J. Plasma Phys. (2020), *vol.* 86, 905860102 © The Author(s) 2020
This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted re-use, distribution, and reproduction in any medium, provided the original work is properly cited.
doi:10.1017/S0022377819000813

Development of a non-parametric Gaussian process model in the three-dimensional equilibrium reconstruction code V3FIT

Eric C. Howell[1,†] **and J. D. Hanson**[2]

[1]Tech-X Corporation, 5621 Arapahoe Ave Suite A, Boulder, CO 80303, USA

[2]Department of Physics, Auburn University, Leach Science Center, Auburn, AL 36832, USA

(Received 29 May 2019; revised 30 October 2019; accepted 31 October 2019)

A non-parametric Gaussian process regression model is developed in the three-dimensional equilibrium reconstruction code V3FIT. A Gaussian process is a normal distribution of functions that is uniquely defined by specifying a mean function and covariance kernel function. Gaussian process regression assumes that an unknown profile belongs to a particular Gaussian process and uses Bayesian analysis to select the function the give the best fit to measured data. The implementation in V3FIT uses a hybrid representation where Gaussian processes are used to infer some of the equilibrium profiles and standard parametric techniques are used to infer the remaining profiles. The implementation of the Gaussian process is tested using both synthetic data and experimental data from multiple machines.

Key words: fusion plasma, plasma confinement

1. Introduction

Equilibrium reconstruction is the process of inferring properties of the magnetohydrodynamic (MHD) equilibrium from experimental measurements. The specification of an equilibrium is the first step for many subsequent analyses, such as transport calculations or stability determination. An accurate experimental equilibrium is needed to compare the theoretical models used in these subsequent analyses with experimental results. Reconstruction is an essential tool in toroidal magnetic confinement fusion research.

During equilibrium reconstruction one must infer multiple radial profiles, that only depend on the flux-surface label, along with the appropriate boundary conditions. Two or more radial profiles are needed to specify an MHD equilibrium. For example, axisymmetric stationary equilibria are solutions of the Grad–Shafranov equation (Grad & Rubin 1958; Shafranov 1966), and they depend on the pressure profile, $p(\psi)$, and the toroidal magnetic field function, $F(\psi)$. Non-axisymmetric equilibria are often determined from their pressure profile and their toroidal current density profile

† Email address for correspondence: ehowell@txcorp.com

https://doi.org/10.1017/S0022377819000813 Published online by Cambridge University Press

(Hirshman & Whitson 1983). During the reconstruction process it is often necessary to introduce auxiliary profiles that relate MHD equilibrium quantities to measured diagnostic signals. A temperature (or density profile with an equation of state) is needed to relate temperatures measured from a Thomson diagnostic to the MHD equilibrium. Soft x-ray measurements, which can be used to infer the shape of fluxes, require an emissivity profile (Ma *et al.* 2018).

Many equilibrium reconstruction codes, such as EFIT (Lao & Ferron 1990) and V3FIT (Hanson *et al.* 2009), define the radial profiles using a parametric representation. The radial profiles are assumed to have a specified functional form characterized by multiple free parameters $\boldsymbol{p}$. The best fit equilibrium is defined by the set of parameters that minimize the error between the observed diagnostic signals, $\boldsymbol{S}^O$, and modelled diagnostic signals $\boldsymbol{S}^M(\boldsymbol{p})$. Often least-squares minimization is used, and the error between the observed signals and the modelled signals is defined as

$$\chi^2 \equiv \sum_i \frac{1}{2}\left(\frac{S_i^O - S_i^M(\boldsymbol{p})}{\sigma_{ni}}\right)^2. \tag{1.1}$$

Here σ_{ni} is the experimental error associated with the ith signal.

This paper discusses a non-parametric approach to reconstructing the radial profiles. Non-parametric approaches infer the profile shape directly from the measurements without specifying a particular functional form for the profile. The approach taken here uses Gaussian process regression (GPR) to infer the amplitude of the profile at a finite number of radial locations. The fundamental assumption of GPR is that the radial profile belongs to a Gaussian process (GP), which generalizes multivariate Gaussian distributions to a continuous domain (Rasmussen & Williams 2006). Bayesian analysis is then used to select the profile in the Gaussian process that gives the best fit.

Gaussian process regression has several advantages over standard parametric regression. The achievable accuracy of a parametric representation of experimental data is limited by the choice of a parameterization. In contrast, certain Gaussian processes contain a complete basis, and their accuracy is limited by measured data (Rasmussen & Williams 2006). The quality of fit of a GP will continually improve with additional measurements. Parametric models have to be specially designed to capture complex features like a pedestal or islands. The design of these complex parametric models can introduce systematic errors into the reconstruction process. With sufficient data a Gaussian process can naturally capture these complex features.

Many of the weaknesses of a parametric representation can be addressed by increasing the complexity of the model and thereby increasing the number of free parameters. However, increasing the number of parameters increases the computational cost of minimizing the error. Increasing the number of free parameters also increases the risk of overfitting. Gaussian process regression minimizes issues associated with overfitting by penalizing overly complex models. This feature of Bayesian inference is commonly known as Occam's razor.

Gaussian process regression has been used in several applications for fusion research. J. Svensson developed a framework for Gaussian process tomography (Svensson & Contributors 2010). This framework has been used to perform soft x-ray (SXR) tomographic analysis of W7-AS and TJ-II stellarator plasmas (Li *et al.* 2013). GPR has also been used to infer edge density profiles on the joint European torus (JET) (Kwak *et al.* 2017), and for uncertainty analysis in transport calculations (Chilenski *et al.* 2015).

https://doi.org/10.1017/S0022377819000813 Published online by Cambridge University Press

This paper is organized as follows. Section 2 provides an introduction to Gaussian process regression. The implementation of the Gaussian process model in V3FIT is discussed in § 3. Several example reconstructions that use both synthetic data and real experimental data are shown in § 4. Finally, a discussion of the results is presented in § 5.

2. Gaussian process regression

Gaussian processes generalize multivariate Gaussian probability distributions to an infinite domain. A GP is a collection of infinitely many random variables such that any finite collection of these random variables is normally distributed. A random function can be represented as an element of a Gaussian process, where the amplitude of the function at every point in the domain is treated as a normal random variable. The amplitudes of the function at any finite collection of points will have a joint multivariate Gaussian distribution.

Univariate normal distributions are uniquely defined by a mean value, μ, and a variance σ^2. Multivariate normal distributions are defined by a mean vector, $\boldsymbol{\mu}$, and a covariance matrix $\boldsymbol{\Sigma}$. Similarly, Gaussian processes are uniquely defined by a mean function, $\mu(x)$, and a covariance kernel $K(x, x')$.

The covariance kernel determines the properties of the Gaussian process. For example random Gaussian noise is generated using the delta function kernel,

$$K_G(x, x') = \sigma_f^2 \delta(x - x'). \tag{2.1}$$

The hyper-parameter σ_f is the standard deviation off the mean function. It characterizes how far the amplitude of the function can vary from the mean function, $\mu(x)$, at each point x. The delta function kernel imposes no correlation in the amplitudes of the function evaluated at two distinct points x_1 and x_2. The variance of the function from its mean value at x_1 is unrelated to the variance of the function at x_2. In the case where the mean function is zero, $\mu(x) = 0$, the delta function kernel implies that there is no correlation in the amplitude of the function across the domain.

The squared exponential kernel is another kernel that is frequently used in GPR,

$$K_{SE}(x, x') = \sigma_f^2 \exp\left(-\frac{(x - x')^2}{2\sigma_l^2}\right). \tag{2.2}$$

This kernel is characterized by two hyper-parameters σ_f and σ_l. Here σ_f is once again the standard deviation off the mean function. The hyper-parameter σ_l is a correlation length. It characterizes how strongly amplitude deviations of the function from the mean function are correlated across the domain. If $|x_1 - x_2| \ll \sigma_l$ then the variations from the mean function at these two points are strongly correlated. If the variation at x_1 is positive then the variation at x_2 will be positive by a similar amount. Conversely if $|x_1 - x_2| \gg \sigma_l$, then variations from the mean function at these two distant points are not correlated. In the case where the mean function is zero, $\mu(x) = 0$, the correlation length describes how rapidly the amplitude of the function varies across the domain. Thus this correlation length introduces a degree of smoothness into the Gaussian process. The Gaussian process defined by the squared exponential kernel has a basis that spans the space of smooth function (Rasmussen & Williams 2006).

