Quantifying Experimental Edge Plasma Evolution via Multidimensional Adaptive Gaussian Process Regression

Abhilash Mathews and Jerry W. Hughes

***Abstract*—The edge density and temperature of tokamak plasmas are strongly correlated with energy and particle confinement and their quantification is fundamental to understanding edge dynamics. These quantities exhibit behaviours ranging from sharp plasma gradients and fast transient phenomena (e.g. transitions between low and high confinement regimes) to nominal stationary phases. Analysis of experimental edge measurements therefore require robust fitting techniques to capture potentially stiff spatiotemporal evolution. Additionally, fusion plasma diagnostics inevitably involve measurement errors and data analysis requires a statistical framework to accurately quantify uncertainties. This paper outlines a generalized multidimensional adaptive Gaussian process routine capable of automatically handling noisy data and spatiotemporal correlations. We focus on the edge-pedestal region in order to underline advancements in quantifying time-dependent plasma profiles including transport barrier formation on the Alcator C-Mod tokamak.**

***Index Terms*—Gaussian process, plasma diagnostics, critical gradients.**

I. INTRODUCTION

THE edge-pedestal of fusion plasmas traverses a vast range of physical scales. At the top of the pedestal are temperatures exceeding the sun's core at nearly 10^7 K while separated sometimes just centimeters from solid plasma-facing components having operational surface temperatures of $\approx 10^3$ K. This physical span in tokamaks includes some of the strongest temperature gradients observed in the universe and is critical towards overall reactor performance, stability, and fuelling [1]–[4]. Fundamental plasma properties (e.g. collisionality, ionization rate) consequently vary strongly both spatially and temporally in this region and require capable numerical analysis tools since quantifying the pedestal is essential towards understanding dynamics near the last closed flux surface (LCFS) and across the scrape-off layer (SOL). The advent of scientific machine learning could potentially advance understanding in this region as well, but requires constructing good statistical estimates of partially observed variables from diagnostic measurements. Automatically generating accurate pedestal plasma density and temperature profiles requires sufficiently sophisticated tools to handle large quantities of noisy observations. Current fitting routines in the fusion community involve a range of methods including nonlinear least squares via modified hyperbolic tangent functions [5] such as in Figure 1, cubic splines [6], and various Bayesian techniques [7]. Past Gaussian process (GP) regression codes typically fixed covariance function length scales [8] and generally permitted only one-dimensional scenarios to build radial profiles. This requires filtering or averaging temporal variation in data which can be limiting in the edge especially when analyzing transient phenomena such as spontaneous transitions of confinement regimes and transport barrier formation [9], [10]. Robustness to capture both mean spatial and temporal variations of edge plasma profiles and associated gradients is therefore sought for improving analysis.

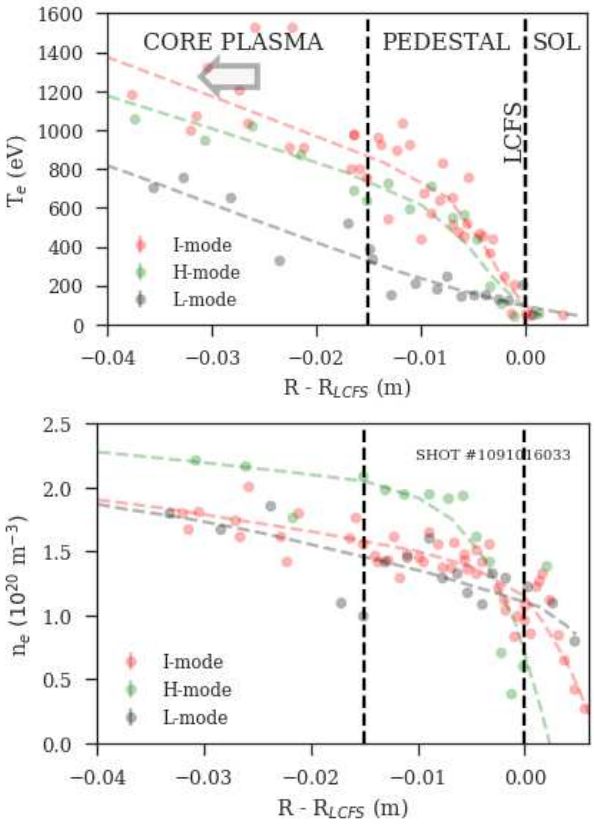

Fig. 1. Electron density and temperature measurements fit by a modified hyperbolic tangent via nonlinear least squares exhibiting strong—albeit limited in fitted structure—profile variation across confinement regimes in a single discharge. Data and conventional fitting of profiles adapted from [11].

