to test the effectiveness of the 2D GP for fast changing data. The termination could be caused by excessive pellet fueling. The artificial data for $\sigma = 0.3$ and $\mu = 0$ is shown in Fig. 5.

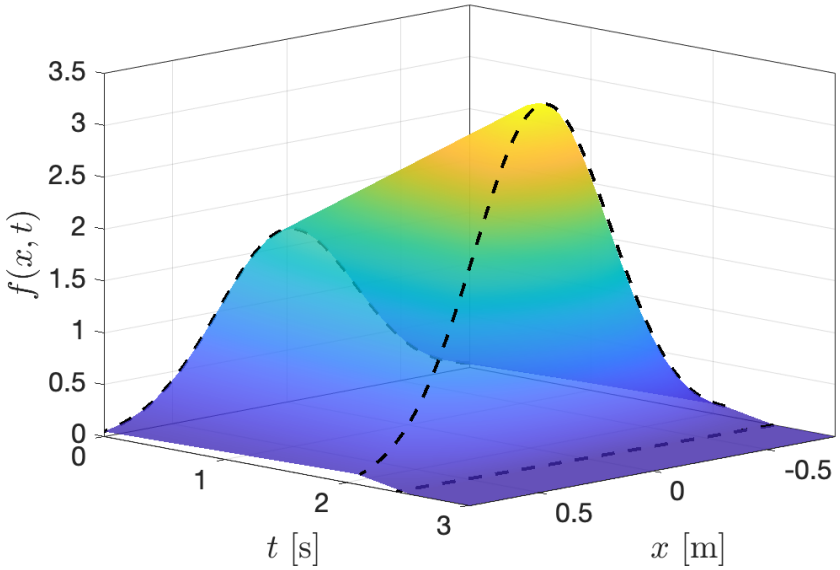


Figure 5: Artificial Gaussian data $f(x,t)$ with a time dependent amplitude $A(t)$ simulating the electron or ion temperatures of a plasma with a linear increase and an abrupt termination of the plasma. The dotted lines illustrate the profile at the times $[0, 2, 2.5]$ s.

3.1 One-dimensional Gaussian Process

In this section the already established usage of 1D GP will be outlined. For the artificial data, the data from Eq. (3.1) at $t = 0\,\mathrm{s}$ is used:

$$f(x) = \frac{A(0)}{\sigma} \cdot e^{-\frac{1}{2}\left(\frac{x}{\sigma}\right)^2} = \frac{0.5}{\sigma} \cdot e^{-\frac{1}{2}\left(\frac{x}{\sigma}\right)^2} \tag{3.2}$$

resulting in a N_x dimensional vector of training data. As noise a constant value of $\varepsilon = 0.01$ is chosen. In Fig. 6 the impact of the hyperparameters l_x and σ_x can be seen. In the first row, σ_x was kept constant at $\sigma_x = 1$, while l_x is varied $l_x = [0.01, 0.1, 1, 5]$ (left to right). In the second row the correlation length is kept constant at $l_x = 1$, while varying $\sigma_x = [0.01, 0.1, 1, 10]$. The shaded area represents the 95% confidence interval (i.e. two standard deviations) calculated from the fit uncertainty as $\pm 1.96\sqrt{\Delta f^*}$ where Δf^* is calculated using Eq. (2.2). It can be seen, that the uncertainty is small at the points where training data is given. For $l_x = 0.01$ the GP overfits the training data. The best fit is achieved for $l_x = 0.1$. A correlation length too large for the training data smoothes the function as it cannot follow the changes in $f(x)$. A correlation length smaller than the distance of the training data leads to overfitting as seen in Fig. 6. The small correlation length allows the GP to find reconstructions with fast changing test data between the given training data. The hyperparameter σ_x decides the variance of the underlying function. Thus the GP cannot cover the whole range of training data for small σ_x. It appears that it is not important whether σ_x is too large as the last two fits in the second row are similar. The comparison of different values for the hyperparameter shows the importance of hyperparameter selection.

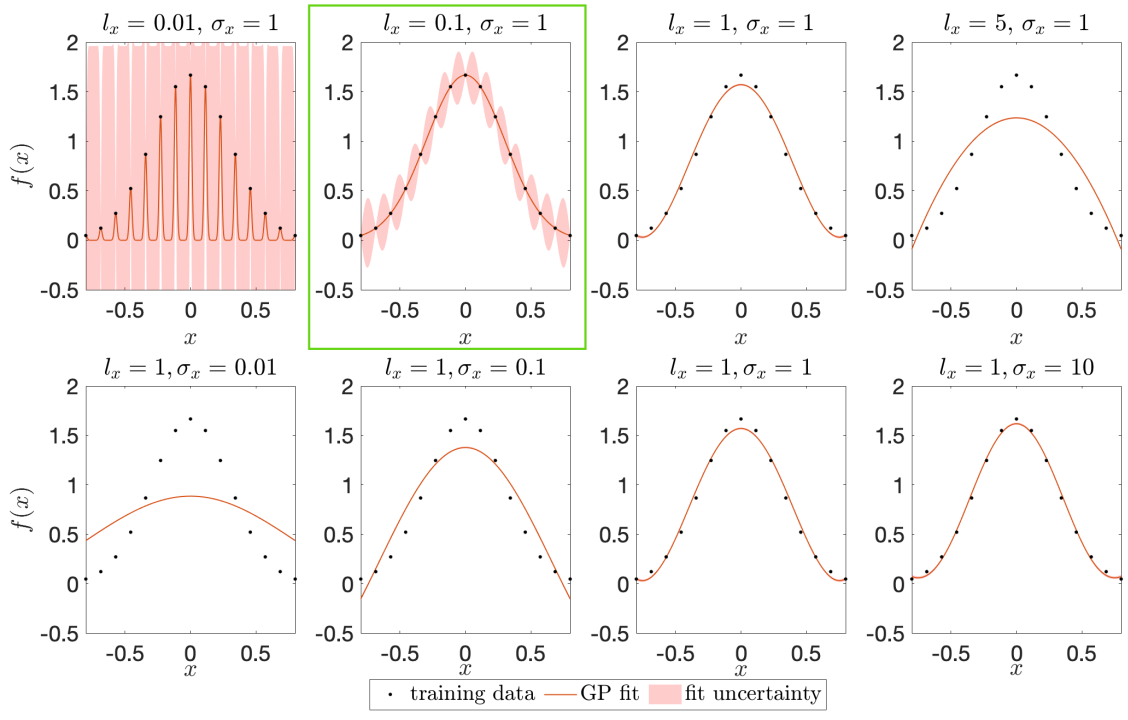


Figure 6: Impact of hyperparameters l_x and σ_x of the SE kernel for artificial Gaussian data $f(x)$ with a constant noise value of $\varepsilon = 0.01$. First row shows the variation of l_x, while the second row displays the effect of varying σ_x. The shaded areas depict the 95% confidence interval. The green box indicates the best fit of the used fixed hyperparameter.

3.2 Two-dimensional Gaussian Process

The difference between 2D GP and 1D GP lies in the dimensions of both training and test data. An additional temporal dimension with N_t points is added compared to the 1D GP. The training data then consists of $N_x \times N_t$ inputs. Thus, a new kernel is needed. This new multidimensional kernel can be constructed by multiplying two one-dimensional SE kernels [20], i.e.

$$k_{\mathrm{SE}}(\boldsymbol{x}, \boldsymbol{t}, \boldsymbol{x'}, \boldsymbol{t'}) = k_{\mathrm{SE},x}(\boldsymbol{x}, \boldsymbol{x'}) k_{\mathrm{SE},t}(\boldsymbol{t}, \boldsymbol{t'}) = \sigma_x^2 \exp\left(-\frac{1}{2}\left(\frac{x - x'}{l_x}\right)^2\right) \sigma_t^2 \exp\left(-\frac{1}{2}\left(\frac{t - t'}{l_t}\right)^2\right)$$

is a new kernel with four hyperparameters l_x, σ_x, l_t and σ_x, allowing for different correlation lengths for each dimension. The kernel above can be also derived from the Gibbs kernel in Eq. (2.4) with $D = 2$ representing one spatial and a temporal dimension with constant characteristic length scales, respectively.

3.2.1 Optimization of spatial hyperparameter

Here, the hyperparameter optimization for 2D GP consists of two steps. First, the optimization of the spatial hyperparameters l_x and σ_x is implemented by maximizing the marginal likelihood, as in the 1D GP. As seen in section 3.1 the hyperparameter l_x determines the correlation length in the spatial dimension, whereas σ_x defines the variance. As the training data is a $N_x \times N_t$ matrix the covariances would be stored in a covariance tensor $\boldsymbol{\mathcal{K}_{xx}}$ of size $(N_x \times N_t)^2$. For the purpose of maximizing the marginal likelihood, the inverse $\boldsymbol{\mathcal{K}_{xx}^{-1}}$ and determinant $|\boldsymbol{\mathcal{K}_{xx}}|$ of the covariance tensor are needed. For easier computation, the covariance tensor is reshaped, resulting in a matrix $\boldsymbol{K_{xx}}$ where the covariance matrices for each time-slice are aligned next to each other. However, when calculating the determinant $|\boldsymbol{K_{xx}}|$ the logarithm of the marginal likelihood diverges towards infinity. This is suspected to occur due to the calculation of the logarithm of the determinant of $\boldsymbol{K_{xx}}$. In general, the computational cost scales as $\mathcal{O}(N^3)$ [20], which can pose a problem for large training data size, as is the case with 2D GPs. In general, values of the covariance matrix are < 1, hence the determinant will be close to zero (singular), respectively

$$\lim_{N \to \infty} |\boldsymbol{K_{xx}}| = 0.$$

So it can be concluded that for large sizes of training data the determinant of $\boldsymbol{K_{xx}}$ approaches zero, therefore the logarithm diverges towards negative infinity. As a work around, the marginal likelihood was calculated for a single time-slice. Consequently, the covariance matrix is N_x^2 dimensional, similar to the 1D case, which significantly reduces the computational cost. Alternatively, [20] proposes the addition of a small multiple of the identity matrix for avoiding the convergence to zero. However, this approach and its effect on the marginal likelihood was not tested in this work and is left open for future research.

3.2.2 Optimization of temporal hyperparameter

The plasma parameter can vary strongly whith time, which leads to the idea of using a time dependent temporal hyperparameter $l_t = l(t)$, enabling the regression of fast and slow changes. Optimizing the temporal correlation length by maximizing the marginal likelihood would result in a constant correlation length, which cannot capture the different length scales. In this section, it will be attempted to construct a function $l(t)$, that enables a variation of the temporal correlation

length for the temporal variation of the plasma parameters. To ensure that the correlation length encaptures fast changes of the plasma parameter, the hyperparamater needs to be small for large changes and vice versa. The rate of change can be derived from the partial deriviation of the plasma parameter with respect to time.

