Single Gaussian process method for arbitrary tokamak regimes with a statistical analysis

J Leddy\textsuperscript{1,*}, S Madireddy\textsuperscript{2}, E Howell\textsuperscript{1} and S Kruger\textsuperscript{1}\textsuperscript{,}\textsuperscript{\copyright}

\textsuperscript{1} Tech-X Corporation, Boulder, CO, United States of America
\textsuperscript{2} Argonne National Laboratory Lemont, IL, United States of America

E-mail: jleddy@txcorp.com

Received 18 February 2022, revised 11 July 2022
Accepted for publication 15 August 2022
Published 26 August 2022

Abstract
Gaussian process regression is a Bayesian method for inferring profiles based on input data. The technique is increasing in popularity in the fusion community due to its many advantages over traditional fitting techniques including intrinsic uncertainty quantification and robustness to over-fitting. This work investigates the use of a new method, the change-point method, for handling the varying length scales found in different tokamak regimes. The use of the Student’s t-distribution for the Bayesian likelihood probability is also investigated and shown to be advantageous in providing good fits in profiles with many outliers. To compare different methods, synthetic data generated from analytic profiles is used to create a database enabling a quantitative statistical comparison of which methods perform the best. Using a full Bayesian approach with the change-point method, Matérn kernel for the prior probability, and Student’s t-distribution for the likelihood is shown to give the best results.

Keywords: tokamak, equilibrium reconstruction, Gaussian process regression, EFIT, Thomson scattering

(Some figures may appear in colour only in the online journal)

1. Introduction
Achieving fusion energy using magnetic confinement requires understanding plasmas that are hotter than the Sun and their relationship to the magnetic field. This is not an easy task because diagnostics in this harsh environment requires understanding the physics of the diagnostic, and the plasma itself. This is known as an inverse problem: understanding the plasma requires understanding the diagnostic, which requires understanding the plasma.

For magnetic confinement, the small Larmor radius expansion gives to first order force balance: $J \times B = \nabla p$ \cite{1}. In tokamaks, this gives the Grad-Shafranov equation and solving the inverse problem described above is known as equilibrium reconstruction \cite{2, 3}. Equilibrium reconstruction has emerged as a critical tool for tokamak control \cite{4} and in plasma performance optimization \cite{5-8}. The most widely-used code for equilibrium reconstruction is EFIT \cite{2, 3}, which has been used at the DIII-D \cite{3, 9}, C-MOD \cite{10}, JET \cite{11}, HIT \cite{12}, START \cite{13, 14}, MAST \cite{15}, KSTAR \cite{16, 17}, JT-60U \cite{18}, NSTX \cite{19}, TORE SUPRA \cite{8, 20}, HT-7 \cite{21}, HL-2A \cite{22}, and EAST \cite{23, 24} devices. Equilibrium reconstruction is a critical part of the development of the fusion.

However, while traditional equilibrium reconstructions using only the magnetic data can be robust and routine, using additional data such as motional Stark effect \cite{6}, Thomson scattering diagnostic \cite{25, 26}, or soft x-ray \cite{27} are more problematic both due to the additional complexity additional of modeling internal diagnostics and diagnostic limitations. The best equilibrium reconstructions typically require
considerable manual intervention. For burning plasmas, where the harsh requirements and increased need for plasma control, more robust methods are needed.

Using a Bayesian approach to these problems offers many advantages [28, 29]. It more naturally allows one to account for the uncertainty in the data and for prior knowledge to be incorporated into the problem. The specific technique used in this paper will be Gaussian Process Regression, a technique that is becoming widely used in the uncertainty quantification and machine learning literature [30]. Its introduction into the fusion community was by Svensson [31] for soft x-ray tomography, but has since grown to include other diagnostics [31–35]. Its use within the fusion community has broadened to be used for verification and validation [36–39], and in equilibrium reconstruction as well [40, 41]. The latter is our ultimate goal.

One of the biggest drawbacks in using GPR is the same as that of using traditional fitting methods: how to account for very different profiles in the fitting. Most notably for tokamak fitting is the difference between the L-mode (low-confinement mode) and H-mode regimes. The H-mode regime is characterized by an edge pedestal that has a transport barrier giving short radial scale lengths. In the terminology of GPR (discussed more below), this regime is describable with a non-stationary kernel and its corresponding hyperparameters. That is, the short radial scale lengths imply that the radial correlation of the observations change. There are a few ways of handling this. The most popular method both within the larger ML community is known as warping, and is equivalent to a coordinate transformation in applied math, or non-uniform grids in numerical analysis [42]. In the fusion community, a tanh warping function was used to develop non-stationary hyperparameters for H-mode profiles by Chilenski et al [36]. This has been used by many authors [32, 34, 35, 37, 38, 43] since then.

The disadvantage of the warping approach to radially varying hyperparameters is the same problem that confronts the traditional fitting approach: one needs to know the type of fitting to perform ahead of time. This problem has been addressed in multiple ways. One method is to determine if the profile is L-mode or H-mode ahead of time, and then perform the appropriate traditional fitting [44], or hyperparameter fitting [39]. Other methods have also been developed to address this issue in the form of more complex warping functions: Beta CDF [45] and improved Kumaraswamy [46]. These have been shown to be capable of learning a wide family of bijective transformations via the addition of more hyperparameters. In this paper, we explore a different approach using a change-point kernel with the goal of using a single method regardless of the tokamak parameter regime. Our methodology for these investigations will also be different: we will focus strictly on synthetic analytic profiles using synthetically-generated data. Using this method allows us to analytically explore the effects of the distribution of the noise and the influence of outliers in the data in a quantitative way.

Our paper proceeds as follows. We first briefly review the basics of Gaussian Process Regression to introduce the basic concepts and terminology. Although this has been discussed in the fusion literature above, as well as other fields such as machine learning [30], we will be introducing new methods and need to clarify the context of these new methods. We next apply our new methods to the synthetic data characteristic of tokamak regimes and contrast with prior methods in the fusion community. Finally, we draw conclusions and point to future work.

2. Background

Traditional method to fitting a profile \( f(\psi) \) typically involves finding a function that depends on a set of parameters and minimizing the error. In tokamak devices, profiles are usually written in terms of the poloidal magnetic flux, \( \psi \), as many important physical quantities of interest are a function of only this parameter. For example, one can write \( f(\psi) = \sum f_i \alpha_i \), where \( f_i \) is a set of \( N \) parameters, and \( \alpha_i \) is the fitting function for those parameters, and then minimize the error related to the data. This is the traditional method of EFIT which uses spline functions for the fitting functions with L-mode profiles, and splines and a tanh function for H-mode profiles. Advanced techniques for fitting using the OMFIT [47] are discussed in [48, 49].