Towards this task a deep multidimensional heteroscedastic GP routine is outlined to provide automated fitting and uncertainty estimates from the Thomson scattering diagnostic

This work is supported by the U.S. Department of Energy (DOE) Office of Science under the Fusion Energy Sciences program by contract DE-SC0014264 and the Natural Sciences and Engineering Research Council of Canada (NSERC) by the doctoral postgraduate scholarship (PGS D)

on the Alcator C-Mod tokamak. Evolution of Both plasma density and temperature is tracked across the edge-pedestal region with varying length scales extant in experimental data including the formation of both particle and energy transport barriers. This technique has the capability to routinely process thousands of discharges automatically to yield profile statistics and be run across novel experiments. The applied methodology is described in section 2 (with accompanying mathematical proofs in the Appendix), demonstrated on relevant scenarios exhibiting transport barriers from real experiments in section 3, and final summarizing remarks are presented in section 4.

II. METHODOLOGY

The technique applied for reconstructing edge-pedestal plasma profiles is an adaptive heteroscedastic multidimensional GP routine. Each of these terms are now individually defined and outlined to introduce the scope of this method.

A. Gaussian process

A GP is a supervised learning method capable of solving classification and regression problems. The latter fitting capability via nonlinear probabilistic regression is the main focus of this paper. In particular, the underlying assumption is that the variable (e.g. plasma density) being predicted at a certain location is normally distributed and spatiotemporally correlated with neighbouring points, indicating that partial observations provide information at nearby locations for conditioning future predictions. The function space definition of a GP is that any finite collection of the random variables modelled follow a joint Gaussian distribution, i.e. $(y_i, y_j) \sim \mathcal{N}(\mu, \Sigma)$ and any subset is given by $y_i \sim f(\mathbf{x}_i) + \sigma_n \mathcal{N}(0, I)$ [12], where σ_n^2 is the noise variance. A GP is specified entirely up to its second-order statistics as denoted by the mean, $\mu(\mathbf{x}_i)$, and covariance, $\Sigma(\mathbf{x}_i, \mathbf{x}_j)$. Consequently, it can be proven (c.f. Appendix) that conditional predictions at $\mathbf{x}_b$ based upon observations at $\mathbf{x}_a$ are analytically given by [12]:

$$\boldsymbol{\mu}_{\mathbf{y}_b|\mathbf{y}_a} = \boldsymbol{\mu}_b + \Sigma_{b,a}{\Sigma_{a,a}}^{-1}(\boldsymbol{y}_a - \boldsymbol{\mu}_a) \tag{1}$$

$$\Sigma_{\mathbf{y}_b|\mathbf{y}_a} = \Sigma_{b,b} - \Sigma_{b,a}{\Sigma_{a,a}}^{-1}\Sigma_{a,b} \tag{2}$$

Suitably selecting the kernel function, $k(\mathbf{x}_i, \mathbf{x}_j)$, representing the full covariance is critical to the GP since it specifies the correlation between any pairs of random variables. Constraining kernels will consequently limit the range of behaviour that can be captured by the GP which may only be physically warranted in certain scenarios. To remain robust to tracking a wide range of spatiotemporal behaviour, an adaptive heteroscedastic kernel is optimized against experimental observations. Applicability depends upon the case, but utilizing alternative distributions (e.g. log-normal) and latent variable transformations can also permit scenarios with non-Gaussian residuals [13], although Gaussianity is assumed here.

B. Adaptivity

GP regression is a nonparametric method without an explicit functional form. Nonparametric in this context means that there are no fixed number of constraining model parameters but instead the fitting routine becomes increasingly constrained as training data increases. Correlations between observed data points are based upon the prescribed covariance function. Various kernels have been proposed to embed this structure ranging from a Gaussian function (for expectedly smooth behaviour) to periodic functions (for expectedly cyclic behaviour) to combinations of multiple kernels (e.g. for automatic relevance determination) [14]. Despite the generic regression technique lacking a strictly fixed functional form, the optimized hyperparameters defining the kernel are typically constrained themselves. For example, a standard stationary isotropic Matérn kernel is defined by