In case of fitting the spatio-temporal evolution of the electron density n_e measured by Thomson scattering [27], the time derivative of the line integrated density $1/L \int n_e \mathrm{d}L$ will be needed. The line integrated density represents the averaged electron density and is measured with an interferometer [28]. Therefore, the hyperparameter is calculated by

$$\tilde{l}(t) = \left| \frac{\partial}{\partial t} \left(\frac{1}{L} \int n_e \, \mathrm{d}L \right) \right|^{-1} \tag{3.3}$$

with $L = 2a$ being the length of the laser path through the system. The maximum possible temporal correlation time τ will be defined by the energy confinement time τ_E as this parameter indicates how long energy is confined in the plasma [7]. Temporal changes of transport phenomena occur on this timescale. Therefore, the maximum of $l_t(t)$ is chosen to be

$$\max(l(t)) = 3\tau_E \tag{3.4}$$

for a stationary plasma. A stationary plasma is a plasma with constant energy, temperature and density. A good estimate for τ_E in W7-X and LHD is $\sim 100\,\mathrm{ms}$ [29, 30]. Therefore, in the following first implementation the maximum of $l_t(t)$ will be kept fixed to 300 ms. For the minimum of $l_t(t)$ the temporal resolution Δt of the measured data is chosen, i.e.

$$\min(l(t)) = \Delta t. \tag{3.5}$$

As seen in Fig. 6 a correlation length smaller than the temporal resolution would lead to overfitting of the reconstruction. For the considered data of electron density, the temporal resolution is $\Delta t \sim 30\,\mathrm{ms}$. The hyperparameter will be scaled using Eq. (3.4) and (3.5) resulting in

$$l(t) = \frac{3\tau_E - \Delta t}{\max(\tilde{l}(t))} \tilde{l}(t) + \Delta t.$$

It is important to note that Eq. (3.3) is not defined, when the partial time derivative is zero, i.e. if the plasma is stationary or if the evolution of the plasma parameter has an extremum. For a stationary plasma the hyperparameter should be large, to ensure a long correlation, while for extrema a small correlation length is needed. Furthermore, a stationary plasma or extremum was defined for 5% deviation of $\partial_t(1/L \int n_e \mathrm{d}L) = 0$. This results in the function

$$l(t) = \begin{cases} 3\tau_E & \text{, stationary} \\ \Delta t & \text{, extremum} \\ \frac{3\tau_E - \Delta t}{\max(\tilde{l}(t))} \tilde{l}(t) + \Delta t & \text{, else} \end{cases}$$

where $\tilde{l}(t)$ is described by Eq. (3.3). With this construction $l_t(t) > 0$ for all t, which is why the prefactor in the Gibbs kernel in Eq. (2.4) is ignored for the first application.

4 Results

4.1 Application to noisy trivial data

For the analysis of the 2D GP it was first applied to trivial training data $f(x,t) = 1 + \varepsilon$ where the noise is $\varepsilon \sim \mathcal{N}(0, 0.1)$ for an equidistant grid of $x \in [0, 10]$ and $t \in [0, 10]$. Therefore, the uncertainty of the artificial data is $\Delta f = 0.1$. We want to test the convergence behaviour of the fit and see if the mean of the fit converges to $m(f) = 1$ despite the noise. The hyperparameters are kept constant at $[l_x, \sigma_x] = [10, 1]$ and $[l_t, \sigma_t] = [10, 1]$ to ensure that the correlation length spans the complete training data and thereby preventing the noise being fitted. The number of training data is varied to $N = N_x \times N_t = [5^2, 10^2, 20^2, 30^2, 40^2, 50^2, 60^2]$. The test data is kept fixed to $N^* = 60^2$. In Fig. 7 the training data, the 2D GP fit as well as its uncertainty and the residuals for $N = 10^2$ and $N = 60^2$ are shown. When looking at the 2D GP fit of Fig. 7a2 and Fig. 7b2 a constant value of the reconstruction $f^* \approx 1$ for all x^* and t^* can be seen. This corresponds to the mean $m(f) = 1$ of the trivial training data.

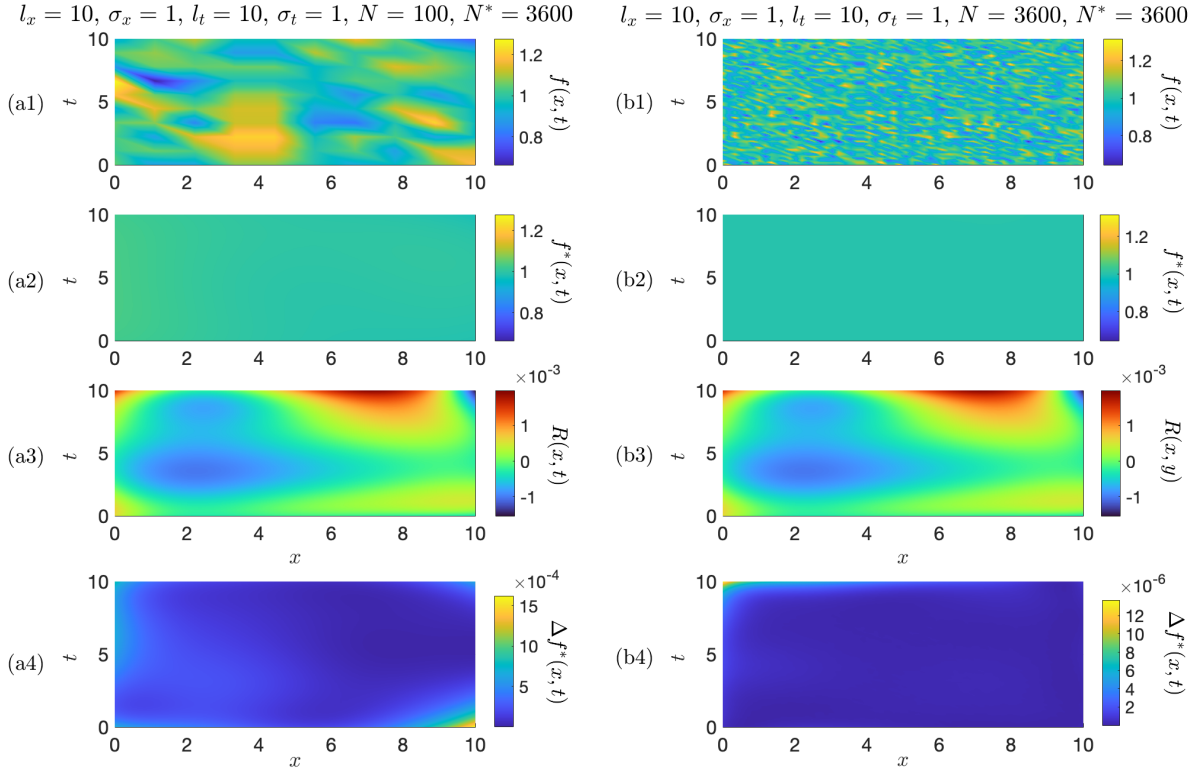


Figure 7: (a1) and (b1) show the trivial training data $f(x,t)$ with normally distributed noise $\varepsilon \sim \mathcal{N}(0, 0.1)$, uncertainty $\Delta f = 0.1$ and their 2D GP reconstruction $f^*(x,t)$ is displayed in (a2) and (b2). The residuals $R(x,t)$ are depicted in (a3) and (b3). (a4) and (b4) show the uncertainties $\Delta f^*(x,t)$. The two columns differ in the size of training data as (a) has $N = 10^2$ and (b) has $N = 60^2$.

The residuals

$$R(x,t) = f^*(x^*, t^*) - f(x,t) \tag{4.1}$$

compare the reconstruction by the 2D GP $f^*(x^*, t^*)$ to the training data $f(x,t)$. In this case, they give the deviation from the GP fit to the true value of $m(f) = 1$. The residuals $R(x,t)$

of artificial Gaussian data are depicted in Fig. 7a3 and Fig. 7b3. The uncertainties of the 2D GP fit, shown in Fig. 7a4 and Fig. 7b4, are smaller than the noise ε. The distribution of ε and $R(x,t)$ for some of the 2D GP's are shown in Fig. 8. It can be seen that the normally distributed noise is slightly reflected in the residuals, which is expected when looking at Eq. (4.1). The distribution of residuals gets narrower as the sample size increases, meaning the GP fit converges to the mean of the training data.

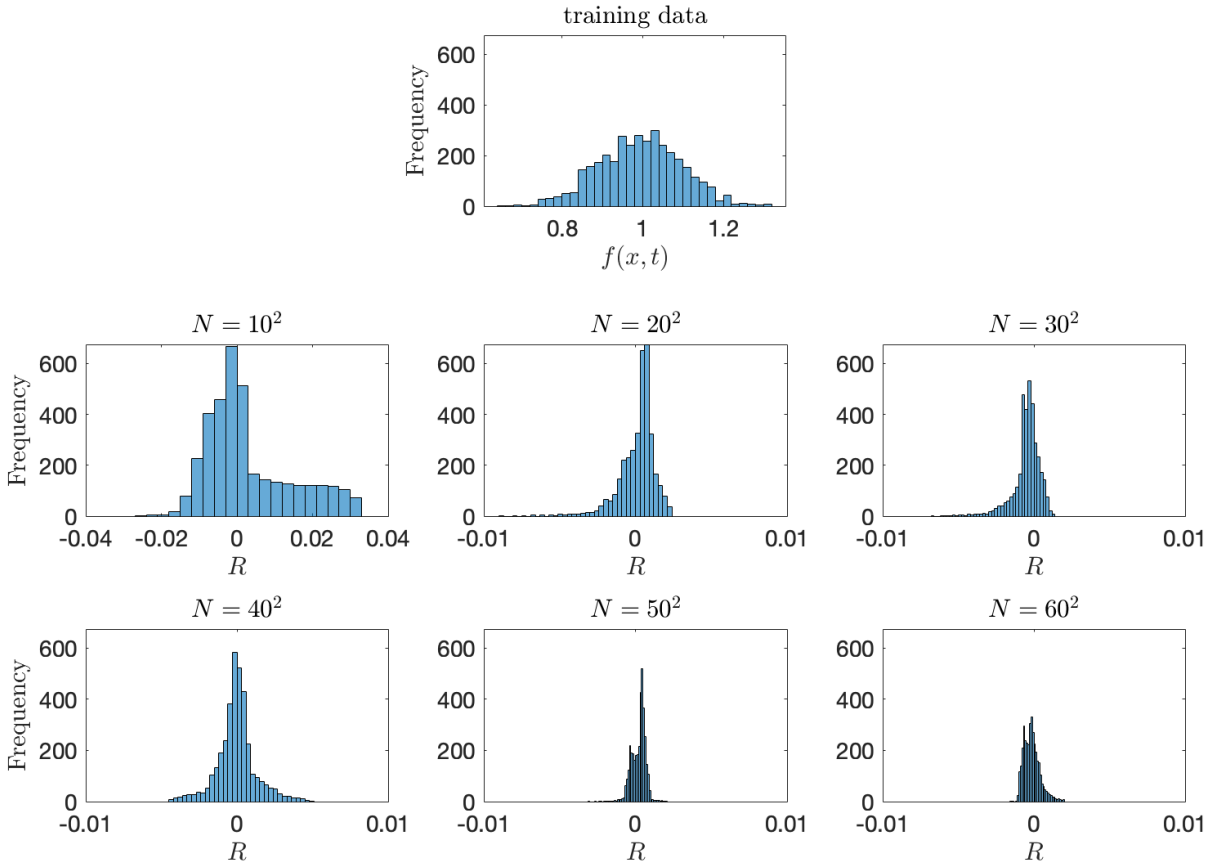


Figure 8: Distribution of noise ε for trivial training data $f(x,y) = 1 + \varepsilon$ with $\varepsilon \sim \mathcal{N}(0, 0.1)$, as well as the resulting frequencies of residuals R of the 2D GP for different sizes of training data N.

In addition, the mean residuals for each 2D GP fit were computed by

$$\overline{R} = \frac{1}{N} \sum \sqrt{R(x,t)^2} \tag{4.2}$$

and are plotted over the size N of the training data in Fig. 9. We can observe a decrease of $\overline{R}$ until $N = 30^2$ leading to convergence to $\overline{R} \sim 3 \cdot 10^{-4}$, which coincides with the convergence of the reconstruction to the mean. This shows a higher number of training data leads to a more accurate reconstruction due to the higher coverage of the underlying function. There are slight oscillations in the plot of $\overline{R}$. The possible reason for this could be the construction of the training data. The training data with a smaller sample size is randomly selected from the training data with a size of $N = 60^2$. Therefore, the selection at $N = 20^2$ could have chosen training data with less noise and, consequently, less deviation from the mean. The errors of $\overline{R}$ result from the

propagation of uncertainty

$$\Delta\overline{R} = \frac{1}{N}\sum\sqrt{\Delta R^2} = \frac{1}{N}\sum\sqrt{(\Delta f^* + \Delta f)^2} \tag{4.3}$$

with the fit and training data uncertainty Δf^* and Δf, respectively. The fit uncertainties are calculated using Eq. (2.2) and are summarized in Tab. 1. They range between $4 \cdot 10^{-4}$ and $5 \cdot 10^{-7}$ and decrease for more training data. However, the fit uncertainties exhibit fluctuations which seem to coincide with the fluctuations in $\overline{R}$. As the fits uncertainties are multiple orders of magnitude smaller than the given uncertainty of the training data $\Delta f = 0.1$, the mean residual uncertainty is $\Delta\overline{R} \approx \Delta f = 0.1$. The values of $\overline{R}$ in Fig. 9 do not show any error bars, because of their small values < 0.01 compared to $\Delta\overline{R} \approx 0.1$.