In the Bayesian view, the process is fundamentally different. One first calculates the probability of a function being correct, and the most probable function is used as that fit. Mathematically, this is written as the \( p(f(\psi) | d) \): the probability of \( f(\psi) \) given the data \( d \). This is a conceptual shift compared to traditional fitting. It acknowledges that there is uncertainty in the data, and uncertainty in \( f(\psi) \) itself. To find this probability is not straightforward directly. Instead, we use Bayes’ Theorem:

\[
p(f(\psi) | d) = \frac{\text{likelihood}}{\text{marginal likelihood}} \cdot p(f(\psi)).
\]

The advantage of Bayes’ theorem, in practical terms, stems from the fact that it often straightforward to directly express the likelihood in terms of an explicit model, but rarely do we have an explicit model for the posterior. Thus Bayes’ theorem relates a quantity of interest to a quantity the we can directly model. The marginal likelihood is a normalizing constant that ensures the total probability is one. So with Bayes’ theorem it is possible to obtain \( p(f(\psi) | d) \) when it would have otherwise been very difficult.

The first step that we wish to consider is how to encode the space of all functions into the prior probability. Above, we discussed how parameters were used to parameterize a function. A similar method is used to explore the distribution space of the probability functions here. In PDE’s, exploring a functional space is done using discretizations (e.g. finite difference). The problem with this for probability functions is that although we might wish to assume that \( f(\psi) \) lies within a functional space (e.g. Hilbert space), the data, which is subject to noise, does not. Using traditional PDE discretizations would also subject one to systematic bias in determining the distributions. Hence, randomized sampling
of the distribution space is generally preferred. Two methods for exploring this space are polynomial chaos expansions and stochastic processes. In both methods, the conversion of the continuum form of the probabilities above are converted to discrete form. For distributions, hyperparameters, denoted $\theta_{hp}$, are used to explore the probability space for the prior and likelihood functions.

In this paper, we will consider a specific form of a stochastic process known as the Gaussian Process. Intuitively, a Gaussian process is one that can generate random numbers (definition of a stochastic process) that are distributed normally. Mathematically, we write the prior sampling as

$$p(f(\psi)) = GP(\mu(\psi), K(\psi, \psi'))$$

where $\mu$ is the mean function of the process and $K$ is the covariance function, or kernel. A common Gaussian process uses a zero mean and the squared exponential kernel (a.k.a. radial basis function):

$$K_{SE}(\psi, \psi') = e^{-\frac{(\psi - \psi')^2}{2\theta^2}}.$$  

The Gaussian process is distribution over all smooth functions of $\psi$. Each function in the Gaussian process occurs with a certain probability that depends on the two hyperparameters $\theta_1$ and $\theta_2$. The hyperparameter $\theta_1$ characterizes how rapidly the probability decreases as the function departs from the mean function ($0$ in this case), and $\theta_2$ characterizes the spatial correlation of the function at two points $\psi$ and $\psi'$. A kernel is stationary if it only depends on the distance between two points $K(\psi, \psi') = K(|\psi - \psi'|)$, otherwise the kernel is non-stationary.

In the context of fitting profiles of fusion data, the discussion of whether a Gaussian process is stationary or not relates to the characteristic length scale of the profile, or data, to be inferred. For H-mode profiles, the short length scale of the pedestal region motivates the use of non-stationary kernels; e.g. as used in Chilenski et al [36]. Similarly, inferring magnetic data can be challenging due to rapid changes in the data that require non-stationary kernels [33].

The hyperparameters play an important role in GPR, and the choice of hyperparameters requires special consideration. The simplest approaches is to pick reasonable values for the hyperparameters that are learned from prior experiments. In this approach, a large enough data set can overcome a bad choice of hyperparameters. However, this approach suffers if underlying profile changes drastically in different experimental regimes or if the amount of measured data is limited, both of which are case here. Thus, this approach is not used in practice.

A second approach is the Empirical Bayesian where one uses the data to determine the optimal hyperparameters. Also called Type-II maximum likelihood, in this method the estimation of the hyperparameters from the data acts as a regularizer on the prior over the parameters [50]. There is a family of methods in this approach. One method chooses the likelihood to also be a Gaussian function. The combination of Gaussian functions in both the likelihood, and a squared-exponential kernel enables an analytically-tractable method for maximizing the marginal likelihood. This method was used in the first introduction of GPR to the fusion community [31]. When the likelihood is not a Gaussian, a closed-form analytic solution in general is not possible, and one must utilize numerical methods for this calculation.

A different philosophical approach is the Full Bayesian method in which the prior distribution is fixed before any data are observed. In this approach, the prior and likelihood are still described in terms of hyperparameters as described above, but the full hyperparameter space is explored numerically by integrating out the hyperparameters. Specifically, the likelihood and prior probability functions in equation (1) are parameterized in terms of hyperparameters, and then the hyperparameters are numerically integrated out in a process known as marginalization. There are a wide variety of approaches to marginalization or in using partial approximations to accelerate this integration. In this paper, we will use Markov chain Monte Carlo (MCMC) methods to systematically perform the Bayesian integration. In the fusion community, Chilenski is a notable paper that discusses the Full Bayesian technique as well as using non-stationary hyperparameters. Empirical Bayesian remains popular, however, because of its computational speed compared to Full Bayesian techniques. This is discussed in more detail later.

In this paper, we will systematically explore different approaches to using Bayesian fitting of profiles of varying length scales, and using both Empirical Bays and Full Bays using MCMC. The use of Full Bays will be required as we explore the use of different likelihood functions in fitting noisy data. We turn to the method of determining the best method next.

3. Overview of methods

The Gaussian Process Regression methods that are discussed in this section have been implemented in Python and are available in a public repository: UNified Bayesian Analysis Framework for Fusion ExperimentA.L Data (UNBAFFEILD) [51]. This library is under continuous development and will be augmented with new and useful tools for Bayesian analysis of fusion data.