Gaussian process regression can be understood using Bayesian statistics. Here we provide a qualitative sketch of the derivation. A detailed derivation can be found in the works of Rasmussen & Williams (2006) and Svensson & Contributors (2010).

https://doi.org/10.1017/S0022377819000813 Published online by Cambridge University Press

The derivation begins with Bayes' theorem,

$$p(f(x)|\boldsymbol{S}^O, \boldsymbol{I}) = \frac{p(\boldsymbol{S}^O|f(x), \boldsymbol{I})p(f(x)|\boldsymbol{I})}{p(\boldsymbol{S}^O|\boldsymbol{I})}. \tag{2.3}$$

The posterior distribution $p(f(x)|\boldsymbol{S}^O, \boldsymbol{I})$ is the probability that the experimental profile is the function $f(x)$ given the experimental measurements $\boldsymbol{S}^O$ and other prior knowledge $\boldsymbol{I}$. The likelihood $p(\boldsymbol{S}^O|f(x), \boldsymbol{I})$ is the probability of measuring the signals $\boldsymbol{S}^O$ for a given profile $f(x)$. The prior distribution $p(f(x)|\boldsymbol{I})$ characterizes the knowledge about the experimental profile before measuring the experimental signals. The evidence $p(\boldsymbol{S}^O|\boldsymbol{I})$ is a normalizing factor. Often the goal is to calculate the posterior distribution, but in practice the likelihood is much easier to calculate. Bayes' theorem provides a convenient way to calculate the posterior from the likelihood.

Gaussian process regression starts by assuming that the experimental profile is an element in a Gaussian process. This assumption is equivalent to saying that the prior distribution is a Gaussian process,

$$p(f(x)|\boldsymbol{I}) \equiv \mathcal{GP}(\mu(x), K(x, x')). \tag{2.4}$$

Here, $\mathcal{GP}$ represents a Gaussian process. The experimental measurements are assumed to have normally distributed noise with zero mean, characterized by the covariance $\boldsymbol{\Sigma}_n$. The likelihood is:

$$p(\boldsymbol{S}^O|f(x), \boldsymbol{I}) = \frac{\exp\left(-\frac{1}{2}(\boldsymbol{S}^O - \boldsymbol{L}f(x))^T \boldsymbol{\Sigma}_n^{-1}(\boldsymbol{S}^O - \boldsymbol{L}f(x))\right)}{\sqrt{(2\pi)^N|\boldsymbol{\Sigma}_n|}}, \tag{2.5}$$

where $\boldsymbol{\Sigma}_n$ is the noise correlation matrix, L^i is the mathematical operator that models the ith noise-free signal: $\boldsymbol{S}^M = \boldsymbol{L}f(x)$, with $\boldsymbol{L} = (L^1, L^2, \ldots, L^N)^T$ and N is the number of signals. In this paper a signal is called a (non)linear signal if the operator modelling that signal is a (non)linear operator. The posterior distribution is calculated from the prior and the likelihood using Bayes' theorem. The posterior distribution is then sampled at a finite number of points $\boldsymbol{x}_*$.

If the signals are linear, then the posterior distribution sampled at the points $\boldsymbol{x}_*$ is a multivariate normal distribution with a mean vector $\boldsymbol{\mu}_*$ and covariance matrix $\boldsymbol{\Sigma}_*$ (Svensson & Contributors 2010)

$$\boldsymbol{\mu}_* = \boldsymbol{K}_{*L}(\boldsymbol{K}_{LL} + \boldsymbol{\Sigma}_n)^{-1}\boldsymbol{S}^O, \tag{2.6}$$

$$\boldsymbol{\Sigma}_* = \boldsymbol{K}_{**} - \boldsymbol{K}_{*L}(\boldsymbol{K}_{LL} + \boldsymbol{\Sigma}_n)^{-1}\boldsymbol{K}_{L*}, \tag{2.7}$$

where it is assumed that the prior Gaussian process has a zero mean, $\mu(x) = 0$. The covariance matrices, $\boldsymbol{K}_{pq}$, are calculated from the Gaussian process kernel, and they characterize the covariance between the profile amplitude at different sample locations and the modelled signal values. The subscripts indicate whether the respective argument in the kernel is evaluated at the points $\boldsymbol{x}_*$ or operated on by the operators $\boldsymbol{L}$. The first (second) subscript corresponds to the first (second) argument in the kernel. The (ith, jth) elements of the $\boldsymbol{K}_{pq}$ matrices are

$$K_{**}^{ij} = K(x_*^i, x_*^j) \tag{2.8}$$

$$K_{LL}^{ij} = L^i L'^j K(x, x') \tag{2.9}$$

$$K_{L*}^{ij} = L^i K(x, x_*^j) = K_{*L}^{ji}. \tag{2.10}$$

The last equality uses the fact that $\boldsymbol{K}_{*L} = \boldsymbol{K}_{L*}^T$.

https://doi.org/10.1017/S0022377819000813 Published online by Cambridge University Press

For example consider the case where the value of a function $f(x)$ is measured at two points l_1 and l_2. The operator representing these measurements is a delta function

$$L^i f(x) = \int \delta(x - l_i) f(x)\, \mathrm{d}x = f(l_i) \quad \text{for } i = 1, 2. \tag{2.11}$$

Similarly the goal is to sample the function at three points: s_1, s_2, s_3. In this case the three $\boldsymbol{K}$ matrices are

$$\boldsymbol{K}_{**} = \begin{pmatrix} K(s_1, s_1) & K(s_1, s_2) & K(s_1, s_3) \\ K(s_2, s_1) & K(s_2, s_2) & K(s_2, s_3) \\ K(s_3, s_1) & K(s_3, s_2) & K(s_3, s_3) \end{pmatrix} \tag{2.12}$$

$$\boldsymbol{K}_{LL} = \begin{pmatrix} K(l_1, l_1) & K(l_1, l_2) \\ K(l_2, l_1) & K(l_2, l_2) \end{pmatrix} \tag{2.13}$$

$$\boldsymbol{K}_{L*} = \begin{pmatrix} K(l_1, s_1) & K(l_1, s_2) & K(l_1, s_3) \\ K(l_2, s_1) & K(l_2, s_2) & K(l_2, s_3) \end{pmatrix}. \tag{2.14}$$

The matrix $\boldsymbol{K}_{**}$ characterizes the correlation between the amplitudes of the function evaluated at each of the points in $\boldsymbol{x}_*$. The matrix $\boldsymbol{K}_{LL}$ characterizes the correlation between each of the noise-free modelled signals. The matrix $\boldsymbol{K}_{L*}$ characterizes the correlation between the amplitude of the function evaluated at each point in $\boldsymbol{x}_*$ and each noise-free model signal.

The posterior distribution is a multivariate Gaussian distribution, and it is completely specified by solving (2.6) and (2.7) for $\boldsymbol{\mu}_*$ and $\boldsymbol{\Sigma}_*$. These equations are linear, and they involve the inverse of the matrix $(\boldsymbol{K}_{LL} + \boldsymbol{\Sigma}_n)$. This matrix is a symmetric positive definite matrix, and it can be factored using a Cholesky decomposition.

The kernels used to define a GP often have one or more hyper-parameters. An optional step in GPR is to then find the optimal set of hyper-parameters. This is done by using Bayes' theorem to define a hyper-posterior distribution for the hyper-parameters $\boldsymbol{\sigma}_{hp}$,

$$p(\boldsymbol{\sigma}_{hp}|\boldsymbol{S}^O) \propto p(\boldsymbol{S}^O|\boldsymbol{\sigma}_{hp}) p(\boldsymbol{\sigma}_{hp}), \tag{2.15}$$

and then finding the set of hyper-parameters that maximize this hyper-posterior. A common choice is to use a uniform hyper-prior $p(\boldsymbol{\sigma}_{hp})$. In this case the hyper-posterior distribution is,

$$p(\boldsymbol{\sigma}_{hp}|\boldsymbol{S}^O) \propto p(\boldsymbol{S}^O|\boldsymbol{\sigma}_{hp}). \tag{2.16}$$

Here, the hyper-posterior is proportional to the Bayesian evidence in (2.3).

The evidence is calculated by marginalizing the posterior distribution for the data over all possible functions,

$$p(\boldsymbol{S}^O|\boldsymbol{\sigma}_{hp}) = \int \mathrm{d}f p(\boldsymbol{S}^O|f, \boldsymbol{\sigma}_{hp}) p(f|\boldsymbol{\sigma}_{hp}). \tag{2.17}$$

The result of the integration yields the expression for log of the evidence

$$\ln p(\boldsymbol{S}^O|\boldsymbol{\sigma}_{hp}) = -\tfrac{1}{2}[N \ln 2\pi + \ln|\boldsymbol{K}_{LL} + \boldsymbol{\Sigma}_n| + (\boldsymbol{S}^O)^T (\boldsymbol{K}_{LL} + \boldsymbol{\Sigma}_n)^{-1} \boldsymbol{S}^O]. \tag{2.18}$$

Maximizing the evidence is equivalent to minimizing the bracketed quantity in the equation for the log of the evidence. The bracketed quantity is composed of three

https://doi.org/10.1017/S0022377819000813 Published online by Cambridge University Press

terms. The first term is a positive normalizing constant that accounts for the number of signals, N, in the Gaussian process. The second term characterizes the complexity of the model, and it can be either positive or negative. The third term characterizes the quality of the fit, and it is positive definite. The competition between the second and third terms prevents overfitting by penalizing overly complex models. Increasing the complexity of the model will generally improve the quality of fit, i.e. the third term will be smaller. However, increasing the complexity also increases the second term. Thus an increase in complexity is only advantageous if the resulting improvement in the fit outweighs the cost of the complexity.