In conclusion, additional training data results in smaller residuals and reduced uncertainty. However, they eventually approach a constant value rather than zero, indicating that further training data becomes unnecessary. This demonstrates that the 2D GP can accurately reconstruct the underlying data despite the presence of noise.

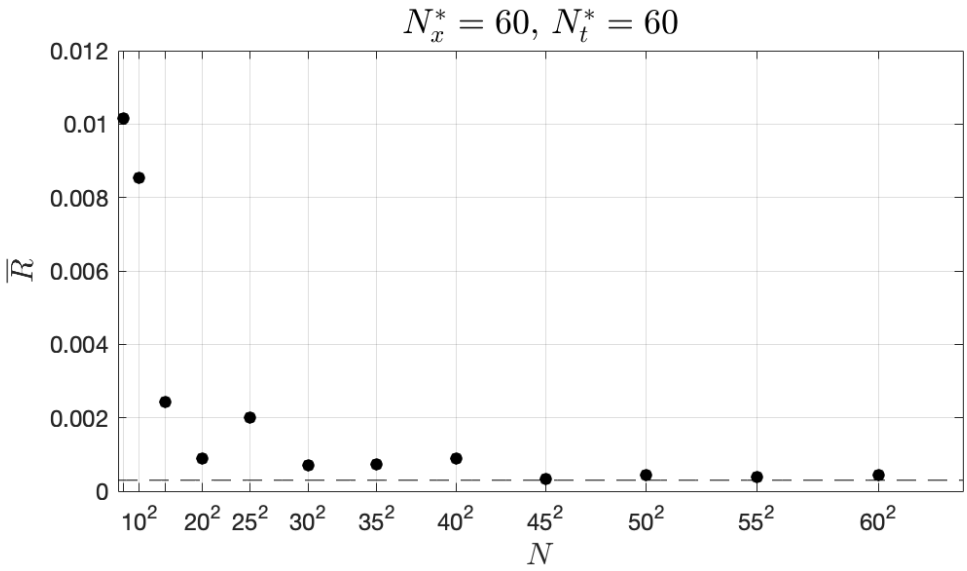


Figure 9: Mean residuals $\overline{R}$ of the 2D GP fits in dependence on the size of training data N. Application on noisy trivial data with size of test data $N^* = 3600$ for all sizes of training data N.

Table 1: Averaged uncertainties $\overline{\Delta f^*}$ of the 2D GP reconstruction with $\Delta f = 0.1$ for different sizes of training data N.

N	5^2	10^2	15^2	20^2	25^2	30^2
$\overline{\Delta f^*}$	$4 \cdot 10^{-4}$	$1.6 \cdot 10^{-4}$	$1.9 \cdot 10^{-5}$	$2.7 \cdot 10^{-6}$	$7 \cdot 10^{-6}$	$4 \cdot 10^{-6}$

N	35^2	40^2	45^2	50^2	55^2	60^2
$\overline{\Delta f^*}$	$1.3 \cdot 10^{-6}$	$1.2 \cdot 10^{-6}$	$5 \cdot 10^{-7}$	$4 \cdot 10^{-7}$	$2.8 \cdot 10^{-7}$	$5 \cdot 10^{-7}$

4.2 Application to artificial Gaussian data

4.2.1 Influence of noise

In actual measurements of plasma parameters, e.g. electron density, uncertainties of the measurements exist, which need to be taken into consideration in the GP as noise. In Eq. (2.1) the noise ε is included in the term $\varepsilon_n^2 \boldsymbol{I}$. The effect of noise levels needs to be analyzed. Thus, constant noise is added to the artificial data described in section 3, acting as uncertainties of measurements. The noise is set to $\varepsilon \in [0.01, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1]$. As already shown in Fig. 6, the hyperparameters have an influence on the accuracy of the 2D GP reconstruction, therefore a spatial hyperparameter optimization by maximization of the marginal likelihood was used. While the spatial hyperparameters $[l_x, \sigma_x]$ were optimized, the temporal hyperparameters were kept constant at $[l_t, \sigma_t] = [1, 1]$, as the temporal optimization is not yet fully elaborated. This leads to higher residuals at the decrease of the artificial data at $t = 2.5\,\mathrm{s}$ (see Fig. 10). The spatial hyperparameter optimization was done as described in section 3.2.1, where the covariance matrix is considered for one single time-slice. Due to their construction the artificial data do not exhibit strong spatial variations. However, this is not the case for the experimental data. In this work, it can be seen that this simplification worked well for the artificial data, as well for the experimental data shown in section 4.3.2. This simplification should be reconsidered for other applications with strong spatial changes.

In Fig. 10 the residuals of the GP reconstruction to the training data are shown for the noise values $\varepsilon = [0, 0.01, 0.1, 1]$, where it can be seen that both the absence of noise and too much noise leads to strong deviations in the 2D GP reconstruction to the training data.

The smallest residuals are found for a minimal noise $\varepsilon = 0.01$. It can be assumed that the residuals get smaller for even smaller noise levels. For $\varepsilon = [0.01, 0.1, 1]$ the greatest deviation can be seen at $t = 2\,\mathrm{s}$ coinciding with a strong change in the amplitude of the artificial data. The deviations are increased by the fixed correlation length $l_t = 1$, as it does not allow the 2D GP to "react" to the abrupt changes. When considering the effect of the hyperparameters, explained in section 2, the reason for this is the presence of spurious correlations between data that do not actually correlate. Using Eq. (4.2) the mean residuals are plotted over the given noise ε in Fig. 11. The errorbars were calculated with Eq. (4.3), therefore growing linearly due to $\Delta f = \varepsilon$. The residuals shown in Fig. 11 appear to converge towards a constant value of approximately 0.25. However, to test the convergence of the residuals, the residuals are also calculated for both $\varepsilon = 10$ and $\varepsilon = 20$. The results are $R(10) \approx (0.4 \pm 0.1)$ and $R(20) \approx (0.56 \pm 0.25)$, indicating that the residuals do not converge, when increasing the noise values. The reported errors $\Delta R(10)$ and $\Delta R(20)$ are the averaged uncertainties $\overline{\Delta f^*}$ of the GP reconstructions.

It becomes clear, that for bigger noise values the 2D GP is given more flexibility in finding the most likely fit and therefore the residuals are bigger as well. While, as displayed in Fig. 11, the mean residuals are minimized for noise values approaching zero, the omission of noise in the GP leads to strong deviations in the reconstruction. In this case the calculated mean residuals of $\varepsilon = 0$ is $\overline{R}(\varepsilon = 0) \approx (4.11 \pm 0.04) \cdot 10^{11}$, indicating that the GP reconstruction does not fit the data. The noise could therefore provide a necessary flexibility for the GP for finding a fitting reconstruction. The noise of experimental training data is always given by the uncertainty of the measurement. Additionally, the GP fits uncertainty is greater in areas where the training data uncertainty is larger.

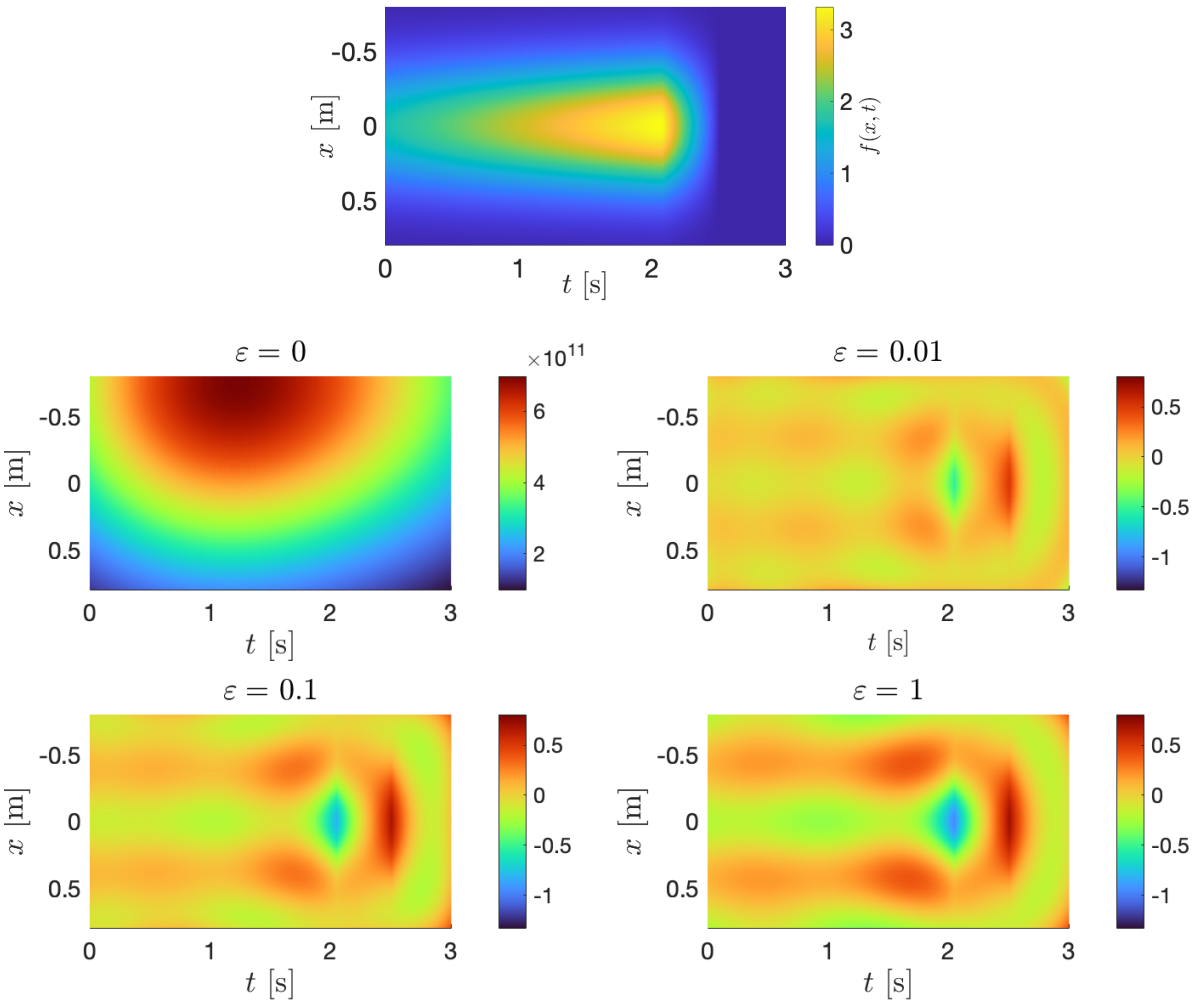


Figure 10: Residuals $R(x,t)$ of 2D GP reconstruction to artificial data (shown at the top) for different noise values ε.

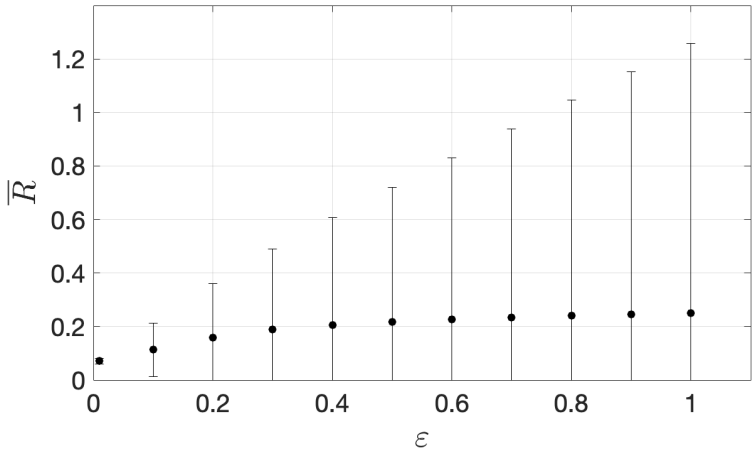


Figure 11: Mean residuals $\overline{R}$ of the training data with $N = 30 \times 50$ and test data $N^* = 50 \times 80$ for varying noise values ε. The error bars show the errors $\Delta\overline{R}$ growing linearly due to the linear growth of $\Delta f = \varepsilon$.