3.1. Generation of synthetic data for tokamak parameter regimes

Our goal is to understand generic issues related to using the GPR method to infer the the density ($n$) and temperature ($T$) profiles of tokamaks from localized measurements such as Thomson scattering. However, the profiles characteristic of L-mode, H-mode, internal transport barrier (ITB) regimes can have varying length scales providing challenges to the GPR method. Experimentally, tokamaks account for this varying resolution by having TS systems with a spatial resolution that varies across the domain in order to capture the sharp features associated with a pedestal. For example, the DIII-D tokamak core TS system has a coarser spatial resolution which is adequate for resolving the slowly varying core $n$ and $T$ profiles.
and a finer resolution in pedestal to resolve the sharp profile gradients in this region [25].

Here, rather than use raw experimental data as in prior work [31–35], we create synthetic data to enable a more accurate understanding of the errors associated with different GPR methods. Specifically, for L-mode profiles we use

\[ f(\psi) = f_o \left(1 - \psi^{\alpha_1}\right)^{\alpha_2} + f_{\text{edge}} \]

where the input parameters are \(f_o\) (value at the center), \(f_{\text{edge}}\) (value at edge), \(\alpha_1\) controls the gradient near the center, and \(\alpha_2\) controls the gradient near the edge. The H-mode profile is given by:

\[ f(\psi) = f_o \left(1 - \psi^{\alpha_1}\right)^{\alpha_2} + f_{\text{edge}} + \frac{1}{2}f_{\text{ped}} \left(1 - \tanh \frac{\psi - \psi_{\text{ped}}}{w_{\text{ped}}} \right) \]

where \(f_{\text{ped}}\) and \(w_{\text{ped}}\) control the pedestal height and width.

The synthetic data is created by taking 40 points uniformly across the domain, and then another 40 points uniformly across the pedestal region to account for the way in which experimental measurements are often localized in the pedestal. The data at the discrete profile points is then perturbed using a random numbers to generate noise assuming a Gaussian distribution with a specified width (standard deviation). In this paper, the width varies from 10% to 33% of the data value itself. In addition, the noise can be shifted uniformly; i.e. given a non-zero mean in the Gaussian, to represent systematic errors. Outliers are synthetically created by choosing random points and using random values that have a Gaussian width of 2–3 times the data value. If a negative outlier is generated, then it is made positive. As seen below, this can create an additional systematic error. The synthetic ITB profiles are similar to the H-mode profiles except that the tanh profile is moved to the center of the plasma.

Though not discussed in detail in this paper, it is useful to consider the noise that will be seen in actual experimental data for Thomson scattering. The noise on the data is set by counting statistics of photons, which depends on a variety of plasma parameters, such as density, temperature, the quantum efficiency of the detector, and the optical line of sight. Different Thomson systems utilize different hardware, and even within a single system there are variations in the collection efficiency. There is also background noise from Bremsstrahlung, line radiation, and stray light, which also depend on the plasma parameters [52]. There are also systematic errors that arise from mapping of the measurements onto 2D flux surfaces, and even some device-specific due to error fields. Fortunately, we know a lot about the errors in the data, and so we should be able to take advantage of this knowledge when building a model for the noise. As much of this noise is based on the current state of the plasma, the noise is heteroscedastic, and there are developed methods for incorporating noise models in GPR for this type of noise [53, 54]. We plan to explore these methods and incorporate experimental noise analysis in a future paper where the GPR methods discussed within this current paper are applied to experimental data.

3.2. Choosing a kernel

As discussed in section 2, the standard basic kernel for GPR is the squared exponential kernels as given by equation (3). The hyperparameters in these kernels do not have to be chosen a priori, but can instead be inferred from the data themselves. Typically, a stationary kernel is used, but for H-mode profiles, Chilenski et al [36], using a a non-stationary kernel derived by Gibbs [55], used a tanh function for the hyperparameters. This addition can make the computation of the fit expensive, and additionally must be constrained or convergence becomes difficult. We desire a robust fitting routine, so we opted for a second option that is available in GPflow [56], the software that we use, called the change-point kernel. This is effectively a piecewise function for the kernel where multiple individual kernels can be defined and locations can be specified for when one kernel changes to another. The derivatives of the total kernel is ensured to be continuous by the use of a transfer function between the piecewise parts of the kernel, which smoothly transitions between the sub-kernels.

The H-mode pedestal occurs in the edge region and so we allow one kernel to describe \(\psi \in [0, 0.9]\) and \(\psi \in [1, 0, 1.1]\), and then a separate kernel defines \(\psi \in [0.9, 1.0]\). Each of these distributions has independent hyperparameters and therefore can represent a profile that has varying correlation length across the profile. Since we use the data to infer the hyperparameter values, this actually also works well to describe an L-mode profile that has a single effective correlation length. In that case, both kernels would end up with similar values for the hyperparameters. Figure 1 demonstrates the way this kernel has been constructed, as well as the data that illustrate the need for the change-point kernel.

The individual kernels chosen for kernels A and B (as indicated in figure 1) are the well-established Matérn covariance functions. These are designed to indicate the correlation between points given some distance, \(d = |\psi - \psi'|\), via the following equation:

\[ C_{5/2}(d) = \theta_\nu^2 \left(1 + \frac{\sqrt{5}d}{\theta_\nu} + \frac{5}{3}\frac{d^2}{\theta_\nu^2}\right) \exp\left(-\frac{\sqrt{5}d}{\theta_\nu}\right), \]

which is a particular instantiation of the general Matérn kernel with \(\nu = 5/3\). In this formula, \(\theta_\nu\) is indicative of the width of the covariance function (like the standard deviation in a Gaussian) and \(\theta_\nu\) is a normalization parameter that indicates the degree of the correlation between points. These two values are hyperparameters of the kernels that can be independently optimized for each kernel (A & B). In regions of large gradients/fluctuations we expect \(\sigma_\nu\) to be small, while in regions of small gradients we expect it to be larger.

3.3. Handling outliers

Outlying data can negatively affect the fit and must be handled with care. One obvious but labor intensive approach is to manually go through the data to identify outliers or ‘bad’ data and remove them. This is, however, not ideal for multiple reasons. Firstly, it takes time and effort, but we would like the...
fitting to be an automated process. Secondly, sometimes data can look bad but be accurate, and it is not always possible to determine which is the case. Finally, we would rather be able to incorporate all the data but keep track of the errors instead of eliminating data because they look bad. Therefore, we conclude that we want to keep all the data, but we do not want outliers to negatively affect the fit. Instead, we want them to affect the error of the fit. In the Bayesian framework, this is handled by the likelihood probability.