3. Implementation of GPR model in V3FIT

V3FIT is a three-dimensional equilibrium reconstruction code (Hanson *et al.* 2009). The code is designed to be fast, with the goal of being able to run reconstructions in between experimental discharges. The code is also designed to be modular and extensible. This enables rapid development of new functionality needed to meet the requirements of different experiments. VMEC (Hirshman & Whitson 1983) is the primary equilibrium solver used by V3FIT; however, the code modularity allows V3FIT to use other equilibrium solvers with minor code modifications. Recent modifications to V3FIT allow it to use the SIESTA three-dimensional (3-D) equilibrium code (Cianciosa *et al.* 2018).

V3FIT finds the parameters $\boldsymbol{p}$ that minimize the cost function

$$g^2(\boldsymbol{p}) \equiv \sum_i \omega_i \left(\frac{S_i^O - S_i^M(\boldsymbol{p})}{\sigma_{ni}} \right)^2 . \tag{3.1}$$

The cost function g^2, which is closely related to the χ^2 error, measures the difference between the experimentally determined signals, $\boldsymbol{S}^O$, and the modelled signals, $\boldsymbol{S}^M$, for an equilibrium defined by the parameters $\boldsymbol{p}$. Note that the form for g^2 is consistent with an assumption that the signal noise is uncorrelated – the signal covariance matrix is diagonal. The σ_{ni} are the square roots of the signal variances. Weighting factors, ω_i, allow one to emphasize or deemphasize selected signals. The cost function g^2 is a positive definite quantity, and V3FIT uses a modified quasi-newton algorithm to find the equilibrium parameters that minimize g^2.

The implementation of the GPR model in V3FIT is designed to work in conjunction with the standard parametric representation. The guiding philosophy is to use GPR to infer profiles where it is easy and efficient to do so. A parametric representation is then used to represent the remaining profiles. For speed and efficiency the implementation of the GPR model is currently limited to use with linear diagnostics. These diagnostics can be modelled by a linear operator acting on the radial profile. Some examples of profiles that can be modelled with our GPR model include the emissivity inferred from soft x-ray data, the temperature inferred from Thomson or ECE measurements, and a density profile inferred from interferometry. The use of soft x-ray data to infer the temperature profile is an example of a nonlinear diagnostic where our GPR model is not applicable. Here the soft x-ray emissivity is nonlinearly related to the temperature.

Let us look in more detail at the observed soft x-ray signals and their relation to the soft x-ray emissivity. The VMEC code labels flux surfaces with a minor-radius variable s with $0 \leqslant s \leqslant 1$. The soft x-ray emissivity ϵ depends on the position within

https://doi.org/10.1017/S0022377819000813 Published online by Cambridge University Press

the plasma, $\boldsymbol{r}$. The αth soft x-ray signal is viewed along the chord $\boldsymbol{r}_\alpha(t) = \boldsymbol{r}_{\alpha i} + t(\boldsymbol{r}_{\alpha f} - \boldsymbol{r}_{\alpha i})$ between the initial position $\boldsymbol{r}_{\alpha i}$ and the final position $\boldsymbol{r}_{\alpha f}$. The observed signal is modelled as

$$\boldsymbol{S}_\alpha^O \propto \int \mathrm{d}t \epsilon(\boldsymbol{r}_\alpha(t)). \tag{3.2}$$

The soft x-ray emissivity is assumed to be constant on a flux surface, described by the profile function $\epsilon(s)$. Any position in the plasma can be mapped to a particular flux surface through the function $s(\boldsymbol{r})$. Thus the observed signal can be written as a linear function of the emissivity profile,

$$\boldsymbol{S}_\alpha^O \propto \int \mathrm{d}t \epsilon(s(\boldsymbol{r}_\alpha(t))). \tag{3.3}$$

The covariance matrix element characterizing the correlation between two x-ray signals is

$$K_{LL}^{\alpha\beta} \propto \int\int \mathrm{d}t\, \mathrm{d}t' K(s(\boldsymbol{r}_\alpha(t)), s(\boldsymbol{r}_\beta(t'))). \tag{3.4}$$

The restriction to linear diagnostics allows for the direct calculation of the posterior mean and covariance using (2.6) and (2.7). Here the posterior distribution is also a Gaussian process, and thus it is completely defined by the mean and covariance. The posterior mean profile is the best fit to the experimental data, and the covariance matrix characterizes the uncertainty in the fit.

A new term is added to V3FIT's cost function that allows V3FIT to optimize the hyper-parameters,

$$g^2 \equiv \sum_{i=1}^{N} \omega_i \left(\frac{S_i^O - S_i^M}{\sigma_{ni}} \right)^2 + \sum_{j=1}^{T} \omega_{N+j} \Theta_j^2. \tag{3.5}$$

Here N is the number of experimental signals, T is the number of radial profiles modelled by independent Gaussian processes and Θ_j^2 is related to the negative log evidence for the jth Gaussian process,

$$\Theta_j^2 = \tfrac{1}{2}[N \ln 2\pi + \ln |\boldsymbol{K}_{LL} + \boldsymbol{\Sigma}_n| + (\boldsymbol{S}^O)^T (\boldsymbol{K}_{LL} + \boldsymbol{\Sigma}_n)^{-1} \boldsymbol{S}^O]_j + c_j. \tag{3.6}$$

The nonlinear optimization routine in V3FIT assumes that the cost function is a quadratic form. The constant c_j is added to ensure that Θ_j^2 is positive for each Gaussian process, and thus Θ_j is real.

The added term to the cost function treats each Gaussian process and its hyper-parameters in a way that is similar to the experimentally measured signals. This treatment allows for the simultaneous optimization of both the model parameters and the Gaussian process hyper-parameters. The dependence of the cost function on the hyper-parameters enters through the $\boldsymbol{K}_{LL}$ matrix.

Figure 1 shows an outline of the V3FIT algorithm. First an initial guess is made for each of the model parameters and Gaussian process hyper-parameters. The model parameters are then used to solve for an initial magnetic equilibrium. Then GPR is used to calculate GP modelled radial profiles. After that each of the model signals are calculated for each diagnostic, and then g^2 is calculated. V3FIT then checks to see if

https://doi.org/10.1017/S0022377819000813 Published online by Cambridge University Press

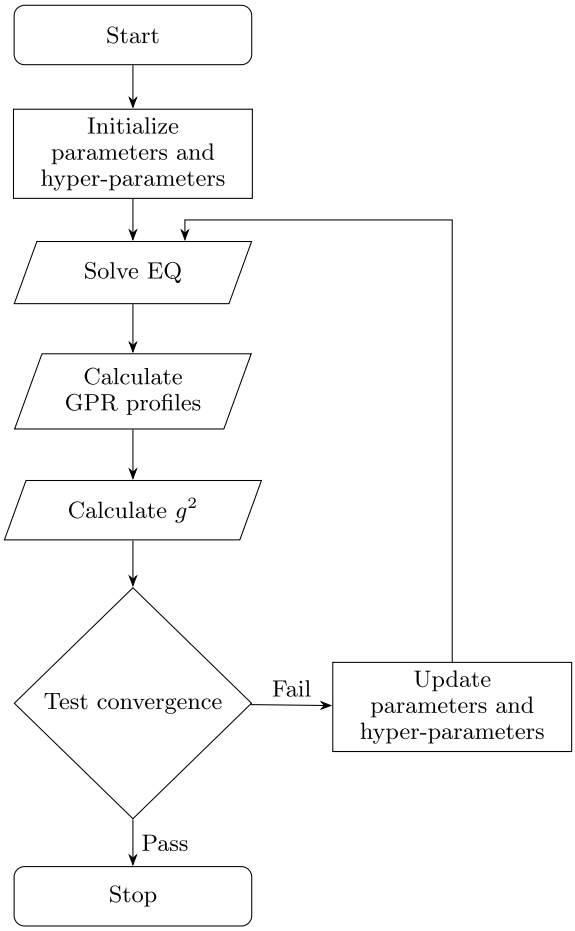

FIGURE 1. Flow chart of the modified algorithm.

an optimal set of parameters and hyper-parameters have been found. If a converged solution has been found then V3FIT exits. Otherwise V3FIT calculates a new set of parameters and repeats the process. In order to update the parameters V3FIT numerically calculates a parameter Jacobi matrix. During this calculation V3FIT has to vary each of the parameters and hyper-parameters. The variations of the parameters often require that the equilibrium be resolved and/or the GPR profiles be recalculated.