4.2.2 Influence of sample size

For the 2D GP the size of the training data $N = N_x + N_t$ is also of importance. As the training data is two-dimensional, both N_x and N_t can be varied. The correlation between $\overline{R}$ and N is shown in Fig. 12. Thereby, the size of the training data $N_{x/t}$ was varied for the spatial and temporal dimensions, respectively, while keeping the other constant at $N = 100$. In order to see only the effect of the size of training data, the hyperparameter are kept fixed at $[l_x, \sigma_x] = [1, 1]$ and $[l_t, \sigma_t] = [1, 1]$. A constant noise value of $\varepsilon = 0.1$ was applied because it showed small residuals in the previous section. Due to the chosen noise value the mean residuals are close to 0.1, which corresponds to the values in Fig. 11. For both $\overline{R}(N_x)$ and $\overline{R}(N_t)$ it can be seen that an increasing size of training data leads to decreasing $\overline{R}$. For a larger set of training data they both approach the same value of mean residual $\overline{R} \approx (8.412 \pm 0.003) \cdot 10^{-2}$ with the residuals of N_t variation being greater than for N_x variation. The oscillations in $\overline{R}(N_t)$ could be due to the choice of artificial training data and N. Due to a normal distribution of the training data, for lower N_t the strong variation of amplitude at $t = 2\,\mathrm{s}$ in the artificial data might not be sufficiently covered.

In addition, the runtime increases with the size of the training data. Increasing the amount of training data from $N_t = 10$ to $N_t = 100$ extends the computational time by 1.89 h due to the increase of training data N by 9000 points.

In conclusion, the accuracy of the reconstruction of training data with a 2D GP is higher for more training data, as the reconstruction of small changes is possible. If there are large variations in the training data, it is beneficial to have more training data at these locations or times. However, increasing the amount of training data results in longer GP runtimes. It is important to note that over a certain amount of training data adding more will not necessarily improve the GP's performance, as the residuals approach a constant value. With respect to experimental data of plasma parameters, the amount of training data can be increased by combining the data at $r_{\mathrm{eff}} < 0$ and $r_{\mathrm{eff}} \geq 0$ due to the radial symmetry.

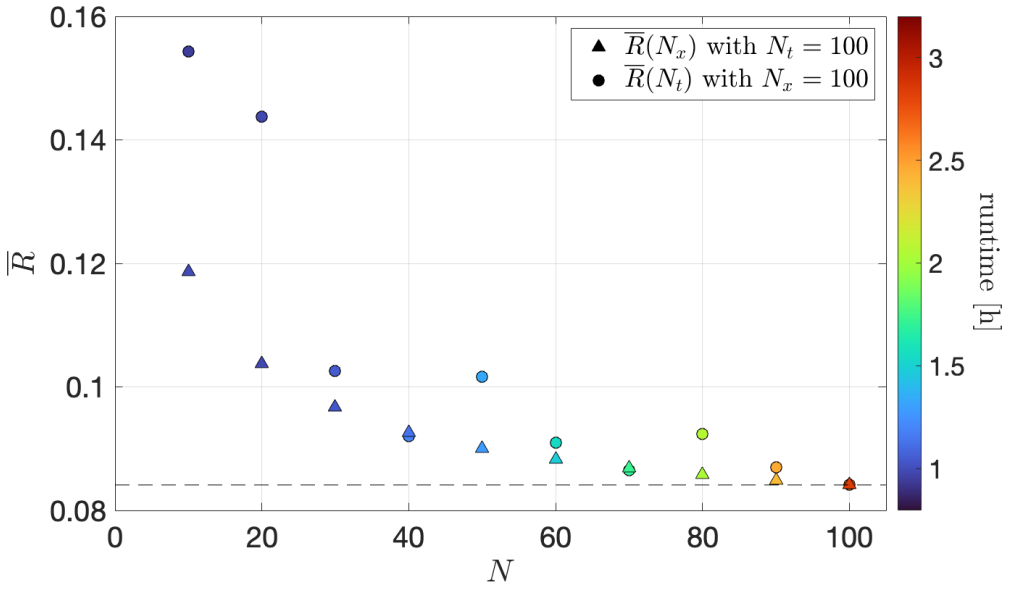


Figure 12: Mean residuals $\overline{R}$ for variation of training data size N_x (triangles) and N_t (circles) at noise $\varepsilon = 0.1$. The runtime of the individiual GP is indicated by the marker color.

4.2.3 Implementation of a time dependent temporal hyperparameter

Section 3 presented a new method for determining the temporal hyperparameter $l_t(t)$, which is applied to artificial data (see Fig. 13a) in this section. To better distuingish stationary and non-stationary phases, the artificial data was slightly adjusted by adding a constant amplitude in $t = 0\,\mathrm{s}$ to $t = 0.5\,\mathrm{s}$. The sample size of the training data is $N = 20 \times 60$ leading to a temporal resolution and minimum for $l_t(t)$ of $\Delta t = 0.05\,\mathrm{s}$. The 2D GP reconstruction is displayed in Fig. 13b with the time dependent hyperparameter shown Fig. 13c.

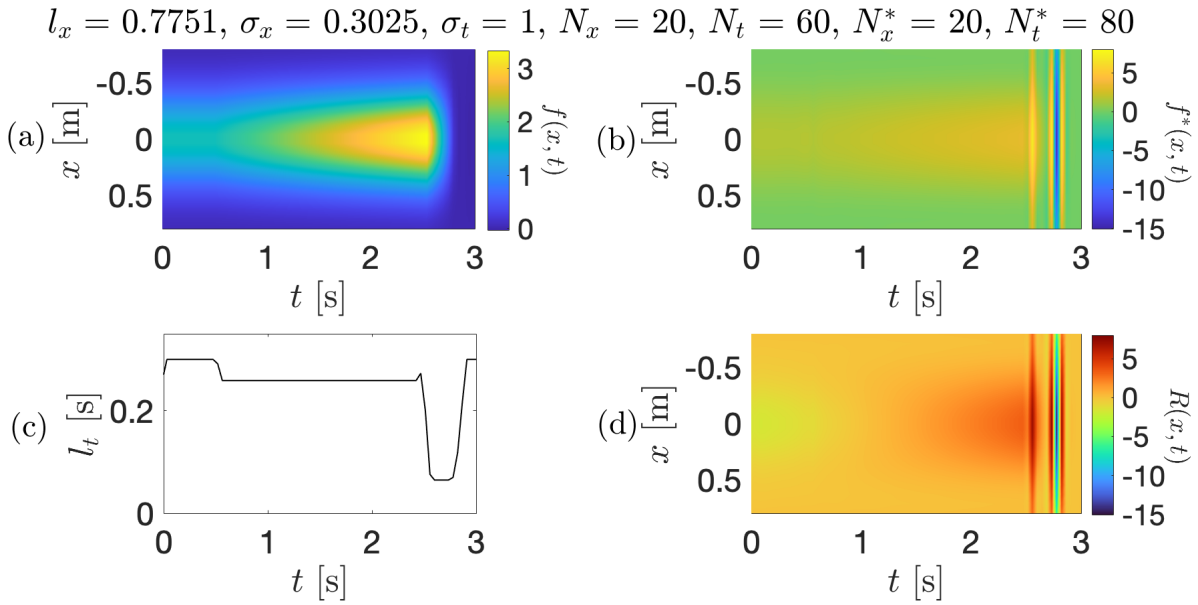


Figure 13: (a) shows artificial training data $f(x,t)$ with a sample size of $N = 20 \times 60$. (b) displays the 2D GP reconstruction $f^*(x,t)$ with size of test data $N^* = 20 \times 80$. (c) depicts the time dependent hyperparameter $l_t(t)$ calculated with the line integrated density of the training data. (d) shows the residuals $R(x,t)$ of the 2D GP reconstruction to the training data.

The calculation of $l_t(t)$ worked as desired, as the hyperparameter is large for stationary training data and small for large temporal changes. However, inspecting the 2D GP reconstruction in Fig. 13b, it can be seen that the reconstruction exhibits strong oscillations, ranging from approximately -15 to 8. These oscillations lead to large deviations from $t \approx 2.51\,\mathrm{s}$ to $t \approx 2.87\,\mathrm{s}$, where both the training data and the hyperparameter $l_t(t)$ suddenly decrease. The residuals in Fig. 13d show that for all other times the 2D GP reconstruction fits well to the training data.

The observed oscillations in the 2D GP reconstruction suggest that the hyperparameter minimum may have been chosen too small, resulting in overfitting (as seen in Fig. 6). Because of this, the minimum and maximum of $l_t(t)$ were altered, with the boundaries listed in Tab. 2. This test was conducted to examine the influence of different minima and maxima of $l_t(t)$ and has no physical reason. It was found that optimization of the spatial hyperparameters by maximization of the marginal likelihood resulted in larger and sometimes too large values when the time dependent correlation length $l_t(t)$ is used. Therefore, spatial hyperparameters are first optimized for a fixed value of l_t and then used as fixed hyperparamaters with optimization of temporal hyperparameters. The resulting 2D GP reconstructions as well as the residuals and corresponding hyperparameters are depicted in Fig. 14b – 14d. In the 2D GP reconstruction it can be seen that the increase of the minimum of $l_t(t)$ leads to a decrease of oscillations.

Table 2: Averaged residuals $\overline{R}$ (rounded to the third decimal place) of 2D GP reconstructions of artificial Gaussian training data for different maxima and minima of the time dependent hyperparameter $l_t(t)$.

	I	**II**	**II**	**IV**	**V**
$\max(l_t(t))$ [s]	0.06	0.15	0.3	0.3	0.3
$\min(l_t(t))$ [s]	0.06	0.05	0.05	0.15	0.25
$\overline{R}$	0.565	1.362	2.398	0.572	0.566

As previous shown in Fig. 6 can a correlation length that is too small result in overfitting. To test whether the sudden change of the hyperparameter $l_t(t)$ or the small value of the minimum causes the oscillations, the maximum of $l_t(t)$ was lowered while keeping the minimum fixed. It can be seen in Fig. 14b, that this leads to a slight decrease of oscillations as well. Additionally, as shown in the first column of Fig. 14, the 2D GP reconstruction does not oscillate when a constant value of $l_t = 0.06\,\mathrm{s} = \text{const.}$ is chosen for the hyperparameter, contradicting the assumption of a too small hyperparameter. The averaged residuals $\overline{R}$ for each 2D GP reconstruction are listed in Tab. 2. The smallest averaged residual depicts the reconstruction with fixed hyperparameter, followed by the reconstruction with minimal changes of $l_t(t)$. The oscillations of the 2D GP reconstruction could be decreased further when the hyperparameter minimum approaches its maximum, resulting in a constant hyperparameter $l_t(t) = l_t$.

The implementation of a time dependent hyperparameter $l_t(t)$ lead to strong oscillations in the 2D GP reconstruction, where the hyperparameter has strong changes. In addition, the spatial hyperparameter optimized by maximizing the marginal likelihood deviate when a time dependent hyperparameter $l_t(t)$ is used. A comparison with a small fix hyperparameter $l_t = 0.06\,\mathrm{s}$ confirmed that the oscillations are not the result of overfitting due to a too small hyperparameter. The oscillations are assumed to be caused by strong and rapid changes in $l_t(t)$. Due to the oscillations, optimizing the temporal hyperparameter does not perform better than using a fixed hyperparameter. Optimizing the temporal hyperparameter by maximizing the marginal likelihood would not allow for a varying length scale and therefore does not provide an alternative.