In general, when letting the data inform the choice in hyperparameters, we want to maximize the log likelihood of the fit given the data. For data with normally distributed noise, a squared exponential likelihood is used as it is appropriate, and it comes with the added benefit that the problem can be reduced analytically to an easily tractable form. However, if there are outliers in the data, and/or the noise is non-Gaussian this is no longer an appropriate choice for the likelihood function. A heavy-tailed distribution can capture the error on the outliers without affecting the quality of the fit for the data points with lower error.

We have tested multiple heavy-tailed distributions including the Student’s t-, logistic, and Laplace distributions. All of these function adequately, as we will show, but the Student’s t-distribution is the one performs best. It is given by:

$$L(x) = \frac{\Gamma \left( \frac{\nu+1}{2} \right)}{\sqrt{\nu\pi}\Gamma \left( \frac{\nu}{2} \right)} \left( 1 + \frac{x^2}{\nu} \right)^{-\frac{\nu+1}{2}}$$  \hspace{1cm} (7)

where $\nu$ is the degrees of freedom and $\Gamma$ is the gamma function. This is a particularly nice function because it approaches a Gaussian as $\nu \to \infty$. We can allow the degrees of freedom hyperparameter to be optimized during the GPR so that if the errors on the data are genuinely Gaussian then the value for $\nu$ will end up being large and this will capture the errors properly. However, if we have outliers that would skew the fit, the $\nu$ parameter will adjust automatically to be smaller, increasing the weight of the likelihood function tails, allowing the quality of fit to remain high while minimizing over-fitting. This does not come without a cost, however, since we can no longer make the same analytic simplifications to optimize hyperparameters that can be done with a pure Gaussian likelihood. Instead, we resort to MCMC to sample the hyperparameters space of both the kernel and the likelihood simultaneously. Instead of optimizing, then, we obtain a distribution for each of the hyperparameters, and then sample from them to obtain possible fits. The final fit we choose is the mean of all the sampled fits.

4. Results

We will discuss two regimes that we fit explicitly: L-mode and H-mode. We do this because these two cases are illustrative of the reasoning behind our choice of kernel and likelihood functions. We then go on to discuss the statistics of many fits.
Figure 2. Fit of an L-mode profile using full Bayesian approach with the change-point kernel described in section 3.2 and the Student’s t-distribution likelihood described in section 3.3.

Figure 3. Histograms of the hyperparameter distributions as found by the MCMC algorithm for the L-mode fit. The y-axes are counts from the MCMC chain steps and the x-axes are the values for the variable in the title of the plots. Although the analytic profile has a smoothly-varying length scale and fixed length scale, several length scales can be found in the noisy data illustrating the problems of standard data fitting techniques.

4.1 Fitting in various regimes

In all of the regimes we discuss below, the kernel and likelihood function are identical. The automated tuning of the hyperparameters is all that differs, and so we will discuss those results and show the distributions for all the hyperparameters. In this way, all of these regimes are fit without any human intervention or regime detection algorithms. The data themselves inform the choice of hyperparameters for the GPR.

4.1.1 L-mode. In the L-mode, as discussed above, the profiles are smooth. Figure 2 shows an example of a GPR fit for an such an L-mode profile. Despite the noise in the data, the fit follows the underlying profile closely. We can hypothesize the affect of such a profile on the hyperparameter distributions. Firstly, it is expected that the hyperparameters of both kernels are similar since the correlation length across the profiles is roughly unchanged. Additionally, we expect the degrees of freedom parameter to roughly correlate with the number and size of the outlying data. Figure 3 shows the distributions for some of the hyperparameters as sampled using MCMC. The hyperparameter $L_p$ represents the correlation length in what would be the pedestal region in H-mode ($\psi \in [0.9, 1]$) and corresponds to the hyperparameter $\theta_l$ for the pedestal kernel in equation (6). This peaks around 0.8, but is roughly uniform across the 0.1–1.2 range. A value of 0.1 is of similar size to the pedestal width itself indicating that all values over this result in a smooth fit across the pedestal region. The hyperparameter $L_c$ represents the correlation length for the core (corresponding to the hyperparameter $\theta_l$ for the core kernel in equation (6)) and edge region and peaks around 0.5. Due to the slight kink even in L-mode profiles, we do expect to see this small decrease in correlation length in the pedestal region, but for H-mode we would expect to see this value to be much lower to account for the steep pedestal. The degrees of freedom hyperparameter, $\nu$, for the Student’s t-distribution likelihood
distribution, as defined in section 3.3. This distribution peaks around 1.5, which means it is much closer to a Cauchy distribution than a Gaussian, so confirms the need for such a heavy-tailed distribution to adjust the fit to account for the outliers.

4.1.2. H-mode. The change in length scales of the pedestal region is the motivation for using the change point kernel discussed in section 3.2. Figure 4 shows the pedestal around $\psi = 1$, and one can see that the remaining profile is simply the same as the L-mode profile with an offset. The combination change point kernel allows this profile to be fit with the same kernel used to fit the L-mode profile in section 4.1.1.

Figure 5 shows the distributions for the hyperparameters for the H-mode fit. The length scales distributions show the key feature of our kernel choice. The hyperparameter $L_c$ is the hyperparameter for the correlation length in kernel A, which spans the majority of the spatial domain, and the majority of the values are between 0.5 and 1. Interestingly, this is a bi-modal distribution confirming the need for the full Bayesian approach to sample this distribution. In the $L_p$ distribution, which is the distribution for the correlation length hyperparameter in kernel B which spans $\psi \in [0, 1]$, most of the values are much smaller, below 0.1. This is what allows the fit to capture the gradient changes in the pedestal region equally as well as the rest of the domain is fit. Again, like in the L-mode fit, the degree of freedom parameter, $\nu$, peaks around 1.5 so that outliers do not shift the mean fit.