4. Testing

This section presents three example reconstructions to illustrate and test the performance of the GPR model. The first test case is a synthetic equilibrium that is based on the Compact Toroidal Hybrid experiment (the experiment is explained below). The case uses a realistic set of diagnostics; however, a fictitious Thomson scattering diagnostic. The Thomson scattering diagnostic is treated as a localized measurement of the electron temperature profile. First, the synthetic equilibrium is used to illustrate the behaviour of the Gaussian process reconstruction as the hyper-parameters are varied. Here, the use of the localized measurements is instructive

https://doi.org/10.1017/S0022377819000813 Published online by Cambridge University Press

and helps illustrate the GPR behaviour. Second, a full reconstruction of synthetic data is performed to test the performance of the GPR reconstruction model.

The second test case uses real data from Compact Toroidal Hybrid experiment. This reconstruction does not use the fictitious Thomson system, but instead uses the experiments two colour soft x-ray system. GPR is used to reconstruct two emissivity profiles, one for each soft x-ray colour. In addition this case uses fixed hyper-parameters to test the performance of model when reasonable but non-optimal hyper-parameters are used.

The final test case uses real data from the Madison symmetric torus (described below). Here GPR is used to reconstruct the temperature profile from an experimental Thomson diagnostic. In this case the optimal set of hyper-parameters are calculated as part of the reconstruction.

Combined, these examples serve as a comprehensive test the GPR model. The model is tested on both synthetic data and experimental data. The model is tested on multiple machines uses different diagnostics. Finally the model is tested with both fixed hyper-parameters and when the hyper-parameters are calculated as part of the reconstruction.

4.1. *Synthetic data*

The implementation of the Gaussian process model is tested using a synthetic equilibrium based on the Compact Toroidal Hybrid experimental (CTH) (Hartwell *et al.* 2017). CTH is a low-aspect-ratio five field period torsatron designed to study current carrying stellarator equilibria. A flexible set of external magnetic coils allows for the creation of a diverse set of vacuum magnetic equilibria with rotational transforms ranging from $\iota_{\mathrm{vac}}/2\pi \approx 0.02$ to $\iota_{\mathrm{vac}}/2\pi \approx 0.35$. The maximum on-axis toroidal magnetic field is $B_0 = 0.7$ T. An ohmic transformer is used to drive up to 80 kA of toroidal plasma current.

The synthetic equilibrium uses a two-power parameterization for toroidal current profile

$$I'(s) = \frac{\mathrm{d}I}{\mathrm{d}s} = I_0(1 - s^{\alpha_I})^{\beta_I}. \tag{4.1}$$

Here s is the flux-surface label, where $s = 0$ labels the magnetic axis and $s = 1$ labels the last closed flux surface; $I(s)$ is the net toroidal current enclosed by the flux surface s, the exponents α_I and β_I are positive real numbers that control the shape of the current profile. For a fixed β_I a larger value of α_I leads to a broader profile. A smaller value of α_I leads to more peaked profile. The net toroidal current $I(1)$ is calculated by integrating $I'(s)$ from $s = 0$ to $s = 1$.

The pressure profile $p(s)$ and the electron temperature profile $T_e(s)$ are also specified using two-power profiles

$$p(s) = p_0(1 - s^{\alpha_p})^{\beta_p} \tag{4.2}$$

$$T_e(s) = T_{e0}(1 - s^{\alpha_T})^{\beta_T}, \tag{4.3}$$

where p_0 (T_{e0}) is the pressure (electron temperature) at the magnetic axis. The density n is inferred from the pressure and electron temperature assuming that the electron and ion temperatures are equal: $p = 2nK_BT_e$ with the Boltzmann constant K_B. The values of the parameters that define the synthetic equilibrium are summarized in table 1. The

https://doi.org/10.1017/S0022377819000813 Published online by Cambridge University Press

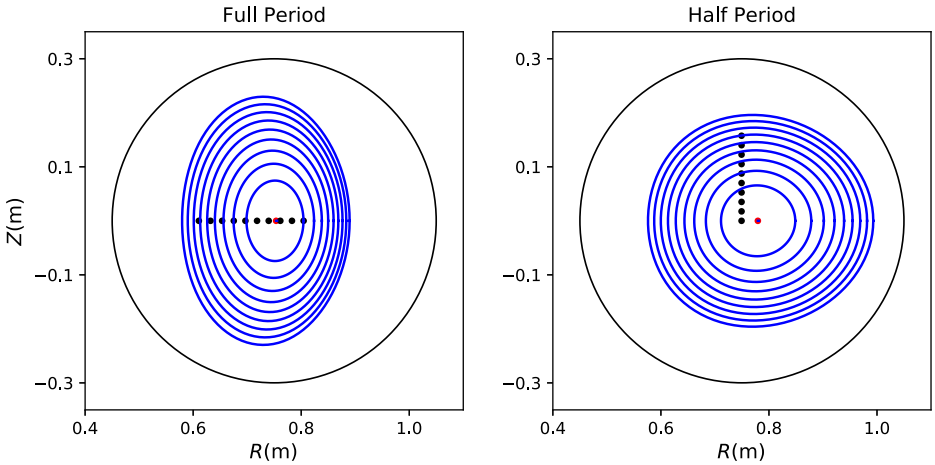

FIGURE 2. The magnetic flux surfaces are plotted for the synthetic equilibrium at the full period and half-period. The location of the magnetic axis is shown in red. The sampling locations of the twenty Thomson scattering channels are indicated by the black dots.

Net toroidal current	$I(1)$	62.32 kA
First current exponent	α_I	3.56
Second current exponent	β_I	6.0
Pressure on axis	p_0	488 Pa
First pressure exponent	α_P	1.25
Second pressure exponent	β_P	4.0
Temperature on axis	T_{e0}	250 eV
First temperature exponent	α_T	2.2
Second temperature exponent	β_T	4.3

TABLE 1. The profile parameters that are used to define the synthetic equilibrium.

magnetic flux surfaces of the synthetic equilibrium are shown at the full period and half-period in figure 2.

The reconstructions use a realistic set of diagnostics that are based on CTH's diagnostics. CTH's full set of magnetic diagnostics are used. CTH's three-channel interferometer is used to constrain the pressure profile via the inferred density. Two fictitious 10-channel Thomson scattering diagnostics, one located at the full period and the other located at the half-period, are used to infer the electron temperature. This Thomson scattering system is used to illustrate the behaviour of the Gaussian processes. The sampling locations of the Thomson scattering system are indicated by the bullets (●) in figure 2.

Synthetic data are calculated for each diagnostic used in the reconstruction. First the noise-free modelled signal is calculated for each synthetic diagnostic. Noisy signals are then generated using these modelled signals assuming 5 % Gaussian noise. The noisy signals are then used as input to a V3FIT reconstruction. The accuracy of the reconstruction is tested by comparing the reconstructed parameter values with their prescribed equilibrium values.

https://doi.org/10.1017/S0022377819000813 Published online by Cambridge University Press

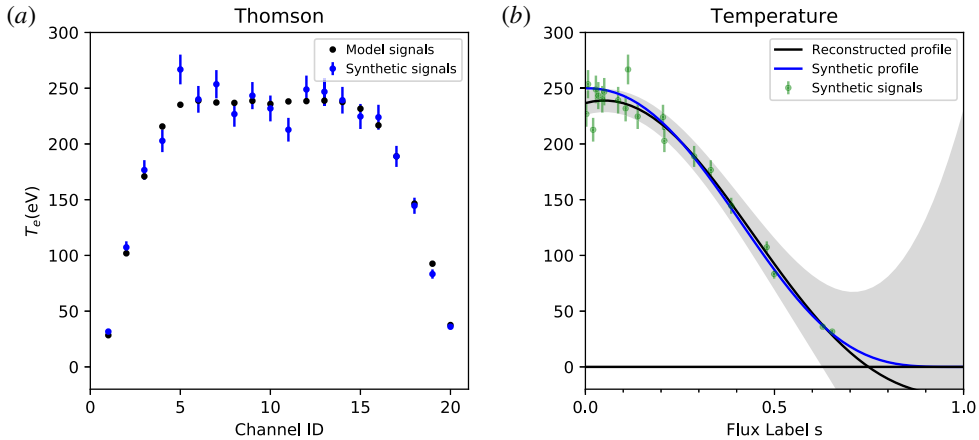

FIGURE 3. The electron temperature measured by the Thomson system is shown in (*a*). The synthetic signals (blue) are the noisy signals generated from the synthetic temperature profile. The error bars indicate 5 % uncertainty in the synthetic measured signals. The model signals (black) are calculated from the reconstructed profile. The reconstructed temperature profile is shown in figure (*b*). The shaded region represents the 2σ standard deviation. The synthetic electron temperature profile is shown in blue. The green markers indicate the location of Thomson sampling points and the measured values of the temperature. The reconstructed GP profile is plotted as a straight line connecting the value of profile evaluated at the points $\boldsymbol{x}_*$.