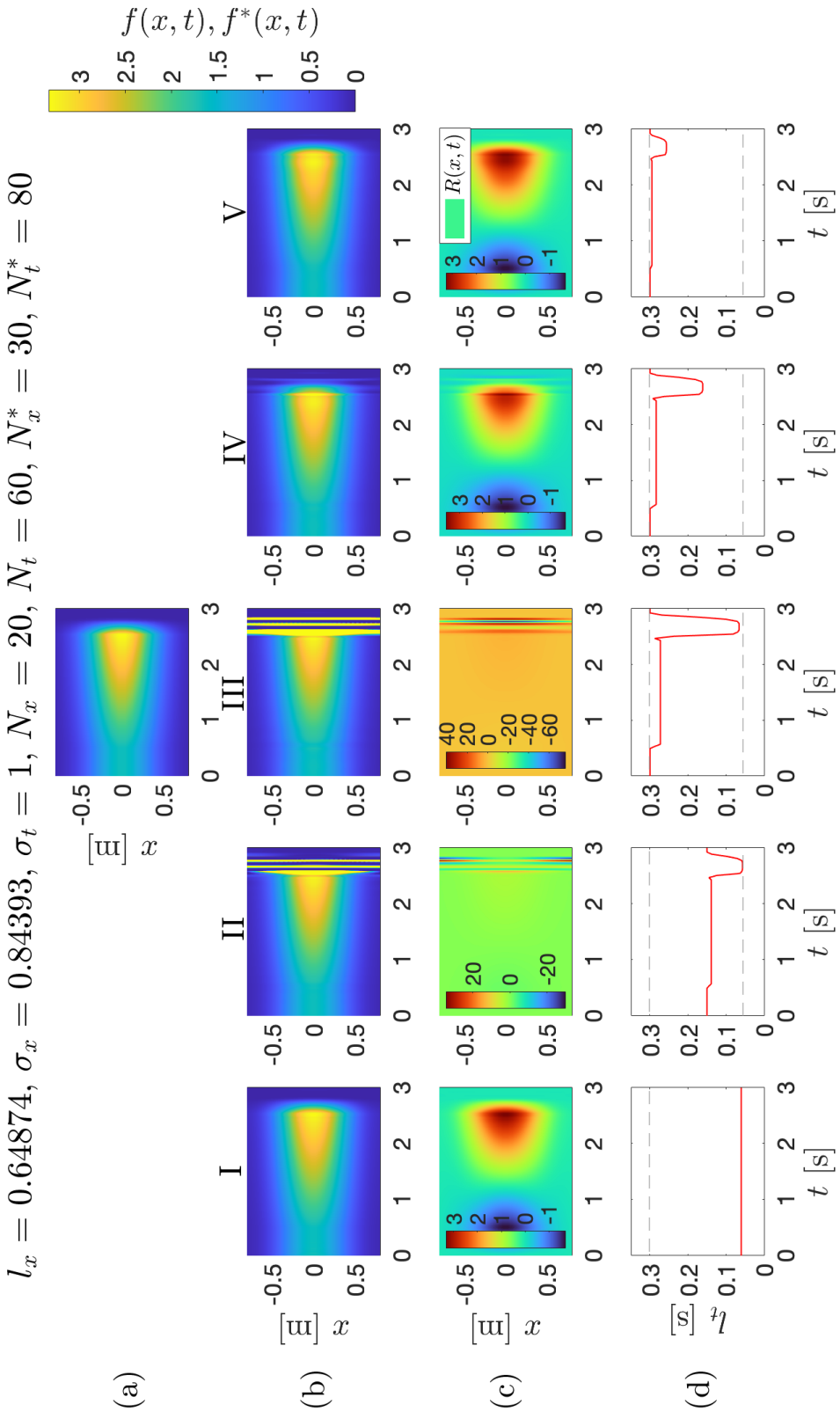


Figure 14: (a) shows the artificial training data $f(x,t)$. (b) depicts different 2D GP fits of the training data. (c) displays the residuals $R(x,t)$ of the 2D GP fits to the training data. (d) depicts the used time dependent hyperparameter $l_t = l(t)$ for different minima and maxima.

4.3 Application of 2D GP to experimental data of LHD

A goal of the development of the 2D GP for experimental plasma physics data is a qualitative reconstruction of fast temporal changes in the data. For a first application of the 2D GP to experimental data, we focus on the spatio-temporal evolution of the electron density n_e. For that, the LHD discharge #185880 is chosen because of its fast temporal changes in density as the injection of pellets into the plasma leads to density peaking. At times $[3.85, 3.91, 4.06, 4.36, 4.57]$ s, a single pellet made of frozen hydrogen is injected. Parameters of this discharge, such as the heating power of Electron Cyclotron Resonance Heating (ECRH) [31] and Neutral Beam Injection (NBI) [32] (P_{ECRH}, P_{NBI}) and radiation power P_{rad}, plasma energy W_{dia}, plasma current I_{p}, electron temperature T_e and line integrated electron density $\int n_e \mathrm{d}L$ are shown in Fig. 15.

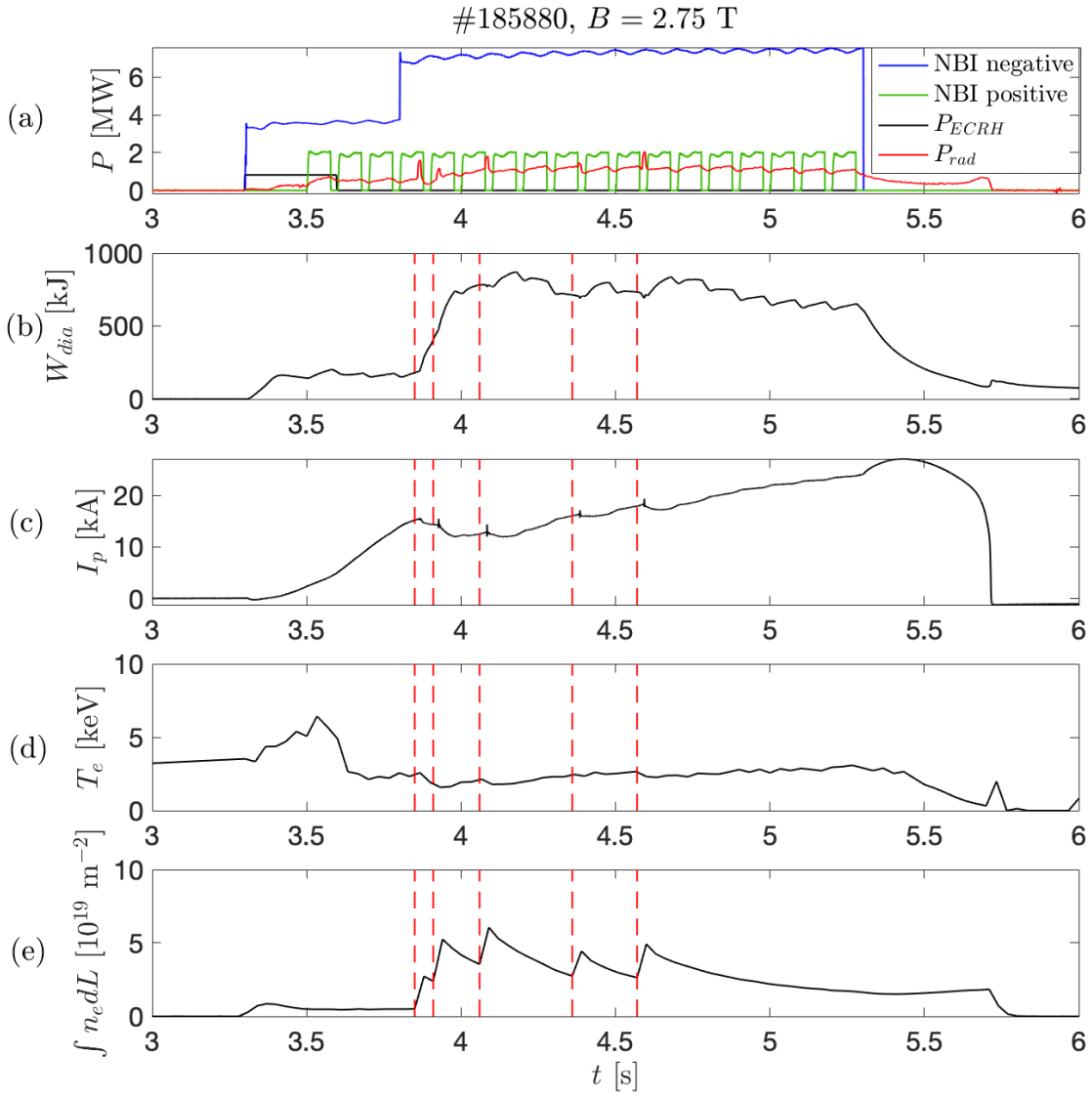


Figure 15: Temporal evolution of discharge parameters of LHD discharge #185880: (a) heating power of positive and negative NBI P_{NBI} and ECRH P_{ECRH} as well as the radiation power P_{rad}, (b) plasma energy W_{dia}, (c) plasma current I_p, (d) electron temperature T_e and (e) line integrated electron density $\int n_e \mathrm{d}L$. The five red dashed lines indicate the times of pellet injection.

The main heating source is NBI. Initially the heating power is $P_{\mathrm{NBI}} \sim 3.5\,\mathrm{MW}$, which is increased to $P_{\mathrm{NBI}} \sim 7\,\mathrm{MW}$ at $t = 3.8\,\mathrm{s}$. Before pellet injection the plasma is stationary with the values of plasma energy and line integrated electron density being $W_{\mathrm{dia}} \sim 150\,\mathrm{kJ}$ and $\int n_e \mathrm{d}L \sim 0.5 \cdot 10^{19}\,\mathrm{m}^{-2}$. The plasma current is ramped up during the whole discharge. The electron temperature peaks at the beginning of the discharge at $\sim 6.5\,\mathrm{keV}$ due to the ECRH used for plasma start-up. The step-up in NBI leads to a significant increase in W_{dia} to 600 kJ – 900 kJ. The periodic fluctuations in W_{dia} are caused by the modulation of NBI. The plasma current increases steadily up to 27 kA. Additionaly, I_{p} exhibits small peaks after each pellet injection. T_e remains quasi-stationary for most of the time, but declines slightly after each pellet injection. The line integrated electron density increases abruptly at the start of the second phase of heating, which is further enhanced by the injection of the first pellet. Each pellet injection results in a peak of the line integrated electron density. Between the pellet injections, the line integrated electron density declines slowly. The end of heating results in a decline of each plasma parameter.

4.3.1 Downsampling of training data

As mentioned in the previous section 4.2.2 the size of the training data plays an important role in the 2D GP, because of the computational complexity scaling with $\mathcal{O}(N^3)$ [20]. This also implies that the runtime increases with the training data. For the reconstruction of the experimental data, the electron density $n_e(r,t)$ measured by Thomson scattering [27] is chosen as the training data. The chosen training data is the spatial-temporal evolution of the electron density (see Fig. 16) in the time interval $t \in [3.3, 5.2]$ s, in which the plasma is heated by NBI. For the spatial interval $r_{\mathrm{eff}} \in [-0.65, 0.65]$ m were chosen, corresponding to the plasma boundary.