4.1.3. H-mode with ITB. The last case we look at explicitly is for an H-mode profile that contains an ITB. This is another region of steep gradient but in the core of the plasma instead of the edge, where the pedestal can exist. In our synthetic data, this will occur where the data is more sparse than the pedestal region, similar to real experiments. This case is not explicitly fit by the kernel we introduce in section 3.2, but it could be extended to do so by adding a third kernel that covers a region...
around \( \psi = 0.5 \). This would increase the number of hyperparameters, therefore the dimensionality of the parameter space through which the MCMC must sample. It is still doable, but is saved for future work. Since this is not done, the purpose of this section is to show that this method performs well even without these additions. Figure 6 shows an example of a fit of this type of profile. Though the gradient of the ITB is not perfectly fit by GPR, an inflection in the fit is seen as the fit attempts to capture the inflection in the underlying profile. This fit better captures an ITB than a parameterized method for H-mode would, because there is no over-fitting, and yet requires no knowledge that such an ITB might exist. This applies to any non-standard profile, as well—GPR is able to smoothly fit unexpected features in the underlying profile without making assumptions about tokamak regimes. Figure 7 shows the hyperparameter marginalized distributions for the major hyperparameters of the H-mode with ITB fit. As expected, the core length scale distribution extends to lower values and has its largest peak at these lower length scales as compared to the pure H-mode case. This is because the steep gradient of the ITB requires these smaller length scales for the fit to capture the associated changes in gradient.

4.2. Statistical analysis

In this section we perform a statistical analysis of various GPR fitting methods on a large set of synthetic data that spans a broad parameter space in the profile shape and noise. The four methods we test here cover two kernels, two likelihoods, and full vs empirical Bayes. As described in section 1, empirical Bayes involves an analytic simplification assuming a Gaussian likelihood, allowing for simple hyperparameter optimization. Full Bayesian, on the other hand, does not make this simplification so a method like MCMC must be used to sample the hyperparameter space (likelihood function is not restricted). The Chilenski kernel [36] is used in the fusion community used for GPR regression of H-mode profiles for equilibrium reconstruction. It has been implemented for use in OMFIT [47] with OMFIT profiles [48]. It is a standard Gibbs kernel, but with the length scale function defined and parameterized by a
tanh function to represent two length scales with a transition region. We refer to this kernel and method as Chilenski. We refer to the kernel proposed in section 3.2 as the change-point kernel in the following sections for brevity. The four methods/cases we analyze are:

(a) Empirical Bayes with Chilenski kernel (Gibbs)
(b) Empirical Bayes with change-point kernel (change-point)
(c) Full Bayes with change-point kernel (implied Gaussian likelihood)
(d) Full Bayes with change-point kernel and Student’s t-distribution likelihood

It is possible for the Chilenski kernel to be used with empirical and full Bayesian methods, but herein we only use empirical Bayes with the Chilenski kernel.

4.2.1. Building database of fits. In order to analyze the quality of fit provided by the methods described in the previous sections, we have performed a large number of fits on a range of synthetic profiles. There are two parts to doing a study like this. First, a figure of merit must be defined to determine and compare the quality of the fits. Secondly, a parameter space must be defined to generate the range of synthetic profiles.

For the first, there are many figures of merit that can be used to describe the quality of fit. We desire one that does not depend on the number of points in the resulting fit, so must therefore include a mean. The root mean square error (RMSE) is a reasonable choice as it represents the standard deviation of the residuals between the fit and the true, underlying profile. It is important to note that the Chilenski method also aims to obtain high quality fit derivatives for transport model calculations. In this paper we are using the RMSE which only compares the quality of the data fit, not including any of the extra features involving derivatives that Chilenski developed. We define this as:

$$\text{RMSE} = \sqrt{\frac{1}{N} \sum_{i=1}^{N} (y_i - p_i)^2}$$  \hspace{1cm} (8)$$

where $y_i$ is the value of the fit and $p_i$ is the true profile value. Because these profiles are synthetically generated with noise added, we know the exact, underlying profile making this analysis straightforward and insightful beyond what we can know analyzing experimental data.

Additionally, we have varied a series of parameter associated with the generation of synthetic data so that in the end we fit 5280 different profiles. The set of parameters are

(a) noise amplitude: $\sigma_n \in [0.1, 0.15, 0.2, 0.25, 0.33]$
(b) noise shift (systematic error): $\mu_n \in [0\%, 2\%, 5\%, 10\%]$
(c) outlier number: $N_{OL} \in [0, 3, 5, 10]$
(d) outlier amplitude: $\sigma_{OL} \in [2, 3, 4]$

The number of points generated for these synthetic profiles is 88, so the number of outliers can also be written as a fraction of the total points: $N_{OL} \in [0\%, 2.2\%, 5.7\%, 11.3\%]$. Profiles were generated for H-mode, L-mode, and H-mode with ITBs for these four parameter variations. Additionally, for H-mode we added the following parameters

(a) edge density: $n_{edge} \in [0.01, 0.05, 0.1, 0.2]$
(b) pedestal width: $w_{ped} \in [0.01, 0.015, 0.02]$

Finally, for H-mode with ITB we added the following additional parameters

(a) ITB width: $w_{itb} \in [0.01, 0.015, 0.02]$
(b) ITB height: $n_{itb} \in [0.5, 1.0, 1.5]$

We then fit these synthetically generated profiles with four methods

(a) Chilenski kernel (Gibbs with tanh length scale function) with empirical Bayes assumptions
(b) Change-point kernel (described in section 3.2) with empirical Bayes assumptions
(c) Change-point kernel with Gaussian likelihood with full Bayesian approach
(d) Change-point kernel with Student’s t-distribution likelihood with full Bayesian approach

Between these profiles, methods, and the figure of merit we were able to build up a database of over 5000 sets of synthetic data from that enable us to perform a quantitative comparison of the different methods.

4.2.2. Worst fits for each method. In order to put the RMSE values into context, we first want to show what constitutes a ‘bad’ fit using these methods. An advantage of GPR is that when hyperparameter optimization is successful, even the bad fits are smooth and generally still acceptable fits. This is not always the case for parameterized fitting methods that can fail spectacularly due to over-fitting. To examine these bad cases, we have found the worst fit (highest RMSE) for each method and then plotted that fit along with the fit for the same data using all the other methods.

These plots are shown in figure 8. The worst fit using the empirical Bayesian method with the Chilenski kernel (top left plot) significantly over fit the data in the core and smoothed over the pedestal. This is due to the outliers that are close to each other making the hyperparameter likelihood function multi-modal—we are most likely seeing a local maximum instead of the global. This indicates a key problem that our outlier solution is able to address. This is addressed by our method, but One could also imagine making the assumption that the core correlation length must always be larger than the pedestal/edge correlation length, thus constraining the Chilenski method so that this type of fit could not occur. However, we invoked the Chilenski method using the same safeguards as in OMFIT [47], which would need to be modified for robustness. The data in this fit are not particularly difficult to fit, as the other methods do a relatively good job in fitting them, though their means (excluding the Student’s t-distribution likelihood method) are pulled above the
true profile due to these outliers in the core region. Therefore the strength of the Student’s t-distribution likelihood is that it addressed multiple problems that outliers can cause in data fitting—affecting the mean and confusing the length scale hyperparameter optimization.