$$k(\mathbf{x}_i, \mathbf{x}_j) = \sigma_k^2 \frac{2^{1-\nu}}{\Gamma(\nu)} \left(\sqrt{2\nu} \frac{|\mathbf{x}_i - \mathbf{x}_j|}{\rho} \right)^{\nu} K_\nu \left(\sqrt{2\nu} \frac{|\mathbf{x}_i - \mathbf{x}_j|}{\rho} \right) \tag{3}$$

where Γ is the gamma function, K_ν is the modified Bessel function of the second kind, and ρ and ν are non-negative globally constant hyperparameters which control spatial range and smoothness, respectively. A Matérn kernel is $\nu - 1$ times differentiable and reduces to a Gaussian kernel in the limit $\nu \to \infty$ while becoming an exponential kernel when $\nu = 1/2$ [15]. It resultantly covers a wide class of kernels and confers flexibility. Nevertheless, the hyperparameters are quite restrictive if simply constants [16], [17]. Therefore, a version of the generalized nonstationary Matérn kernel is encoded for GP regression [18], [19]:

$$k(\mathbf{x}_i, \mathbf{x}_j) = \sigma_k^2 \frac{2^{1-\nu}}{\Gamma(\nu)} \frac{|\rho_i|^{1/2}|\rho_j|^{1/2}}{\sqrt{\frac{1}{2}\rho_i^2 + \frac{1}{2}\rho_j^2}} \left(2\sqrt{\frac{2\nu|\mathbf{x}_i - \mathbf{x}_j|^2}{\rho_i^2 + \rho_j^2}} \right)^{\nu} K_\nu \left(2\sqrt{\frac{2\nu|\mathbf{x}_i - \mathbf{x}_j|^2}{\rho_i^2 + \rho_j^2}} \right) \tag{4}$$

where ρ varies across the entire multidimensional domain and adapts to optimize the length scale based upon the experimental data being trained upon in each individual plasma discharge. The kernel accomplishes learning point estimates of local smoothness by representing the primary GP's locally isotropic length scale hyperparameter by a secondary GP with radial basis function (RBF) kernel that allows global variation of ρ across the spatiotemporal grid. It is this second-level GP which introduces the notion of a deep process and adaptivity to the overall regression technique. A stationary kernel is purely a function of $\mathbf{x_i} - \mathbf{x_j}$, while additional local dependence exists in (4) through ρ which introduces nonstationary behaviour [19]. A k-means clustering algorithm is used for training of the secondary GP which parametrizes the nonstationary covariances with latent length-scales [19].

Figure 2 demonstrates a basic 1-dimensional example of a sinusoidal function with imposed discontinuity. The advantage conferred by the adaptive length scale can be quantitatively observed by comparing the log marginal likelihood (LML) [12] between a standard Matérn kernel and one with locally

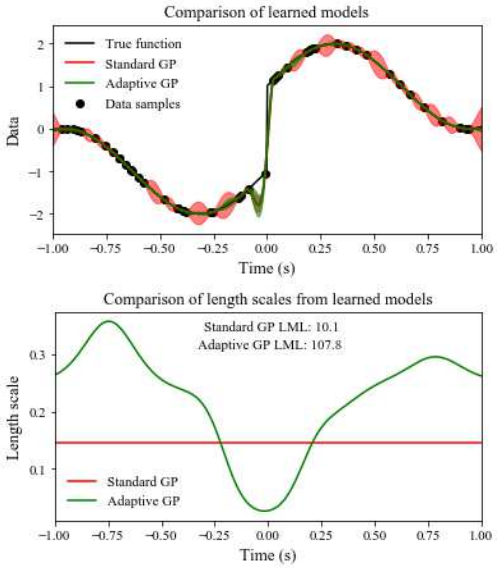

Fig. 2. Comparison of two separate GPs on 1-dimensional data: one employs a standard Matérn kernel while the other includes an adaptive length scale. Fits to the original data samples (top) and computed length scales (bottom) are displayed courtesy of J.H. Metzen.