First, to illustrate the behaviour of GPR, we start by only reconstructing the electron temperature profile. Here the equilibrium current and pressure profiles are specified to have their true equilibrium values, and the electron temperature profile is reconstructed using GPR. V3FIT's minimization routines are used to converge on the optimal set of hyper-parameters by minimizing the negative log evidence.

The reconstructed temperature at the optimized set of hyper-parameters is shown in figure 3. Figure 3(*a*) shows the electron temperature at each of the Thomson scattering sampling locations. The noisy synthetic signals, which are used in the reconstruction, are shown in blue. The error bars indicate 5 % uncertainty in the synthetic measured signals. The modelled signals, shown in black, are calculated from the reconstructed temperature profile. Most of the modelled signals agree with their corresponding synthetic measured signals within one or two standard deviations.

Figure 3(*b*) shows a comparison of the reconstructed temperature profile (black) and the prescribed temperature profile (blue). The 2σ uncertainty region for the reconstructed temperature profile is indicated by the shaded region. The noisy signals used in the reconstruction, along with their corresponding errors bars, are indicated by the green markers. The reconstructed profile agrees with the synthetic profile within 2σ across the entire domain.

The reconstructed temperature profile begins to diverge from the synthetic profile for $s > 0.6$, and the reconstructed profile is negative for $s \gtrsim 0.75$. This discrepancy between the reconstructed temperature profile and the synthetic profile is understood by observing that there are no measurements of the temperature for $s \gtrsim 0.6$. In this region the slope of the synthetic temperature profile drastically changes and the temperature asymptotes to zero at $s = 1$. There are no measurements that capture this transition, and the Gaussian process uses a smooth gradient scale length to extrapolate

https://doi.org/10.1017/S0022377819000813 Published online by Cambridge University Press

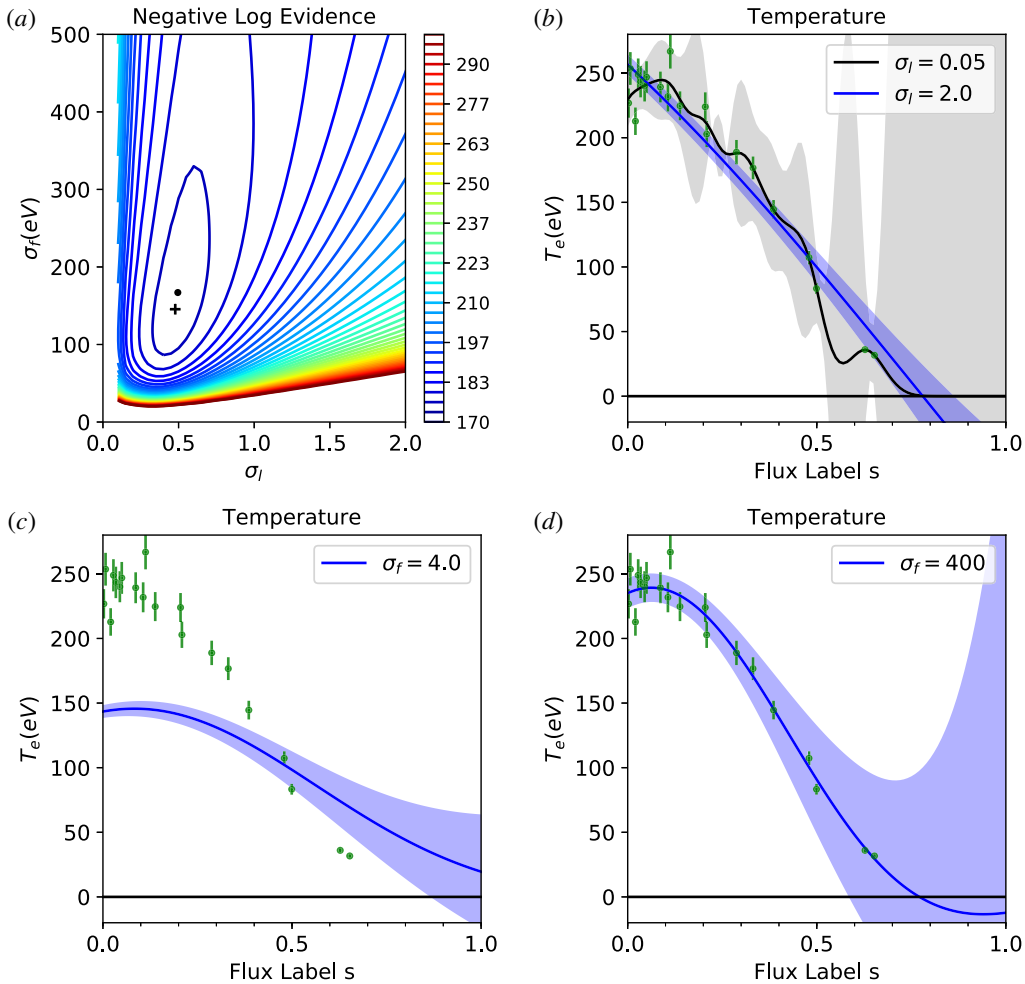

FIGURE 4. (*a*) The negative log evidence is shown as a function of the hyper-parameters σ_l and σ_f. The true minimum is indicated by the (+) where the converged set of hyper-parameters used by V3FIT is indicated by the (●). (*b*) The dependence of the GP best fit as a function of the correlation length scale σ_l is illustrated. A small correlation length scale $\sigma_l = 0.05$ is shown in black, and a longer correlation length scale $\sigma_l = 2.0$ is shown in blue. Both fits use the optimal value $\sigma_f = 145.5$. The shaded regions indicate the 2σ uncertainty for each fit, and the green dots indicate the training data. The dependence of the GP best fit as function of the hyper-parameter σ_f is illustrated in (*c*,*d*). An unrealistically small $\sigma_f = 4$ is used in (*c*) and an unrealistically large $\sigma_f = 400$ is used in (*d*). Both plots use the optimal correlation length scale $\sigma_l = 0.478$.

the temperature profile beyond the last measurement. The Thomson diagnostic in this case has equally spaced viewing locations in physical space. This spacing leads to clustering of the measured data at small s in flux coordinate space.

The reconstructed temperature profile shown in figure 3 is calculated using a optimized set of hyper-parameters: $\sigma_f = 167$, $\sigma_l = 0.495$. Figure 4(*a*) shows contours of the negative log evidence as a function of the hyper-parameters. The contours show a well-defined minimum around $\sigma_f = 146 \pm 5$, $\sigma_l = 0.48 \pm 0.05$ indicated by (+). The

https://doi.org/10.1017/S0022377819000813 Published online by Cambridge University Press

V3FIT optimized values of the hyper-parameters are indicated by (●). The negative log evidence has a shallow minimum and the difference between the two sets of hyper-parameters is due to V3FIT's convergence criteria. More stringent convergence criteria could be used to more accurately converge on the minimum; however, in practice this is not necessary. As illustrated in figure 3 a reasonable value for the hyper-parameters usually results in a good fit.

As illustrated in figure 4(*a*) the negative log evidence rapidly increases at small σ_f and small σ_l. The maximum contour level in the figure has been set to 300 for clarity, but the negative log evidence far exceeds this limit at small σ_f and small σ_l. At small σ_f the increase in the negative log evidence results from a bad fit. Here, the GP profile is too tightly constrained to be zero. At small σ_l the Gaussian process is strongly penalized for having too small of a correlation length scale. The small correlation length scale causes overfitting, and at this point the GP is essentially fitting the noise in the data. Conversely the negative log evidence gradually increases at large σ_f and σ_l. The increase at large σ_f results from penalizing the GP for having too large of a variance. At large σ_l the correlation length is too large, resulting in a bad fit to the data.

Figure 4(*b–d*) further illustrates the dependence of reconstructed temperature profile on the hyper-parameters. The reconstructed temperature profile is shown in figure 4(*b*) for two values of the correlation length σ_l. Over-fitting is observed when a small correlation length is used (black). Here the Gaussian process has the flexibility to fit to the noise in the data (shown in green). The uncertainty in the fit, indicated by the grey shaded region, is large between the measured data. When a large correlation length is used (blue) the GP does not have the flexibility to conform to the profile. Here the Gaussian process is effectively fitting a straight line to the data. Both fits in figure 4(*b*) use the optimized value $\sigma_f = 145.5$.

The hyper-parameter σ_f quantifies the prior standard deviation from zero in the prior Gaussian process. This is illustrated in figures 4(*c*) and 4(*d*), which show the GP fit for two values of σ_f at the optimized correlation length $\sigma_l = 0.48$. Figure 4(*c*) uses the value $\sigma_f = 4.0$ to illustrates the behaviour when σ_f is too small. Here the GP is tightly constrained to be close to zero, and the resulting fit under-predicts the synthetic temperature. Figure 4(*d*) uses $\sigma_f = 400$ to illustrate what happens when σ_f is too large. Here, the GP still produces a good fit; however, the uncertainty in the fit is larger than the uncertainty at the optimized value of σ_f. This behaviour is most easily observed by comparing the shaded regions in figures 3(*b*) and 4(*d*) at $s = 1$. At the optimized set of hyper-parameters the 2σ uncertainty region at $s = 1$ is bounded by $T_e = 100$ eV in contrast to the case of $\sigma_f = 400$ the uncertainty region is bounded by $T_e \approx 175$ eV.