The worst fit using the other three methods show outlying cases where the data happen to have all the outlying data points above the underlying profile. This is an edge case since for the synthetic data the outliers have equal chances of lying above and below the true profile. In all these cases the mean fit are pulled by the outliers away from the underlying profile by these outliers. Additionally, these synthetic data profiles also sit in the parameter space where the noise offset is non-zero. This is to be expected, since none of these methods are able to intrinsically detect systematic errors.

Importantly, all of the proposed (non-Chelinski) methods, even in the worst case, capture the pedestal gradient well and therefore are able to identify pedestal location and width. These are key parameters that EFIT uses to continue forward with the equilibrium calculation. Additionally, even these worst cases result in smooth fits that would be unlikely to cause calculation errors as a part of an equilibrium reconstruction workflow.

4.2.3. Effect of outlier mitigation. We introduce in section 3.3 a method for minimizing the impact of outliers on the fit by using a Student’s t-distribution likelihood. This heavy-tailed distribution accounts for outliers without shifting the mean. Figure 9 shows the benefits of such an approach compared to the rest of the analyzed methods.

It is clear from this plot that the full Bayesian approach provides much better fits on average than the empirical Bayesian approach. Additionally, each of these lines is seen to trend up in error as more and more outliers are introduced, except for the method that includes outlier mitigation via the Student’s t-distribution likelihood. This method is roughly invariant to the number of outliers, at least in the range shown here which represents a maximum of over 11% outliers in the data. The large difference in the performance of the Full Bayes compared to the Empirical Bayes can be clearly seen in these plots, as the RMSE is an order of magnitude lower for the...
Figure 9. Plots of the average and standard deviation for the RMSE of the fits across all synthetically generated profiles illustrate the effect of outliers on the RMSE for the fits. Each fitting method is shown in a different color and the size of the outliers is $2\sigma$ (left), $3\sigma$ (middle) and $4\sigma$ (right). Note the flatness of the fit using the Student’s t-distribution likelihood, indicating a relative independence of the fit on number of outliers.

Figure 10. These plots show the distribution of RMSE for the various methods for H-mode and L-mode. It is important to note that the RMSE axes differ in scale for the plots—the full Bayesian approach consistently produces fits with lower RMSE than the empirical Bayesian approach.

Full Bayes approaches. After examining the hyperparameter distributions, we expect that this difference is largely due to both the irregular shapes of these distributions (some are multi-modal, others highly skewed), so simply choosing the maximum quantity (as is effectively done in Empirical Bayes) does not capture the variety in the fits that taking the mean of a set of Full Bayesian fits does. These non-Gaussian distributions with multiple peaks show why it can be difficult for a global optimization method (type II maximum likelihood) to capture the best fit.

4.2.4. Fitting L-mode versus H-mode. One of the objectives behind developing the fitting method described in this paper (change-point kernel with student-T likelihood) is to fit H-mode and L-mode equally well without having to know a priori anything about the operational mode of the tokamak from which the data come. This is accomplished through GPR with variable length scales in the kernel and hyperparameter optimization/sampling. However, we want to test how well these methods perform at fitting L-mode and H-mode to make sure that the methods are as effective as they should be.

Figure 10 shows the distribution of RMSE for both L-mode and H-mode separately, and for each fitting method separately. There are a few key points worth emphasizing from these plots. Firstly, it is clear that the empirical Bayesian does a worse job, regardless of kernel, for fitting H-mode than it does for fitting L-mode. This is likely due to the large parameter space involved and ending up in a local maximum instead of global for the likelihood maximization. Also, the fit for H-mode is more difficult as the gradients are steeper, so the parameter space is likely also steeper and multi-modal. Secondly, the Full Bayesian approach sees minimal difference between fitting L-mode and H-mode, and the overall RMSE is lower for these than for the empirical Bayesian fits. The multi-modal behavior seen in this figure is due to the large parameter space encapsulated within the data. The peaks at higher RMSE are due to noise offsets and large outliers, while the peaks at smaller RMSE are easier profiles to fit. Particularly, we expect that the peaks in the change-point Full Bayes Student-t plot are due to the lack of smoothing provided by outliers throughout all the data sets. As outliers have a negligible impact the RMSE when the Student-t likelihood is used, we are more clearly able to see the effect of the other parameters as peaks in this histogram.

5. Discussion

This paper used synthetic data with the ability to control for the amount of noise, the degree of outliers, and the effect of systematic bias to perform a statistical analysis of various GPR methods. These methods include the non-stationary empirical Bayes method introduced by Chilenski et al [36], as well as an alternative to non-stationary hyperparameters:
the change-point kernel (a linear combination of two stationary kernels). The effect is similar, however, in that it allows multiple length scales to be chosen for GPR and applied to different locations on the $\psi$ axis. The change-point kernel we introduced is, however, more flexible than a Gibbs kernel with a tanh length scale function as it allows any number of length scales that can change at arbitrary positions in $\psi$.

The most important new method introduced here is the Student’s t-distribution for the likelihood function. This allows the fitting to handle outliers reliably and without any prior knowledge as to which points are outliers and how many there may be. The use of the Student’s t-distribution rules out the Empirical Bayes method used here, but is easily implemented in our Full Bayesian method using GPyflow. We compared Full Bayesian with the standard Gaussian likelihood to the Student’s t-distribution and found that both were more accurate than the Empirical Bayesian techniques.

To be truly quantitative in our judgements, a statistical analysis using the various methods we described to compare their effectiveness across a range of profiles including L-mode, H-mode, and H-mode with ITB. This also included a variety of settings on the noise including amplitude, mean, outlier number, etc to allow an analysis of over 5000 sets of generated data. However, fundamentally the synthetic data is generated with Gaussian noise such that it should match well to the commonly used square exponential kernel. The noise was varied to allow a shifted Gaussian, representing systematic errors in a diagnostic, as well as outliers, representing bad channels in a diagnostic.