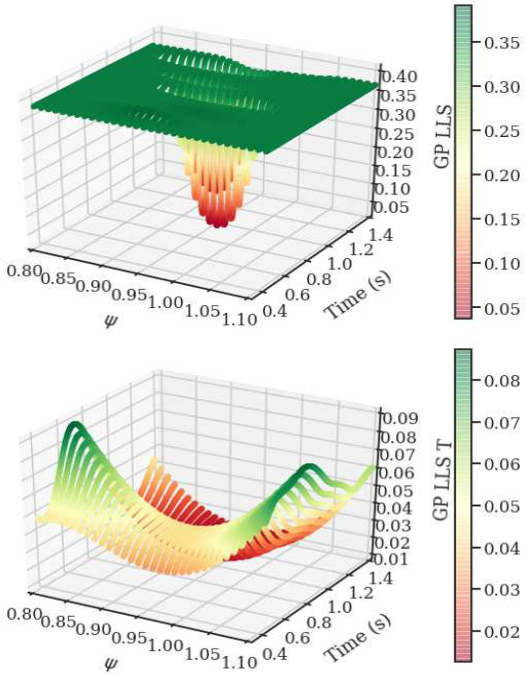

Fig. 3. Two examples of adaptively learning length scales that vary across the spatiotemporal domain by training identical GP models on experimentally measured electron density (top) and temperature (bottom) from the Thomson scattering plasma diagnostic.

adaptive length scale. The order of magnitude improvement of LML (which is a logarithmic quantity) in Figure 2 occurs because a stationary Matérn kernel is forced to decrease its constant global length scale considerably while an adaptive length scale permits reducing its value locally only near the discontinuity. The adaptive length scales not only provide the capability to better capture singular or transient phenomena on otherwise slowly-varying profiles but importantly improves uncertainty estimates across the domain. We set $\nu = 3/2$ which allows for potentially stiff behaviour and this value can be modified in the source code, if sought, and kept as a variable for optimization to capture an entire spectrum of kernel functions. The user can freely specify upper and lower bounds on length scales to be learned across the grid which are given uniform prior distributions during training. As a preview for the application of these methods, the length scales can be extended to multidimensional scenarios as depicted in Figure 3 where the learned input length scale varies across the entire spatial and temporal domain. The exact same kernel was initialized for both the top and bottom data sets used in Figure 3, but since the datasets were themselves different, the locally learned length scales vary. A custom stochastic optimizer based on differential evolution is used in these examples for GP hyperparameter-tuning and finally polished off with gradient-based descent. Finding a global optimum in the likelihood function is not guaranteed, therefore this is helpful because the loss function is highly multimodal and generally challenging for simple gradient-based methods acting on non-convex problems.

C. Heteroscedasticity

Heteroscedasticity in this context refers to learning intrinsic scatter associated with the noisy variable, y, and introducing non-constant variances. The full covariance function applied in the GP can be broken down into an adaptive kernel and noisy variance component:

$$\Sigma(\mathbf{x}_i, \mathbf{x}_j) = \underbrace{k(\mathbf{x}_i, \mathbf{x}_j)}_{\text{adaptive}} + \underbrace{\sigma_n^2(\mathbf{x}_i)\delta_{ij}}_{\text{heteroscedastic}} \tag{5}$$

and heteroscedasticity is mathematically defined by $\sigma_n(\mathbf{x}_i)$ having an explicit dependence on points in the input space. To contrast, homoscedasticity would entail a globally constant σ_n. To more vividly demonstrate the benefit of heteroscedasticity, a 1-dimensional example is displayed in Figure 4 by applying both homoscedastic and heteroscedastic components. The function to be learned is a linear relationship with variance growing quadratically along the abscissa. In this scenario with a homoscedastic noise model, the GP is forced to learn a constant intrinsic scatter in the underlying data commonly represented by white noise functions of constant amplitude. It is evident that enabling a heteroscedastic covariance function better captures the distribution of observed data. The mean estimates of both models are equivalent, but the predicted variances across the domain are markedly different and the heteroscedastic example obtains a larger LML and better captures intrinsic scatter in the data. The non-constant variances are learned across the domain using a k-means algorithm to once again identify clustering centers to provide characteristic noise estimates even when explicit error bars are absent. These prototype values are then extended across the domain by using a pairwise RBF kernel. Both the adaptive length scale kernel and heteroscedastic components are combined to significantly improve overall fitting and stability of the numerical optimization, e.g. avoid zero variances while training. This