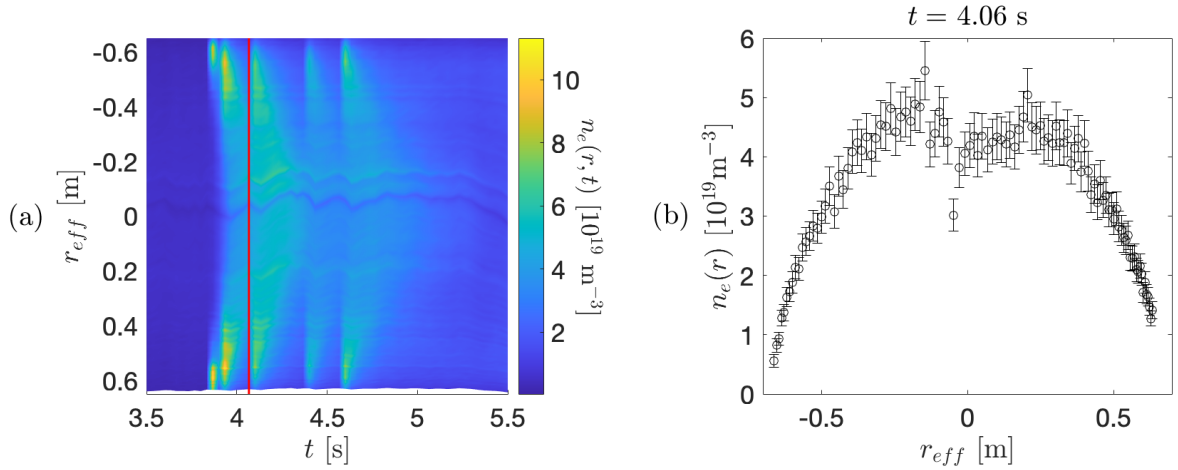


Figure 16: (a) Spatial-temporal evolution of electron density $n_e(r,t)$ of the LHD discharge #185880. (b) Electron density profile $n_e(r)$ at $t = 4.06\,\mathrm{s}$ indicated by the red line in (a).

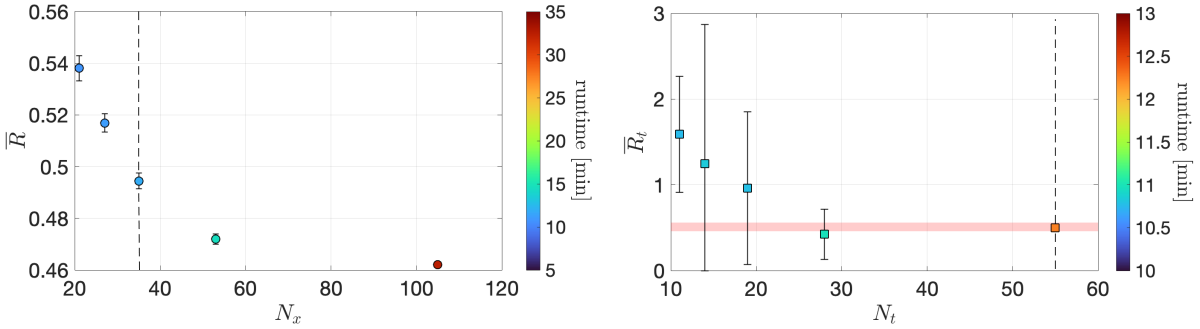
The electron density is measured with a sampling frequency of $\sim 30\,\mathrm{Hz}$, resulting in a total size of $N = 105 \times 55 = 5775$. Hence, it is of importance to see if the size of the training data can be downsampled to save on computational costs but without loosing important information. At first, only the training data in spatial dimension was downsampled by reducing the data by a factor of 2,3,4 and 5, respectively, while keeping the temporal training data set size fixed at the initial value of $N_t = 55$. For each downsampled training data the averaged mean residual $\langle \overline{R} \rangle$ is calculated

over five iterations and is plotted over the size of training data N_x in Fig. 17a. The mean residuals $\langle\overline{R}(N_x)\rangle$ and $\langle\overline{R}(N_t)\rangle$, that are averaged over five iterations, will be referred as $\overline{R}_x$ and $\overline{R}_t$. Without downsampling the mean residual has a value of $\overline{R}_x(105) \approx 0.4621 \pm 0.0011$. It can be seen that $\overline{R}_x$ increases when the spatial training data is downsampled. When downsampling N_x, the residuals steadily increase up to a value of $\overline{R}_x(21) \approx 0.538 \pm 0.005$, i.e. the mean residual increased by

$$\delta = \frac{\overline{R}_x(21) - \overline{R}_x(105)}{\overline{R}_x(105)} = \frac{0.538 - 0.462}{0.462} \approx (16.5 \pm 1.4)\,\%$$

when downsampled to the fifth of its original size. The error was calculated by propagation of uncertainty

$$\Delta\delta \approx \left|\frac{\partial\delta}{\partial\overline{R}_x(105)}\right| \Delta\overline{R}_x(105) + \left|\frac{\partial\delta}{\partial\overline{R}_x(21)}\right| \Delta\overline{R}_x(21) = \frac{\overline{R}_x(21)}{\overline{R}_x(105)^2}\Delta\overline{R}_x(105) + \frac{\Delta\overline{R}_x(21)}{\overline{R}_x(105)} \approx 1.4\%. \tag{4.4}$$



(a) (b)

Figure 17: Mean residuals $\langle\overline{R}\rangle$ of 2D GP reconstruction averaged over five iterations for downsampled training data from LHD discharge #185880. (a) shows $\overline{R}_x$ for downsampled spatial training data with $N_t = 55$ and (b) displays $\overline{R}_t$ for downsampled temporal training data with $N_x = 35$. The chosen size of test data is equivalent to the inital size of training data $N^* = 105 \times 55$. The red area depicts the range of residuals $\overline{R}_x$ and the dashed lines indicate the point with the same sizes of training data.

After downsampling the size of the spatial data, the size of temporal data is downsampled. The choice of downsampling both N_x and N_t is not only a practical one but gives the opportunity to assess the effect of combined downsampling on the quality of the GP reconstruction. For the downsampling of N_t the downsampled spatial sample size is chosen to be $N_x = 35$ because of the lower computational time but still small residuum. The points at $N_x = 35$ in Fig. 17a and $N_t = 55$ in Fig. 17b (indicated by the dashed lines) represent the same point as they have the same sizes of training data and therefore the same mean residual of $\overline{R}_t \approx 0.497 \pm 0.004$. Similar to Fig. 17a the mean residuals increase for downsampled N_t. Here, the training data is downsampled twice, hence the mean residuals are much bigger. The increase of the mean residual results in $\delta \approx (220 \pm 150)\,\%$ with an error propagation calculation similar to Eq. (4.4). For the LHD discharge #185880 the spatial resolution amounts to $\Delta r_{\text{eff}} \sim 15\,\text{mm}$ and the temporal resolution to $\Delta t \sim 30\,\text{ms}$, i.e. the spatial resolution is higher and therefore downsampling spatial training data does not result in such high increase of mean residuals as the downsampling of the temporal training data does. Additional to the increase of the mean residuals, there is a large increase in the errors of $\overline{R}$, due to bigger uncertainties of the 2D GP's. The results are consistent

with the results in section 4.2.2. However, there is also a reduction of the runtime needed to calculate the 2D GP. For example, downsampling the spatial training data by a third also results in a reduction of runtime to almost a third ($11.8664\,\mathrm{s}/32.4512\,\mathrm{s} \approx 0.37$) of the initial runtime, while the mean residual increases by $\delta \approx (7.6 \pm 1.2)\,\%$. If downsampling the training data is necessary needs to be evaluated individually as it reduces the runtime at the cost of information loss. For rapidly changing data, it is not advisable to downsample the training data as it leads to a loss of information.

4.3.2 Comparison 2D GP vs. multiple 1D GPs

In this section the difference between 2D GP and 1D GP for two-dimensional training data (electron density n_e shown in Fig. 16) is analyzed. The training data sample size is reduced by downsampling once in the spatial dimension, considering only every second data point. This results in a sample size of $N = 53 \times 52$. The measurement uncertainty Δn_e is used as noise ε.

For the reconstruction with 1D GPs, the spatial data of each time-slice is fitted individually and put together in order to achieve a 2D reconstruction. Due to the smaller dimension of the 1D GP training data ($N_x = 53$), the spatial hyperparameter are optimized by maximizing the marginal likelihood for each individual time-slice. The size of spatial test data is $N_x^* = 100$, resulting in a combined fit with size $N^* = 100 \times 52$.

In case of the 2D GP, the complete training data is fitted at once, with and without temporal hyperparameter optimization, respectively. A size of $N^* = 100 \times 100$ is used for the test data. For the first 2D GP reconstruction only the spatial hyperparameters are optimized by maximizing the marginal likelihood, while the temporal hyperparameter are kept fixed at $[l_t, \sigma_t] = [0.03, 1]$. This chosen value for the temporal length scale is equal to the temporal resolution $l_t = \Delta t \approx 0.03\,\mathrm{s}$, which provided the best fit according to section 4.2.3. The hyperparameter σ_t is kept fixed at $\sigma_t = 1$ as the variance is already optimized by σ_x and a factor of one has no impact on the kernel. For comparison, the second 2D GP reconstruction is done with optimizaton of the temporal hyperparameters. However, a temporal correlation time of $\tau = 0.02\,\mathrm{s}$ is used, in order to reduce the range of $l_t(t)$ and therefore reduce oscillations in the reconstruction. The calculated time dependent hyperparameter $l_t(t)$ is depicted in Fig. 18. As the spatial optimization does not provide reliable hyperparameters after optimizing the temporal hyperparameters, the same spatial hyperparameters $[l_x, \sigma_x] = [0.0769, 9.7732]$ used in the first 2D GP reconstruction are also used in the second 2D reconstruction.

Fig. 19a – 19d depict the training data and the reconstructions of the 1D and both 2D GPs, while Fig. 19e – 19h show the data and reconstructions uncertainty. The training data shows the presence of horizontal lines of similar electron density permeating through time, indicating the temporal correlation between the data. It can be seen, that these lines exhibit a radial shift over time. The shifting of these lines is caused by the *Shafranov shift* [33]. As mentioned in the introduction, a helical magnetic field is needed in order to prevent an outwards drift of the plasma particles due to a seperation of charges. The plasma particles follow the helical magnetic field lines and counteract the separation of charges. The helical currents are called *Pfisch-Schlüter current* and create an additional vertical magnetic field. The overlap of the vertical magnetic field with the radial component of the magnetic field causes the magnetic axis to shift outwards (*Shafranov shift*) [7]. These lines are faintly reflected in the two 2D GPs but not in the 1D GP, showing that temporal correlations are omitted in the reconstruction with multiple 1D GPs. The uncertainties of the 1D GP reconstruction shown in Fig. 19f do not display any temporal correlations as well. In all reconstructions it can be seen that the GP

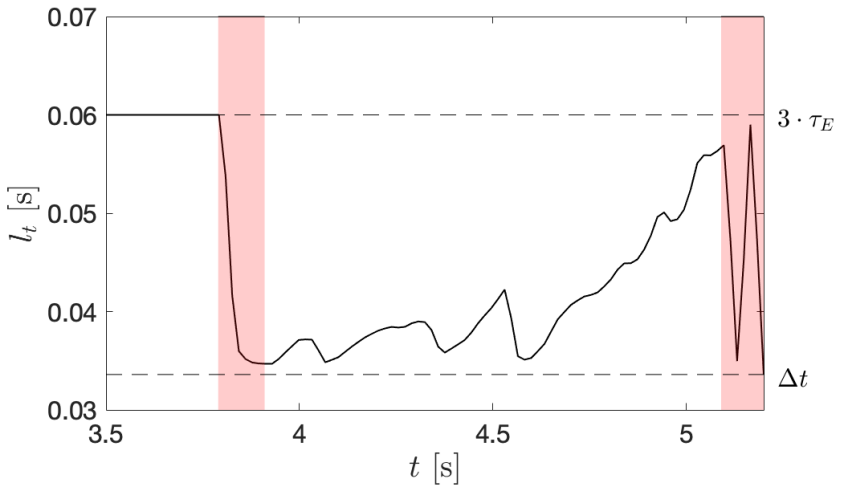


Figure 18: Behaviour of the temporal hyperparameter $l_t(t)$ used in the 2D GP reconstruction of LHD experimental data with temporal resolution $\Delta t \approx 0.0336\,\mathrm{s}$ and temporal correlation time of $\tau_E = 0.02\,\mathrm{s}$. The red areas display the times at which oscillations occur in the 2D GP reconstruction, coinciding with strong variations in $l_t(t)$.

smoothes the training data, especially along r_{eff}. The 2D GP reconstruction in Fig. 19d uses the temporal optimization with a time dependent hyperparameter $l_t(t)$. Similar to the results section 4.2.3, this 2D GP reconstruction also exhibits oscillations. At times where oscillations occur the uncertainty is large as shown in Fig. 19h. The oscillations range from approximately -5 to 18. In Fig. 19d the color range is limited to better display the rest of the reconstruction and enable a comparison between reconstructions. Except for the oscillations, the 2D GP reconstructions with and without optimization of temporal hyperparameter are similar.