Figure 4(*b–d*) is designed to illustrate the physical significance of the hyper-parameters. Extreme values of the hyper-parameters were chosen for illustrative purposes, and a reasonable choice of the hyper-parameters will often yield a good fit. As a general the optimal value of σ_f tends to be comparable to the maximum amplitude of the profile. This is a consequence of Occam's razor, which in this context states that the deviation from the zero mean should be as small as possible while large enough to account for all the measured data. The optimal value of σ_l is a little harder to predict. We have found that $\sigma_l \approx 0.5$ is usually a reasonable guess for smooth profiles that vary across the entire domain. It is worth noting that non-stationary kernels, where the correlation length can vary across the domain, have been useful for capturing sharp transitions in the gradient scale length that happens in many fusion plasma (Li *et al.* 2013; Chilenski *et al.* 2015).

Second, a full reconstruction of the synthetic equilibrium is considered. Here, standard parametric techniques are used to reconstruct all the equilibrium parameters

https://doi.org/10.1017/S0022377819000813 Published online by Cambridge University Press

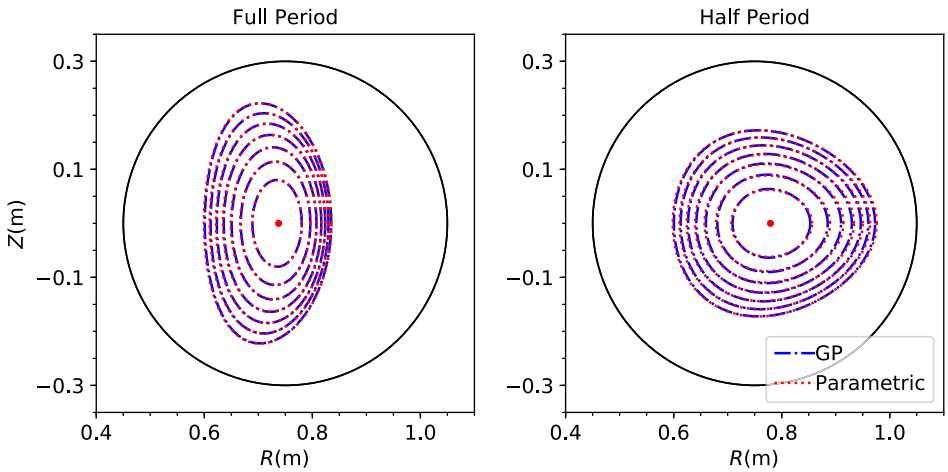

FIGURE 5. The reconstructed flux surfaces are shown for the experimental CTH reconstruction at the full period and the half-period. The flux surfaces from the Gaussian process reconstruction are shown in blue. The flux surfaces from the fully parametric reconstruction are shown in red.

Parameter name	Equilibrium value	Initial guess	Reconstructed value
$I(1)$	62.32 kA	50.0 kA	60.57 ± 3.78 kA
α_I	3.56	2.0	2.79 ± 0.40
p_0	488 Pa	200 Pa	417 ± 178 Pa
Φ_{edge}	59.3 mWb	65.0 mWb	60.3 ± 1.3 mWb
σ_f	NA	200.0	156.8 ± 10.3
σ_l	NA	0.50	0.519 ± 0.035

TABLE 2. The equilibrium parameter values and their reconstructed values are shown. Here Φ_{edge} is the toroidal flux at the last closed flux surface.

except for the electron temperature profile. GPR regression is used to infer the temperature profile. The cost function g^2 is minimized to simultaneously calculate the optimal set of parameters and GP hyper-parameters.

Table 2 shows the true value, the initial guess used to seed the reconstruction, and the reconstructed value for each reconstructed equilibrium parameter.

The largest discrepancy between the reconstructed parameters and their equilibrium values is in the current shaping factor α_I. Here, the disagreement between the reconstructed α_I and the equilibrium value is two standard deviations. The difficulty in reconstructing this parameter is due to the nonlinear dependence of the flux-surface shape on the current shaping factor. Modest changes in α_I have only a small impact on the shape of the flux surfaces. This is apparent in figure 5, which is discussed in the next section. The other reconstructed parameters agree with their equilibrium values within one standard deviation, and overall the reconstructed equilibrium agrees reasonably well with the synthetic equilibrium.

https://doi.org/10.1017/S0022377819000813 Published online by Cambridge University Press

Parameter name	Symbol	GP value	Parametric value
Edge flux	Φ_{edge}	52.91 ± 0.40 mWb	52.28 ± 5.38 mWb
Toroidal current	I_p	38.93 ± 0.24 kA	38.84 ± 0.31 kA
Current shaping factor	α_I	2.09 ± 0.37	1.52 ± 0.16
Vertical offset	Z_0	5.70 ± 0.66 mm	7.23 ± 0.76 mm
Peak pressure	p_0	389 ± 242 pa	364 ± 3700 pa
Pressure shaping factor	α_p	1.82 ± 0.08	$0.00 + 180$

TABLE 3. The values of the reconstructed parameters are shown for the hybrid GP reconstruction and the fully parameter construction. The fully parametric reconstruction uses an additional 20 parameters to model the soft x-ray emissivity profiles in addition to the parameters shown.

4.2. *CTH experimental discharge*

An experimental CTH discharge is used as a second benchmark of the Gaussian process model. In this test case Gaussian processes are used to represent two soft x-ray emissivity profiles. CTH has a three camera two-colour soft x-ray system (Herfindal *et al.* 2014) and a different Gaussian process is used to infer each colour's emissivity profile. Each emissivity profile is approximated as a flux-surface quantity. The soft x-ray system can be used to topographically infer the shape of internal flux surfaces. The shape of these flux surfaces is determined by the internal current profile. Thus, the soft x-ray measurements help to constrain the internal current profile in equilibrium reconstructions (Ma *et al.* 2015, 2018).

Six parameters are inferred in this reconstruction: the edge toroidal flux Φ_{edge}, the net toroidal current I_p, a toroidal current shaping factor α_I, the pressure at the magnetic axis p_0, a pressure profile shaping factor α_p and vertical offset from the mid-plane of the magnetic axis z_0. In this reconstruction a two-power profile is used for the current profile (4.1) and the pressure profile (4.2). The second exponent in the current profile, β_i, and the second exponent in the pressure profile β_p are both assumed to be six. This case uses fixed values for the Gaussian process hyper-parameters. The first colour's kernel has a variance set to $\sigma_{f1} = 2$ and a correlation length scale set to $\sigma_{l1} = 0.4$. The second colour's kernel has a variance set to $\sigma_{f1} = 0.3$ and a correlation length scale set $\sigma_{l1} = 0.4$. These are typical of the hyper-parameter values found for the soft x-ray system when hyper-parameter optimization is used. We refer to this reconstruction as a hybrid reconstruction since some radial profile are represented using a parametric representation and others are represented using a Gaussian process.

A fully parametric reconstruction is also performed to compare with the Gaussian process model. The reconstruction uses two ten-segment linear splines to represent each of the emissivity profiles. The knots of the splines are equally spaced in s, and the amplitudes of the knots are treated as free parameters. In total this model has 26 free parameters.