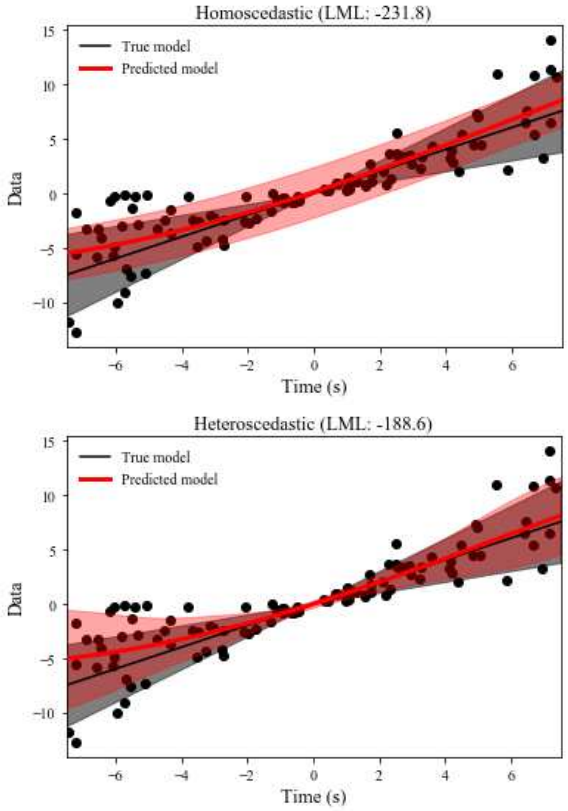

Fig. 4. Comparison of two separate GPs on 1-dimensional data both using an RBF kernel but one utilizes an additional homoscedastic component (top) and the other a heteroscedastic component (bottom) in the full covariance structure. Courtesy of J.H. Metzen.

heteroscedastic term in the full data-driven covariance function is modular to an extent and can be optionally subtracted away to output confidence intervals ($\mathbb{V}[f_*]$) instead of prediction intervals ($\mathbb{V}[y_*]$) which are wider and account for intrinsic scatter in observations. (Note that $\mathbb{E}[f_*] \equiv \mathbb{E}[y_*]$.)

D. Multidimensional

GPs do not require a fixed discretization scheme over the physical domain and hence easily handle spatially and temporally scattered diagnostic measurements involved in nonlinear regression. The generalized kernel above is encoded to handle data spanning any number of dimensions. Due to the highly non-convex optimization problem associated with finding optimal hyperparameters across the multidimensional space, the training applies stochastic optimization (Storn algorithm) with gradient descent at the end to help avoid becoming trapped in local minima. An associated error checking routine has been developed and available in the code on GitHub to automatically identify regions of the domain, if any, over which the GP did not successfully converge during training and requiring further optimization. Generally, multidimensional sampling of the GP is performed by applying

$$f_* = \mu_* + B\mathcal{N}(0, I) \tag{6}$$

where $BB^T = \Sigma_*$ and B is a triangular matrix known as the Cholesky decomposition of the covariance matrix.

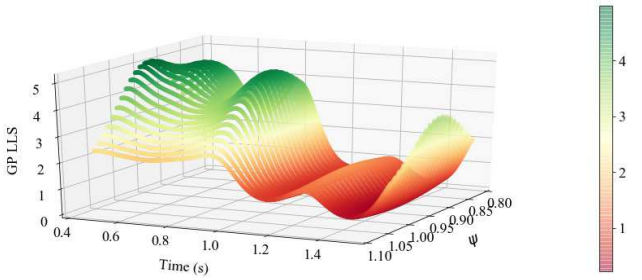

Fig. 5. Corresponding length scales, ρ, learned across the spatial and temporal domain based on the adaptive heteroscedastic GP training upon spatiotemporally evolving data.

To account for particular constraints such as monotonicity or positivity with respect to the mean in sampled profiles, modified or truncated normal distributions can be enabled in the code to permit constrained GP sampling to yield physically relevant results. Finally, the technique's flexibility allows it to automatically handle vastly different data sets (e.g. thousands of separate discharges fitted in parallel) to construct large scale multidimensional profile databases with minimal user input.

III. APPLICATION TO EXPERIMENTAL DATA

The aforementioned adaptive heteroscedastic GP is now directly applied on experimental data originating from the Thomson scattering diagnostic on Alcator C-Mod which consists of two Nd:Yag lasers, each pulsing at 30 Hz with an approximate spatial resolution of 1.0 cm and 1.3 mm in the core and edge, respectively [20]. Accordingly, we can spatially resolve electron dynamics on scales of the poloidal ion gyroradius. We note that the GP itself is a continuous regression model and can provide fittings at arbitrary spatial and temporal resolution. Shot 1091016033 is analyzed which is partly displayed with time-averaged profiles in Figure 1 and exhibits L-, H-, and I-mode behaviours within this single discharge as detailed in [11]. Ion cyclotron range of frequencies (ICRF) heating of up to 5 MW is applied in the experiment with power primarily deposited in the core on the hydrogen minority species. The on-axis magnetic field is 5.6 T with a plasma current of approximately 1.2 MA. There is a wide range of plasma behaviour associated with time-varying density and temperature pedestal structure even in this single discharge including transport barrier formation and confinement mode transitions necessitating a suitably robust regression method. The tools outlined above are therefore now demonstrated on shot 1091016033, and can be easily automated for edge data analyses across any set of discharges.