In general, the 2D GP reconstruction shows temporal correlations, which are not present in the reconstruction with multiple independent 1D GP fits for each time-slice. This behaviour is also represented in the reconstructions uncertainties, which are of great interest when fitting plasma parameters. Furthermore, the 2D GP reconstruction allows a larger size of test data along the temporal dimension. Thus, the 2D GP reconstruction is an improvement over the independent 1D GP reconstructions.

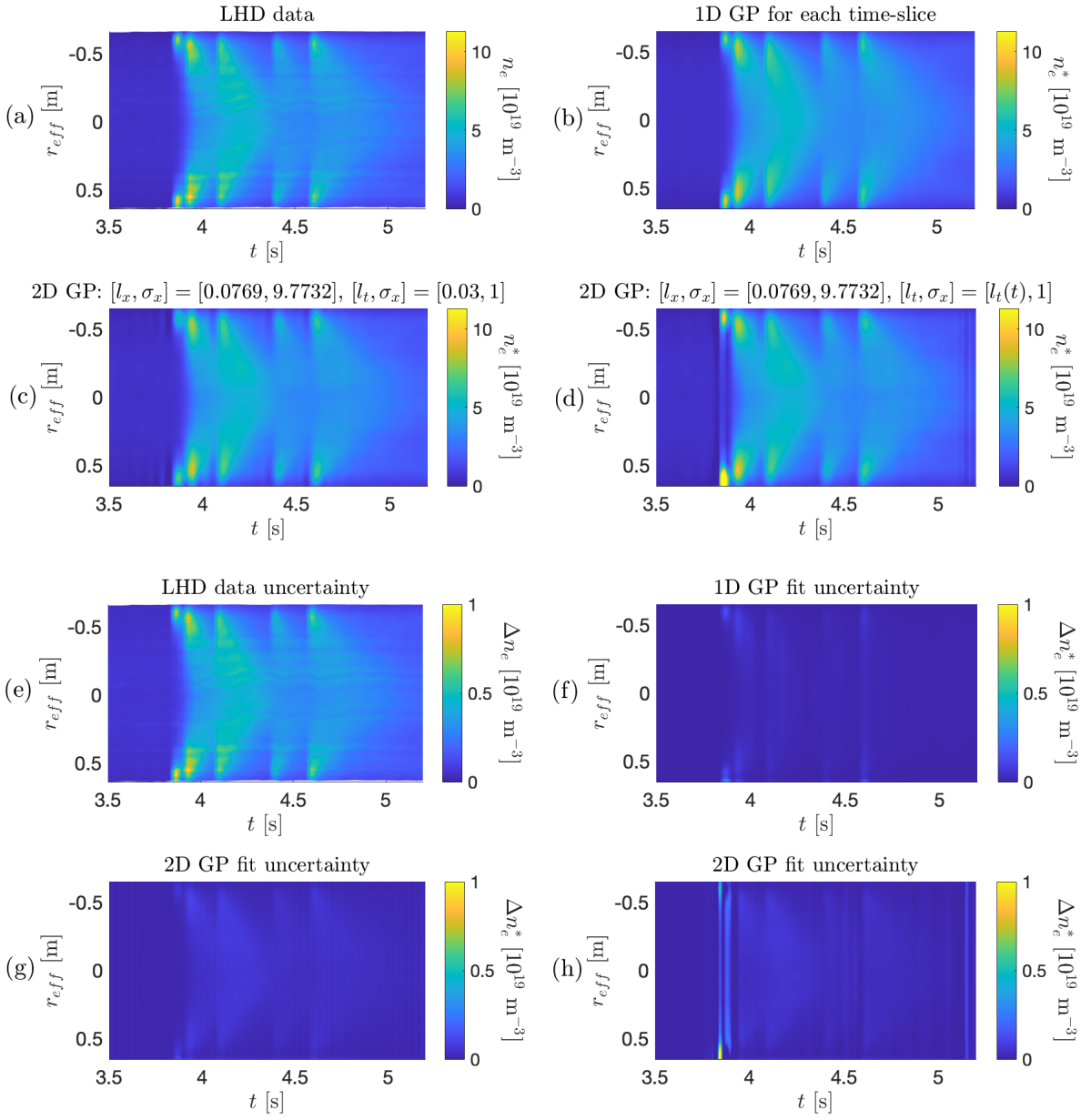


Figure 19: (a) shows the experimentally measured electron density n_e. (b), (c) and (d) depict the 1D GP reconstruction, 2D GP reconstruction with fixed temporal hyperparameter and 2D GP reconstruction with temporal hyperparameter optimization in this order. (e), (f), (g) and (h) display the absolut uncertainties, respectively.

5 Outlook

Going beyond the reconstruction of training data, the GP enables the calculation of the derivatives. The partial derivatives w.r.t. to space and time are needed for the calculation of transport dynamics as shown in the diffusion Eq. (1.2). The true spatial derivative of the 1D articifial data in Eq. (3.2) is

$$\frac{\partial f(x)}{\partial x} = -\frac{A(0)}{\sigma^3} \cdot x \cdot e^{-\frac{1}{2}(\frac{x}{\sigma})^2}. \tag{5.1}$$

This section outlines the calculation of the partial derivative using a 1D GP reconstruction. It provides insight into how spatial and temporal derivatives can be calculated using 2D GP reconstruction and how they could be used to determine diffusion coefficients. Fig. 20b compares the derivative derived from the GP fit in red with the true derivative from Eq. (5.1) in black. The GP fit derivative corresponds with the true derivative. There are slight deviations at $x = -0.8$ and $x = 0.8$, which result from the absence of training data.

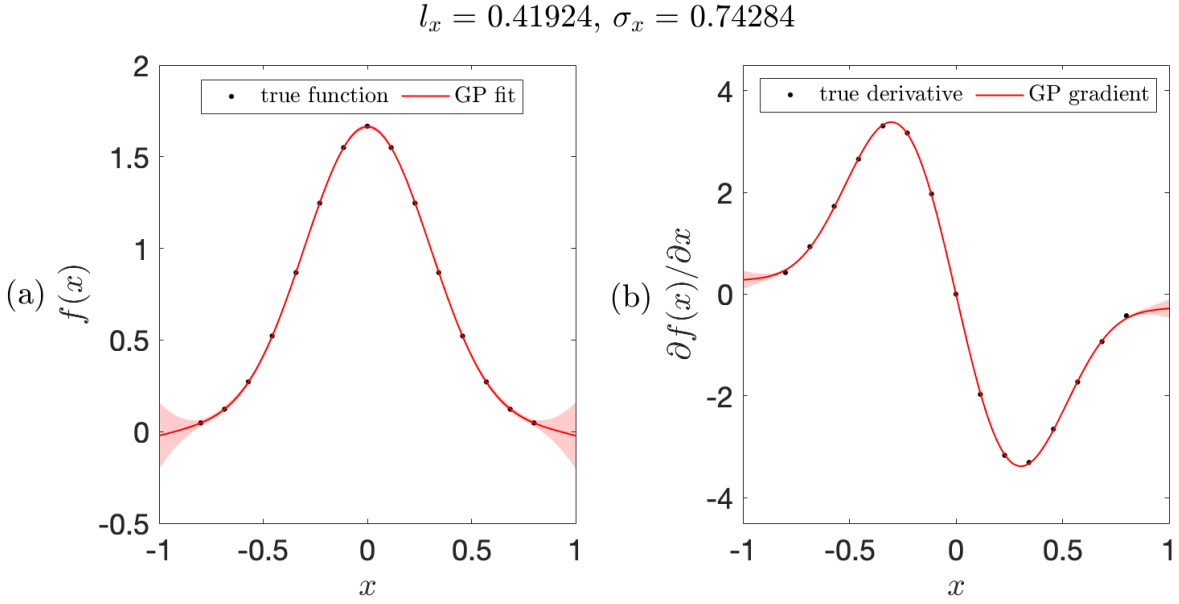


Figure 20: (a) shows the GP fit of the artificial Gaussian data $f(x)$ with hyperparameter optimization. (b) displays the true derivative of the artificial data and the derivative derived by the GP. The red shaded area shows the 95% confidence interval.

The application of 1D GP on the electron density profile n_e of LHD discharge #185880 at the time $t = 4.06\,\mathrm{s}$, where the third pellet is injected, is depicted in Fig. 21a. For the training data, the data at $r_{\mathrm{eff}} < 0\,\mathrm{m}$ (low-field) was mirrored, to get a larger sample size. It can be seen that the experimental data is very noisy and has larger measurement errors than the previous constructed artificial data. Hence, the GP's uncertainty is larger. Training data for $r_{\mathrm{eff}} > 0.7\,\mathrm{m}$ are considered as outliers and not implemented in the GP due to the unreliability of the measurements in this region. Additionally, the boundary condition of a disappearing gradient $\partial_r r'_{\mathrm{eff}} = 0$ at the plasma center $r'_{\mathrm{eff}} = 0$ was added. Fig. 21b displays the derivative of the GP fit in red. The GP fit and reconstructed gradient have the same uncertainty and therefore the same 95% confidence interval. The GP partial derivative shows that the condition of the disappearing gradient is fulfilled at $r'_{\mathrm{eff}} = 0$. For comparison, the plot also shows *Matlabs* finite difference method *gradient()* applied to the training data. Compared to the GP derivative reconstruction, the partial derivative calculation with a finite difference method does not have a reasonable result due to data noise.

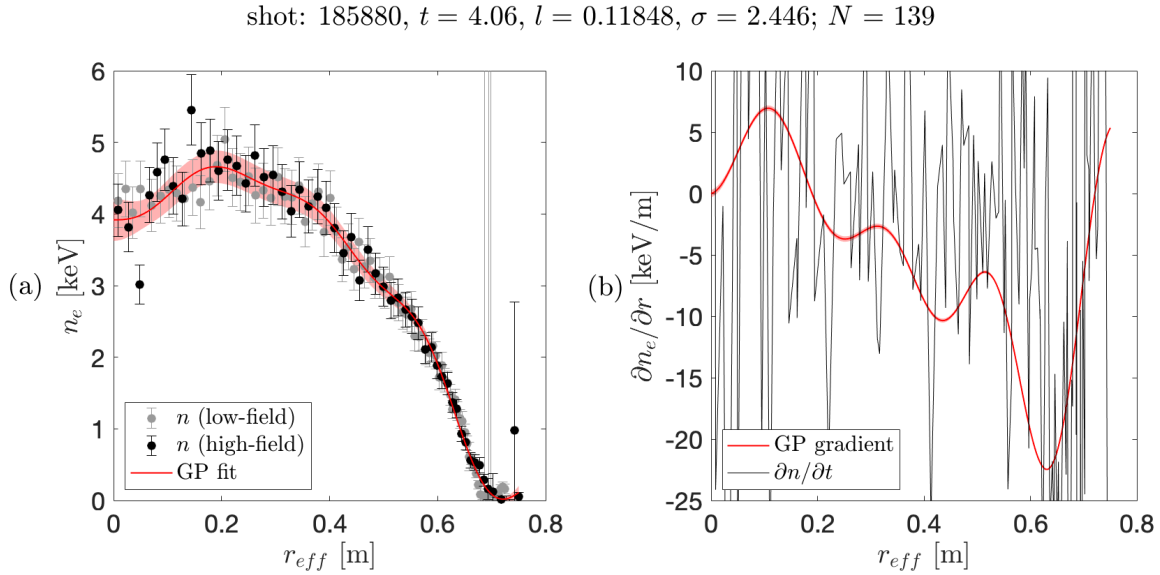


Figure 21: (a) shows the application of 1D GP with hyperparameter optimization to the electron density n_e. (b) compares the derivative derived by GP and the derivative by finite difference methods. The red shaded area shows the 95% confidence interval.