Table 3 shows the reconstructed values of the six equilibrium parameters for the hybrid Gaussian process reconstruction and the parametric reconstruction. The values of equilibrium parameters for the two reconstruction agree within 2σ error. The soft x-ray emissivity is needed to constrain the current shaping factor α_I, and the fact that the two parameters are close indicates that the method is successfully using the GP process profiles to constrain the equilibrium profiles. The large uncertainty in the

https://doi.org/10.1017/S0022377819000813 Published online by Cambridge University Press

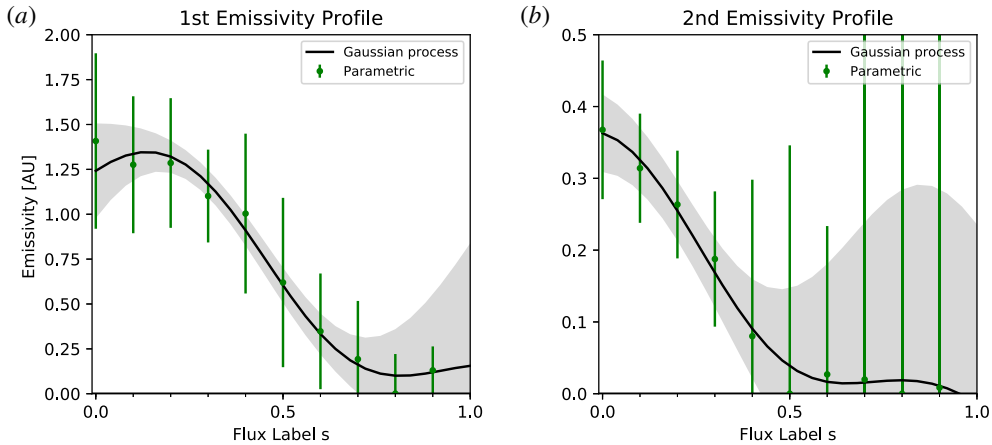

FIGURE 6. The reconstructed soft x-ray emissivity profiles are shown for each of the two colour filters. The black line is the emissivity profile reconstructed using the Gaussian process model and the shaded region indicates the 2σ uncertainty region. The profiles reconstructed using the fully parameter profile are shown in green. Here only the amplitudes of the profile at the line spline knots are shown for clarity. The error bars also indicate the 2σ uncertainty.

peak pressure and the pressure profile shaping factor is due to the fact that these reconstructions do not use any direct measurements of the temperature, density or pressure. While the uncertainty in these parameters is large, including these pressure parameters in the reconstruction gives the χ^2-minimization routine extra degrees of freedom which aids converge.

Figure 5 shows reconstructed magnetic flux surfaces produced by the two methods at the full period and half-period. The observation that the flux surfaces nearly lie on top of each other is another indication that the two reconstructions are in good agreement. This also illustrates the weak dependence of the flux-surface shape on the current shaping parameter α_I.

The reconstructed soft x-ray emissivity profiles for the two colours are shown in figure 6. The mean of the GP posterior distribution is shown black, and the 2σ uncertainty region is indicated by the grey shaded region. The amplitudes of the linear spline knots for the parametric reconstructions are indicated by the green bullets, and the error bars also indicate the 2σ uncertainty. The reconstructed profiles inferred are in good agreement; however, the uncertainty in parametric reconstructed profiles is larger than the uncertainty in GP reconstructed profiles throughout most of the domain. The second soft x-ray emissivity profile (figure 6*b*) has small amplitude in the outer flux region $s > 0.5$, and the corresponding soft x-ray measurements have a small signal to noise ratio. Thus the uncertainty in this region is large.

In general the reconstruction of the CTH equilibrium shows good agreement between the two methods. In total the two reconstructions used a total of 135 diagnostics to constrain the equilibrium profiles. The hybrid GP reconstruction used a total six parameters and the final equilibrium had a chi-squared value of $\chi^2 = 132$. This equates to an average chi-squared value of $\langle\chi^2\rangle = 0.98$ per signal. In contrast the fully parametric reconstruction used a total of 26 free parameters. It had a chi-squared value of $\chi^2 = 114$ and an average value of $\langle\chi^2\rangle = 0.85$ per signal.

https://doi.org/10.1017/S0022377819000813 Published online by Cambridge University Press

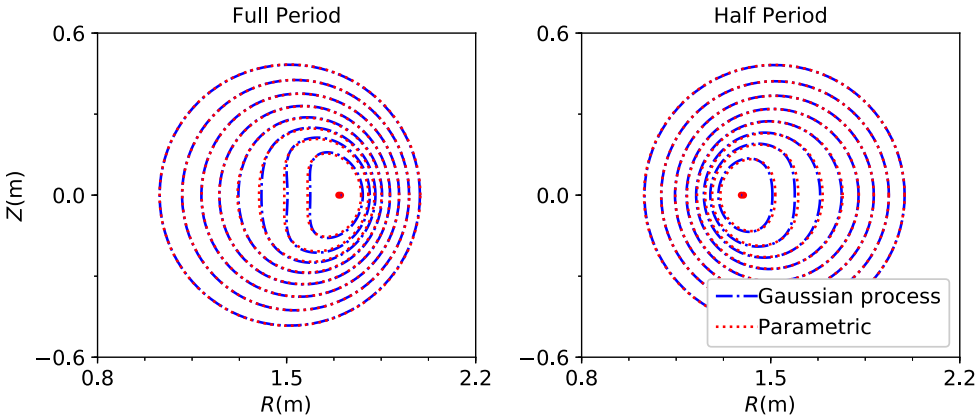

FIGURE 7. Reconstructed flux surface for the Gaussian process reconstruction (blue) and the fully parametric reconstruction (red) at two toroidal angles.

4.3. *MST single helicity axis state*

A reconstruction of a single-helical-axis (SHAx) state in Madison Symmetric torus (MST) (Dexter *et al.* 1991) is used as a second experimental test case. MST is normally operated as a reversed field pinch with a minor radius of $a = 0.52$ m and a major radius of $R = 1.52$ m. The toroidal magnetic field has a maximum value of $B_\phi \leqslant 0.5$ T at the magnetic axis, the field decreases away from the magnetic axis, and it reverses sign near the edge. During standard operation MST is characterized by an axisymmetric equilibrium. At high plasma current, $I_p \approx 500$ kA, the MST plasma spontaneously transitions to a three-dimensional SHAx equilibrium (Bergerson *et al.* 2011). V3FIT has previously been used to reconstruct these SHAx states using the standard parametric representation (Koliner *et al.* 2016). Here one of these previously reported SHAx equilibria is used to benchmark the hybrid Gaussian process model in V3FIT.

This MST reconstruction use approximately 200 diagnostics, which include both a poloidal array and a toroidal array of magnetic diagnostics, a single-colour multi-chord soft x-ray system, a multipoint Thomson scattering diagnostic and a far-infrared interferometry/polarimetry diagnostic. The only difference between the two reconstructions is that the parametric reconstruction uses a cubic spline to represent the temperature profile, where the hybrid reconstruction uses a GP to reconstruct this profile. Here, the optimal hyper-parameters are calculated as part of the minimization of the error. Both reconstructions use cubic splines to represent the pressure profile and the safety-factor profile. A two-power profile is used to represent the soft x-ray emissivity profile. The density profile is inferred from the pressure profile and the electron temperature profile. This is in contrast to the reconstructions presented in the paper by Koliner *et al.* (2016), where the density profile was directly reconstructed and the temperature profile was inferred from pressure and density profiles. This reconstruction is designed to test the GPR implementation for a Thomson scattering diagnostic using real experimental data; hence, the choice to explicitly model the temperature profile instead of the density profile.

The hybrid GP reconstruction and fully parametric reconstructions provide similar values for all their mutually shared parameters. This is illustrated in figure 7 which shows the flux surfaces from the two reconstructions. The fact that the two sets of

https://doi.org/10.1017/S0022377819000813 Published online by Cambridge University Press

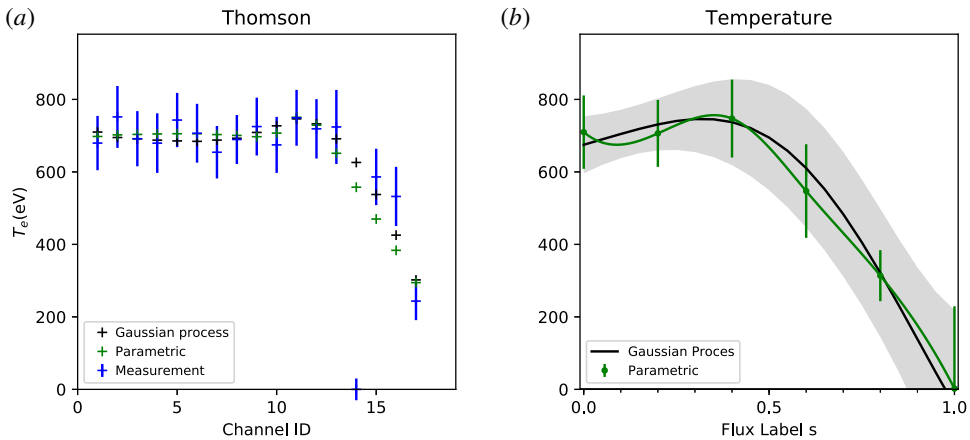

FIGURE 8. (*a*) The experimental and model electron temperature signals for each of the Thomson diagnostic channels. The experimentally measured signals with their corresponding error bars are shown in blue. The model signal value for the hybrid Gaussian process reconstruction is shown in black. The model signal value for the fully parametric reconstruction is shown in green. (*b*) The GP reconstructed temperature profile is shown in black. The grey shaded region represents the 2σ uncertainty in the GP fit. The parametric reconstructed temperature profile is shown in green, the errors indicated the 2σ uncertainty in the amplitude of the spline knots.

flux surfaces lie on top of each other indicates that the two MHD equilibria are in close agreement. However, the Thomson system on MST does not topographically constrain the flux-surface shape, and the temperature profile only weakly constrains the equilibrium flux-surface shape via the pressure profile.