The most important results of this statistical study can be seen in figure 9. H-mode profiles are consistently more difficult to fit with empirical Bayes method, with a consistent RMS error that is approximately 10% higher than L-mode. There are also differences between L-mode and H-mode profiles in the Full Bayesian methods, but the differences are slight. The change-point kernel is slightly better than the non-stationary hyperparameter tanh profiles of Chilenski, but the difference is slight. Full Bayesian methods are an order of magnitude more accurate Empirical Bayesian methods, but comes at a computational cost that is roughly an order of magnitude more expensive. Although we did not test it, we would expect the Chilenski kernel to perform similarly to our own using the full Bayesian method since it also performed similar to our own using empirical Bayes. The Student’s t-distribution function for the likelihood is roughly twice as accurate as the Gaussian likelihood when there are a large number of outliers. The lack of sensitivity to outliers by the Student’s t-distribution can be seen in figure 9. It is important to note that even the worst fits provided by our proposed methods result in smooth, reasonable fits to the data and are robust.

In future work, we will continue to refine these methods as we apply them to experimental data in a similar statistical manner. When working with experimental data, however, we do not have the luxury of knowing the nature of the true profile. Fortunately, the study that we have done here can give us confidence that we are able to capture the underlying profile with a fit, barring the existence of systematic offset in the noise.

Data availability statement

The data that support the findings of this study are openly available at the following URL/DOI: https://gitlab.com/efit-ai/unbaffeld/-/tree/main/unbaffeld/gpr.

Acknowledgments

This material is based upon work supported by the Department of Energy under Award Number(s) DE-SC0021380. We wish to thank Dr Lang Lao, Dr Joseph McClanaghan, and the EFTF-AI Team for useful discussions related to this work.

ORCID IDs

J Leddy https://orcid.org/0000-0002-4492-7858
S Madireddy https://orcid.org/0000-0002-0437-8655
E Howell https://orcid.org/0000-0002-9203-3592
S Kruger https://orcid.org/0000-0001-9174-6227