The goal is an implementation of the reconstruction of the partial derivatives in the 2D GP, as well as an implementation of boundary conditions like the gradient at the plasma center approaching zero. With the reconstructed partial derivatives a determination of diffusivity D is possible according to Eq. (1.2). In [34] a similar approach with a squared exponential (SE) kernel was used to reconstruct the partial derivatives with a GP and thereby calculate the diffusivity of the waterlevel in soil, resulting in good determinations of the diffusivity.

As applying a linear operator to a GP yields another GP, the kernel can be constructed in a way that it satisfies a partial differential equation (PDE). One possibility for a so called convolution kernel (more precisely heat kernel) proposed by [26] is

$$k_n(x,t,x',t';\sigma,D) = \frac{\sigma^2}{\sqrt{4\pi D(t+t')}} \exp\left(-\frac{(x-x')^2}{4D(t+t')}\right),$$

which satisfies the homogeneous diffusion equation ($S(\boldsymbol{x},t) = 0$) in Eq. (1.2). In this approach of physics informed kernels the diffusivity D is treated as a hyperparameter and can be estimated similar to the characteristic length scale by maximization of the marginal likelihood. It is therefore of great interest and could be compared to the calculation by partial derivatives in future work.

6 Summary

In plasma physics, the spatial-temporal evolution of plasma parameters is of interest for transport studies. The goal of this work was the assessment of a 2D Gaussian Process (GP) for the fit of spatio-temporal evolutions of plasma parameters. In this work, a 2D GP was developed using artificial data and applied to experimental electron density measurements from the heliotron LHD. In contrast to 1D GPs and usually applied fitting methods, the spatial and temporal evolution can be fitted at the same time, making the description of spatio-temporal correlations possible. For calculating the covariance matrix a product of two squared exponential kernels was used.

The selection of spatial hyperparameters was done by maximizing the marginal likelihood. For 2D GPs, the size of training data is way larger, so the marginal likelihood $\log p(\boldsymbol{y}|X, \boldsymbol{\theta})$ diverges to infinity because of the term $-\log|\boldsymbol{K}_{\boldsymbol{xx}}|$. The covariance matrix was calculated for one time-slice and used for all times. This approach worked for the investigated cases, when assuming that the spatial hyperparameters do not change strongly over time. This assumption worked in this case, but should be reconsidered when there are strong changes in the density and temperature profiles.

An attempt was made to calculate a time dependent temporal correlation length $l_t(t)$ instead of keeping it fixed at a constant value, using the temporal change of the line integrated electron density $\frac{1}{L}\int n_e \mathrm{d}L$. The range of $l_t(t)$ was defined in the order of magnitude of $\Delta t \sim 30\,\mathrm{ms}$ and $3\tau_E \sim 300\,\mathrm{ms}$. The reconstruction showed strong oscillations when $l_t(t)$ suddenly changes. The oscillations do not appear when a fixed correlation length of $l_t = \Delta t$ is used, showing that the oscillations occur due to strong variations in $l_t(t)$ and not because of a too small correlation length. This can be fixed by adjusting the minimum (Δt) and maximum ($3\tau_E$) of $l_t(t)$. It is easier to adjust the maximum as the confinement time is implemented as a parameter of the 2D GP. However, when the minimum approaches the maximum and vice versa it results in a constant hyperparameter $l_t(t) = l_t$.

In general, application of minimal noise in the 2D GP leads to reconstructions matching the training data. The GP uncertainty and the residuals increase for greater noise values. In the case of no noise, i.e. $\varepsilon = 0$, the 2D GP shows residuals in the order of magnitude of 10^{11}. More training data is another way to improve the 2D GP fit, but it must be considered that more training data strongly increase the runtime. Additionally, the 2D GP approaches a constant value of mean residuals, therefore the fit cannot be improved arbitrarily by increasing the size of training data. It is conducted that if there are a lot of measurements, especially for the profile, it can be considered to downsample the amount of training data to reduce computational cost. Downsampling by a third can decrease the runtime by a third while increasing the mean residual by $\delta \approx (7.6 \pm 12.0)\,\%$. Because there are more spatial measurements than temporal measurements when using Thomson scattering measurements, downsampling the training data in time leads to larger mean residuals and uncertainties. When computational cost is not of importance, downsampling the training data is not recommended in general, to keep the information contained in the data.

By including the temporal dimension in the training data and thus in the covariance matrix, temporal information is taken into account. There are distinguishable differences between the reconstruction of a 2D GP and multiple combined 1D GP reconstructions. The joined 1D GPs do not show any temporal correlation in comparison to the 2D GP. This is also true for the fit uncertainties. We can see that the uncertainty of the 2D GP is more reasonable as it includes more information by considering more training data.

In conclusion, the 2D GP is applied as a method for fitting spatio-temporal data of plasma parameters. It is a non-parametric regression tool, which is suited for the application in plasma physics. Since there can be fast changes in the temporal evolution of plasma parameters, the temporal hyperparameter has to be small. Therefore, this method is not suitable for extrapolating the data beyond l_t. Instead, it can be used for interpolation. The 2D GP proved to be useful for the reconstruction of spatio-temporal experimental data, as it includes the temporal correlations as opposed to the 1D GP. However, the method of optimization for temporal hyperparameters discussed in this work, did not lead to the desired results, because of unexpected oscillations in the reconstructed data. A re-evaluation for the correct scale of $l_t(t)$ is necessary. If the 2D GP proves itself to be a reliable tool for investigating spatio-temporal dynamics of plasma parameters, it can be further expanded by a reliable optimization of temporal hyperparameters. In 1D profile fitting the disappearing flux at $r_{\text{eff}} = 0$ is considered when using 1D GP. As an outlook it is suggested that for the 2D GP, this needs to be included for a more physical treatment. Furthermore, the calculation of spatial and temporal derivatives using the 2D GP would enable the determination of further transport dynamics, e.g. the calculation of the diffusion coefficients.

References

[1] I. Fells. The Need for Energy. *Europhysics News*, **29**(6):193–195, 1998.

[2] A. Formicola, P. Corvisiero, and G. Gervino. The nuclear physics of the hydrogen burning in the Sun. *The European Physical Journal A*, **52**(4):73, 2016.

[3] J. Ongena. Fusion: A true challenge for an enormous reward. *EPJ Web of Conferences*, **98**:05004, 2015.

[4] Max-Planck-Institut für Plasmaphysik. `https://www.ipp.mpg.de/1456183/wendelstein_7_x` [Accessed: 05.03.2024].

[5] O. Kaneko. 15 - Large helical device. In G. H. Neilson, editor, *Magnetic Fusion Energy*, pages 469–491. Woodhead Publishing, 2016.

[6] A. E. Costley. On the fusion triple product and fusion power gain of tokamak pilot plants and reactors. *Nucl. Fusion*, **56**(6):066003, 2016.

[7] U. Stroth. *Plasmaphysik: Phänomene, Grundlagen und Anwendung*. Springer Spektrum Berlin, Heidelberg, 2. edition, 2017.

[8] M. Kikuchi, K. Lackner, and M. Q. Tran. *Fusion Physics*. International Atomic Energy Agency, Vienna, 2012.

[9] S. A. Bozhenkov et al. High-performance plasmas after pellet injections in Wendelstein 7-X. *Nucl. Fusion*, **60**(6):066011, 2020.

[10] A. Dinklage et al. Plasma Termination by Excess Pellet Fueling and Impurities in TJ-II, LHD and Wendelstein 7-X. *Nucl. Fusion*, **59**(7):076010, 2019.

[11] P. W. Terry. Suppression of turbulence and transport by sheared flow. *Rev. Mod. Phys.*, **72**(1):109–165, 2000.

[12] A. Ho et al. Application of Gaussian process regression to plasma turbulent transport model validation via integrated modelling. *Nucl. Fusion*, **59**(5):056007, 2019.

[13] C. K. I. Williams and C. E. Rasmussen. Gaussian Processes for Regression. In D. Touretzky, M.C. Mozer, and M. Hasselmo, editors, *Advances in Neural Information Processing Systems*, volume 8, pages 514–520, Cambridge, MA, USA, 1995. Max-Planck-Gesellschaft, MIT Press.

[14] T. Nishizawa et al. Estimation of plasma parameter profiles and their derivatives from linear observations by using Gaussian processes. *Plasma Phys. Control. Fusion*, **65**(12):125006, 2023.

[15] E. C. Howell and J. D. Hanson. Development of a non-parametric Gaussian process model in the three-dimensional equilibrium reconstruction code V3FIT. *Journal of Plasma Physics*, **86**(1):905860102, 2020.

[16] M. A. Chilenski et al. Experimentally testing the dependence of momentum transport on second derivatives using Gaussian process regression. *Nucl. Fusion*, **57**(12):126013, 2017.

[17] J. Melo. Gaussian processes for regression : a tutorial. Faculty of Engineering, University of Porto, 2012.

[18] A. Solin et al. Modeling and Interpolation of the Ambient Magnetic Field by Gaussian Processes. *IEEE Transactions on Robotics*, **34**(4):1112–1127, 2015.

[19] A. Mathews and J. W. Hughes. Quantifying Experimental Edge Plasma Evolution Via Multidimensional Adaptive Gaussian Process Regression. *IEEE Transactions on Plasma Science*, **49**(12):3841–3847, 2021.

[20] C. E. Rasmussen and C. K. I. Williams. *Gaussian processes for machine learning.* Adaptive computation and machine learning. MIT Press, Cambridge, Massachusetts, 2006.

[21] J. Wang. An Intuitive Tutorial to Gaussian Process Regression. *Computing in Science & Engineering*, **25**(4):4–11, 2023.

[22] K. P. Murphy. *Machine Learning: A Probabilistic Perspective.* MIT Press, Cambridge, Massachusetts, 2012.

[23] M. N. Gibbs. *Bayesian Gaussian Processes for Regression and Classifcation.* Phd thesis, University of Cambridge, 1998. Available: `https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=b5a0c62c8d7cf51137bfb079947b8393c00ed169`.

[24] M. A. Chilenski et al. Improved profile fitting and quantification of uncertainty in experimental measurements of impurity transport coefficients using Gaussian process regression. *Nucl. Fusion*, **55**(2):023012, 2015.

[25] S. Särkkä. Linear Operators and Stochastic Partial Differential Equations in Gaussian Process Regression. In T. Honkela, W. Duch, M. Girolami, and S. Kaski, editors, *Artificial Neural Networks and Machine Learning – ICANN 2011*, pages 151–158, Berlin, Heidelberg, 2011. Springer Berlin Heidelberg.

[26] C. G. Albert and K. Rath. Gaussian Process Regression for Data Fulfilling Linear Differential Equations with Localized Sources. *Entropy*, **22**(2):152, 2020.

[27] H. Funaba et al. Electron temperature and density measurement by Thomson scattering with a high repetition rate laser of 20 kHz on LHD. *Scientific Reports*, **12**(1):15112, 2022.

[28] T. Akiyama et al. Interferometer systems on LHD. *Fusion Science and Technology*, **58**(1):352–363, 2010.

[29] G. Fuchert et al. Increasing the density in Wendelstein 7-X: benefits and limitations. *Nucl. Fusion*, **60**(3):036020, 2020.

[30] M. Fujiwara et al. Plasma confinement studies in LHD. *Nucl. Fusion*, **39**(11Y):1659–1666, 1999.

[31] S. Kubo et al. Extension and characteristics of an ECRH plasma in LHD. *Plasma Phys. Control. Fusion*, **47**(5A):A81, 2005.

[32] K. Tsumori et al. High Power Neutral Beam Injection in LHD. *Plasma Science and Technology*, **8**(1):24, 2006.

[33] C. Suzuki et al. Shafranov shift measurements by a soft x-ray CCD camera for internal diffusion barrier discharges in the Large Helical Device. *Nucl. Fusion*, **50**(6):064013, 2010.

[34] P. K. Rai and S. Tripathi. Gaussian process for estimating parameters of partial differential equations and its application to the Richards equation. *Stochastic Environmental Research and Risk Assessment*, **33**(8):1629–1649, 2019.