Figure 8 compares reconstructed temperature profiles for the two methods. Figure 8(*a*) shows the modelled temperature at each of the Thomson viewing locations. Both the GP temperature profile and the parametric temperature profile agree with the measured data within 2σ of the experimentally measured temperature at each of the Thomson viewing locations. Channel 14 produced a bad signal for this particular shot, and it is not used in the reconstruction. Figure 8(*b*) compares the Gaussian process temperature profile with the parametric cubic spline temperature profile. The two profiles agree within 2σ uncertainty, and the two reconstructions give similar estimates of the uncertainty in the reconstruction.

One advantage of the GP profile is apparent near the core $s \lesssim 0.4$. In this region the parametric reconstruction exhibits oscillatory behaviour characteristic of overfitting, where the Gaussian process is not oscillatory. These oscillations can negatively impact core transport and stability analysis which are sensitive to derivatives of the equilibrium profiles. The oscillations can be removed by reducing the number of knots in the spline and/or adjusting their locations. This process requires trial and error. In contrast the Gaussian process automatically eliminates this behaviour in this particular example.

The parametric reconstruction has an average $\langle \chi^2 \rangle = 1.36$ per signal and uses 26 free parameters. The Gaussian process reconstruction has a comparable average value of $\langle \chi^2 \rangle = 0.47$ (this excludes the log evidence term). The Gaussian process reconstruction uses 20 free parameters and two free hyper-parameters.

https://doi.org/10.1017/S0022377819000813 Published online by Cambridge University Press

5. Discussion and conclusions

This paper introduced a new hybrid regression model that has been implemented in the 3-D reconstruction code V3FIT. This hybrid regression model uses a combination of parametric and non-parametric techniques to infer radial profiles from experimental data during equilibrium reconstruction. Gaussian process regression is used to infer radial profiles from diagnostic signals that are modelled using a linear operator acting on the underlying profile. Standard parametric techniques are used to infer the remaining profiles and equilibrium quantities. The linear assumption simplifies the calculation of the Gaussian process posterior, and the resulting mean and covariance matrices can then be calculated using standard linear algebra techniques.

One of the advantages of using Gaussian process regression is that the profile shape is inferred directly from the experimental data. The equilibrium reconstruction process strongly depends on the choices that the user makes. A bad parameterization can introduce systematic errors that exclude certain types of behaviour, introduce oscillations due to overfitting, etc. Gaussian process regression helps address the above issues by reducing the number of inputs that a user has to make. For example, instead of asking the user to specify a functional form for the profile, use of GP only requires that the user specifies a much broader class of functions (defined by specifying a covariance kernel) to which the profile belongs.

Another advantage of the Gaussian process formalism is that the calculation yields both the posterior mean and the posterior covariance matrix for linear models. These two quantities completely specify the posterior distribution of the reconstructed profile. In post-processing one can use these quantities to randomly sample the posterior profile. These randomly sampled profiles can then be used for sensitivity analysis. For example M. Chilenski used temperature and density profiles sampled from a Gaussian process to do error propagation analysis of turbulent transport fluxes (Chilenski *et al.* 2015). A similar analysis can be applied to other fusion calculations.

Comparison between fully parametric reconstructions and the hybrid GP reconstructions show good agreement between the two methods. These comparisons use experimental data from CTH and MST. In both cases the fully parametric reconstructions use linear or cubic splines to model the radial profiles. These profiles are then inferred using GPR in the hybrid reconstruction.

The work presented in this paper represents the first steps to incorporating Gaussian process regression into V3FIT. A simple, yet powerful, covariance kernel is used, and the implementation is only valid for linear diagnostics. The modularity of the code makes it straightforward to extend the functionality of the GP model. Minor code modifications would be needed to implement alternative kernel functions.

Acknowledgements

The authors would like to thank J. Svensson for helpful conversations and the CTH and MST experimental teams for providing data used in this work. This work is supported by Auburn University and by the U.S. Department of Energy Office of Science, Office of Fusion Energy Sciences program under the award no. DE-FG02-03ER54692 and DE-SC0018313.

REFERENCES

BERGERSON, W. F., AURIEMMA, F., CHAPMAN, B. E., DING, W. X., ZANCA, P., BROWER, D. L., INNOCENTE, P., LIN, L., LORENZINI, R., MARTINES, E. *et al.* 2011 Bifurcation to 3D helical magnetic equilibrium in an axisymmetric toroidal device. *Phys. Rev. Lett.* **107** (25), 1–5.

https://doi.org/10.1017/S0022377819000813 Published online by Cambridge University Press

CHILENSKI, M., GREENWALD, M., MARZOUK, Y., HOWARD, N., WHITE, A., RICE, J. & WALK, J. 2015 Improved profile fitting and quantification of uncertainty in experimental measurements of impurity transport coefficients using Gaussian process regression. *Nucl. Fusion* **55** (2), 023012.

CIANCIOSA, M., HIRSHMAN, S. P., SEAL, S. K. & SHAFER, M. W. 2018 3D equilibrium reconstruction with islands. *Plasma Phys. Control. Fusion* **60** (4), 044017.

DEXTER, R. N., KERST, D. W., LOVELL, T. W., PRAGER, S. C. & SPROTT, J. C. 1991 The Madison symmetric torus. *Fusion Technol.* **19** (1), 131–139.

GRAD, H. & RUBIN, H. 1958 Hydromagnetic equilibria and force-free fields. *J. Nucl. Energy* **7** (3–4), 284–285.

HANSON, J. D., HIRSHMAN, S. P., KNOWLTON, S. F., LAO, L. L., LAZARUS, E. A. & SHIELDS, J. M. 2009 V3FIT: a code for three-dimensional equilibrium reconstruction. *Nucl. Fusion* **49** (7), 075031.

HARTWELL, G. J., KNOWLTON, S. F., HANSON, J. D., ENNIS, D. A. & MAURER, D. A. 2017 Design, construction, and operation of the compact toroidal hybrid. *Fusion Sci. Technol.* **72** (1), 76–90.

HERFINDAL, J. L., DAWSON, J. D., ENNIS, D. A., HARTWELL, G. J., LOCH, S. D. & MAURER, D. A. 2014 Design and initial operation of a two-color soft x-ray camera system on the Compact Toroidal Hybrid experiment. *Rev. Sci. Instrum.* **85** (11), 11D850.

HIRSHMAN, S. P. & WHITSON, J. C. 1983 Steepest-descent moment method for three-dimensional magnetohydrodynamic equilibria. *Phys. Fluids* **26** (12), 3553.

KOLINER, J. J., CIANCIOSA, M. R., BOGUSKI, J., ANDERSON, J. K., HANSON, J. D., CHAPMAN, B. E., BROWER, D. L., DEN HARTOG, D. J., DING, W. X., DUFF, J. R. *et al.* 2016 Three dimensional equilibrium solutions for a current-carrying reversed-field pinch plasma with a close-fitting conducting shell. *Phys. Plasmas* **23** (3), 032508.

KWAK, S., SVENSSON, J., BRIX, M. & GHIM, Y.-C. 2017 Bayesian electron density inference from JET lithium beam emission spectra using Gaussian processes. *Nucl. Fusion* **57** (3), 036017.

LAO, L. & FERRON, J. 1990 Equilibrium analysis of current profiles in tokamaks. *Nucl. Fusion* **35**, 1035.

LI, D., SVENSSON, J., THOMSEN, H., MEDINA, F., WERNER, A. & WOLF, R. 2013 Bayesian soft x-ray tomography using non-stationary Gaussian processes. *Rev. Sci. Instrum.* **84** (8), 083506.

MA, X., CIANCIOSA, M. R., ENNIS, D. A., HANSON, J. D., HARTWELL, G. J., HERFINDAL, J. L., HOWELL, E., KNOWLTON, S. F., MAURER, D. A. & TRAVERSO, P. J. 2018 Determination of current and rotational transform profiles in a current-carrying stellarator using soft x-ray emissivity measurements. *Phys. Plasmas* **25**, 012516.

MA, X., MAURER, D. A., KNOWLTON, S. F., ARCHMILLER, M. C., CIANCIOSA, M. R., ENNIS, D. A., HANSON, J. D., HARTWELL, G. J., HEBERT, J. D., HERFINDAL, J. L. *et al.* 2015 Non-axisymmetric equilibrium reconstruction of a current-carrying stellarator using external magnetic and soft x-ray inversion radius measurements. *Phys. Plasmas* **22** (12), 122509.

RASMUSSEN, C. & WILLIAMS, C. 2006 *Gaussian Processes for Machine Learning*. The MIT Press.

SHAFRANOV, V. D. 1966 Plasma equilibrium in a magnetic field. *Rev. Plasma Phys.* **2**, 103.

SVENSSON, J. 2010 Non-parametric tomography using Gaussian processes. *JET Intern. Rep.* pp. EFDA–JET–PR(11)24.

https://doi.org/10.1017/S0022377819000813 Published online by Cambridge University Press