References

[1] Hazelton R D and Meiss J D 1992 Plasma Confinement (Redwood City, CA: Addison-Wesley Publishing Company)
[2] Lao L L, St John H, Stambaugh R D, Kellman A G and Pfeiffer W 1985 Reconstruction of current profile parameters and plasma shapes in tokamaks Nucl. Fusion 25 1611–22
[3] Lao L L, St John H E, Peng Q, Ferron J R, Strait E J, Taylor T S, Meyer W H, Zhang C and You K I 2005 MHD equilibrium reconstruction in the DIII-D tokamak Fusion Sci. Technol. 48 968–77
[4] Ferron J R, Walker M L, Lao L L, St John H E, Humphreys D A and Leuer J A 1998 Real time equilibrium reconstruction for tokamak discharge control Nucl. Fusion 38 1055
[5] Strait E J 1994 Stability of high beta tokamak plasmas Phys. Plasmas 1 1415–31
[6] Levinton F M et al 1995 Improved confinement with reversed magnetic shear in TFTR Phys. Rev. Lett. 75 4417–20
[7] Strait E J et al 1995 Enhanced confinement and stability in DIII-D discharges with reversed magnetic shear Phys. Rev. Lett. 75 4421–4
[8] Zwirgmann W 2003 Equilibrium analysis of steady state tokamak discharges Nucl. Fusion 43 842–50
[9] Lao L L, Ferron J R, Groebner R J, Howl W, St John H, Strait E J and Taylor T S 1990 Equilibrium analysis of current profiles in tokamaks Nucl. Fusion 30 1035–49
[10] In Y, Ramos J J, Hubbard A E, Hutchinson I H, Porkolab M, Snipes J, Wolfe S M and Bondeson A 2000 Resistive $n = 1$ modes in reversed magnetic shear Alcator C-Mod plasmas Nucl. Fusion 40 A1463–8
[11] O’Brien D P, Lao L L, Solano E R, Garribba M, Taylor T S, Cordey J G and Ellis J J 1992 Equilibrium analysis of iron core tokamaks using a full domain method Nucl. Fusion 32 1351–60
[12] Nelson B A, Jarboe T R, Orvis D J, McCullough L A, Xie J, Zhang C and Zhou L 1994 Formation and sustainment of a 150 kA tokamak by coaxial helicity injection Phys. Rev. Lett. 72 3666–9
[13] Sykes A (The START Team, The NBI Team, The MAST Team and The Theory Team) 1999 The spherical tokamak programme at Culham Nucl. Fusion 39 1271–81
[14] Appel L C, Bevir M K and Walsh M J 2001 Equilibrium reconstruction in the START tokamak Nucl. Fusion 41 169–80
[15] Appel L C et al 2006 A unified approach to equilibrium reconstruction Proc. 33rd EPS Conf. on Controlled Fusion and Plasma Physics p.2–160
[16] Lee B, Pompfrey N and Lao L L 1999 Physics design of poloidal field, toroidal field and external magnetic diagnostics in KSTAR Fusion Technol. 36 278–88
[17] Park Y S et al 2011 KSTAR equilibrium operating space and projected stabilization at high normalized beta Nucl. Fusion 51 053001
[18] Okawa T et al (JT-60 Team) 2000 Heating and non-inductive current drive by negative ion based NBI in JT-60U Nucl. Fusion 40 435–43
[19] Sabbagh S A et al (NSTX Research Team) 2001 Equilibrium properties of spherical torus plasmas in NSTX Nucl. Fusion 41 11–11
[20] Li Y G, Lotte P, Zwillingmann W, Gi C and Imbeaux F 2011 EFIT equilibrium reconstruction including polariometry measurements on Tore Supra Fusion Sci. Technol. 59 397–405
[21] Li J et al 2003 Long pulse enhanced confinement discharges in the HT-7 superconducting tokamak by ion Bernstein wave heating and lower hybrid wave current drive Phys. Plasmas 10 1653–8
[22] Hongta H, Jinhua Z, Jiaqi D and Qiang Li 2006 Study of plasma MHD equilibrium in HL-2A tokamak Plasma Sci. Technol. 8 397–401
[23] Jinping Q et al 2009 Equilibrium reconstruction in EAST tokamak Plasma Sci. Technol. 11 142–5
[24] Li Q G, Ren Q L, Qian J P, Lao L L, Ding S Y, Chen Y J, Liu Z X, Lu B and Zang Q 2013 Kinetic equilibrium reconstruction on EAST tokamak Plasma Phys. Control. Fusion 55 125008
[25] Carlstrom T N et al 1992 Design and operation of the multipulse Thomson scattering diagnostic on DIII-D (invited) Rev. Sci. Instrum. 63 4901–6
[26] Carlstrom T N, Fooke J H, Nilson D G and Rice B W 1995 Design of the divertor Thomson scattering system on DIII-D Rev. Sci. Instrum. 66 493–5
[27] Christiansen J P, Callen J D, Ellis J J and Granetz R S 1989 Determination of current distribution in JET from soft x-ray measurements Nucl. Fusion 29 703
[28] Fischer R and Dinklage A 2004 Integrated data analysis of fusion diagnostics by means of the Bayesian probability theory Rev. Sci. Instrum. 75 4237–9
[29] Fischer R 2004 Bayesian experimental design—studies for fusion diagnostics AIP Conf. Proc. 735 76
[30] Williams C K and Rasmussen C E 2006 Gaussian Processes for Machine Learning (Cambridge, MA: MIT Press)
[31] Svensson J (European Fusion Development Agreement and JET Project) 2011 Non-Parametric Tomography Using Gaussian Processes (EFDA-JET) (Germany: EFDA)
[32] Li D et al 2016 Bayesian tomography and integrated data analysis in fusion diagnostics Rev. Sci. Instrum. 87 11E319
[33] Joung S, Kim J, Kwak S, Park K-R, Hahn S H, Han H S, Kim H S, Bak J G, Lee S G and Ghim Y-C 2018 Imputation of faulty magnetic sensors with coupled Bayesian and Gaussian processes to reconstruct the magnetic equilibrium in real time Rev. Sci. Instrum. 89 10K106
[34] Kwak S, Svensson J, Bozhenkov S, Flanagan J, Kempenaar M, Boboc A and Ghim Y-C (JET Contributors) 2020 Bayesian modelling of thomson scattering and multichannel interferometer diagnostics using Gaussian processes Nucl. Fusion 60 046009
[35] Sehyun Kwak J, Bozhenkov S, S, Trimino Mora H, Hoefel U, Pavone A, Krychowiak M, Langenberg A and Ghim Y C (W7-X Team) 2021 Bayesian modelling of multiple plasma diagnostics at Wendelstein 7-X (arXiv:2103.07582)
[36] Chilenski M A, Greenwald M, Marzouk Y, Howard N T, White A E, Rice J E and Walk J R 2015 Improved profile fitting and quantification of uncertainty in experimental measurements of impurity transport coefficients using Gaussian process regression Nucl. Fusion 55 023012
[37] Fischer R, Giannone L, Illerhaus J, McCarthy P J and McDermott R M (ASDEX Upgrade Team) 2020 Estimation and uncertainties of profiles and equilibria for fusion modeling codes Fusion Sci. Technol. 76 879–93
[38] Aaron H, Citrin J, Auriemma F, Bourdelle C, Casson F J, Kim H-T, Manas P, Szepesi G and Weisen H (JET Contributors) 2019 Application of Gaussian process regression to plasma turbulent transport model validation via integrated modelling Nucl. Fusion 59 056007
[39] Mathews A and Hughes J W 2021 Quantifying experimental edge plasma evolution via multidimensional adaptive Gaussian process regression IEEE Trans. Plasma Sci. 49 3841–7
[40] Howell E C and Hanson J D 2020 Development of a non-parametric Gaussian process model in the three-dimensional equilibrium reconstruction code V3FIT J. Plasma Phys. 86 905860102
[41] Sehyun Kwak J, Ford S, O, Appel L and Ghim Y C (JET Contributors) 2021 Bayesian equilibria of axisymmetric plasmas (arXiv:2103.07581)
[42] Snelson E, Ghahramani Z and Rasmussen C 2003 Warped Gaussian processes Advances in Neural Information Processing Systems vol 16, ed S Thrun, L Saul and B Schölkopf (MIT Press)
[43] Sciortino F et al 2020 Inference of experimental radial impurity transport on alcator C-MOD: Bayesian parameter estimation and model selection Nucl. Fusion 60 126014
[44] Xing Z A et al 2021 CAKE: consistent automatic kinetic equilibrium reconstruction Fusion Eng. Des. 163 112163
[45] Snoek J, Swersky K, Zemel R and Adams R 2014 Input warping for Bayesian optimization of non-stationary functions Proc. 31st Int. Conf. on Machine Learning (Proc. Machine Learning Research vol 32) ed E P Xing and T Jebara (Beijing: PMLR) pp 1674–82
[46] Cowen-Rivers A I et al 2020 HEBO pushing the limits of sample-efficient hyperparameter optimisation (https://doi.org/10.1613/jair.1.13643)
[47] Meneghini O et al 2015 Integrated modeling applications for tokamak experiments with OMFIT Nucl. Fusion 55 083008
[48] Logan N C, Grierson B A, Haskey S R, Smith S P, Meneghini O and Eldon D 2018 OMFIT tokamak profile data fitting and physics analysis Fusion Sci. Technol. 74 125–34
[49] This paper discusses the OMFIT profiles module that can also call out to the gptoo1.1a software package that forms the basis of the paper by Chilenski et al [36].
[50] Lachand V and Rasmussen C E 2020 Approximate inference for fully Bayesian Gaussian process regression Proc. 2nd Symp. on Advances in Approximate Bayesian Inference (Proc. Machine Learning Research vol 118) ed C Zhang, F Ruiz, T Bui, A B Dieng and D Liang (PMLR) pp 1–12
[51] Kruger S, Leddy J and Howell E 2022 Unified Bayesian Analysis Framework for Fusion Experimental Data (UNBAFFELD) (available at: https://gitlab.com/efit-ai/unbaffeld)
[52] Hutchinson I H 2002 Principles of Plasma Diagnostics
2nd edn (Cambridge: Cambridge University Press)

[53] Griffiths R-R, Aldrick A A, Garcia-Ortegon M, Lalchand V
and Lee A A 2021 Achieving robustness to aleatoric
uncertainty with heteroscedastic Bayesian optimisation
Mach. Learn.: Sci. Technol. 3 015004

[54] Kersting K, Plagemann C, Pfaff P and Burgard W 2007 Most
likely heteroscedastic Gaussian process regression Proc.
24th Int. Conf. on Machine Learning (ICML’07) (New
York: Association for Computing Machinery) pp 393–400

[55] Gibbs M N 1997 Bayesian Gaussian processes for regression
and classification PhD Thesis University of Cambridge

[56] Buckingham L, Hogan J M, Roe P, Sumitomo J and Towsley M
2008 Comparative studies made simple in GPFlow 2008
Eighth IEEE Int. Symp. on Cluster Computing and the Grid
(CCCGRID) (IEEE) p 699