To begin, the trained GPs can compute expected mean density and temperature along with corresponding uncertainties across the entire spatiotemporal domain covered by the plasma diagnostic. Spatial gradients (or time derivatives) can be directly produced and are displayed in Figure 7 for electron density across the grid as well. The key advantage of applying the adaptive GP for edge profile fitting is its ability to learn spatiotemporal correlations across the entire domain without any temporal averaging nor spatial filtering of data. This

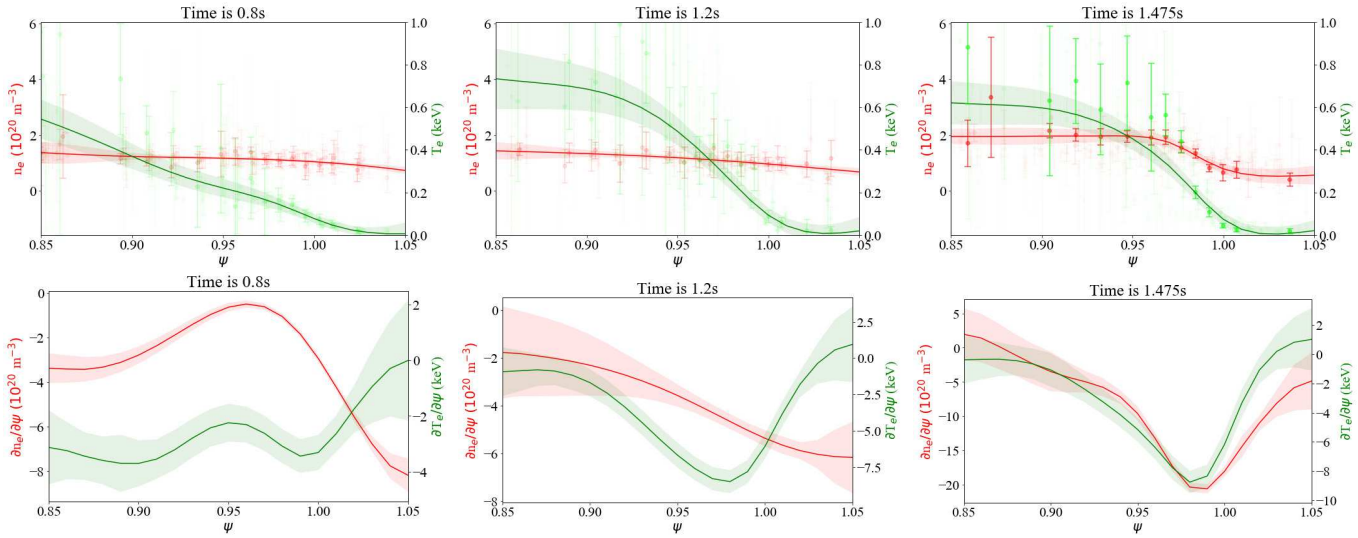

Fig. 6. Electron density and temperature measurements during L- (left), I- (middle), and H-modes (right) fitted by the adaptive heteroscedastic GP without time-averaging experimental data. Proximity of experimental data to the time at which the GP is predicting is indicated by transparency of the data points. For simple demonstrative purposes, 95% prediction intervals are displayed in the top three plots while 95% confidence intervals are applied for the bottom.

freedom in training is evident in the learned variable length scales for density across the discharge as visualized in Figure 5. Particular time slices can also be evaluated by the GPs. Fitted L-, I-, and H-mode profiles from shot 1091016033 are displayed in Figure 6 without employing any time-averaging or filtering of experimental data as required when repeatedly applying a modified tanh function as in Figure 1, which can miss important profile variation even within a confinement regime (e.g. while ramping up ICRF heating). Furthermore, density and temperature gradients along with uncertainties for all quantities are well-defined across the experiment during nominal L-, H-, and I-mode phases corresponding to times of 800, 1200, and 1475 milliseconds, respectively [11].

A major benefit derived from applying the adaptive heteroscedastic GP is its ability to provide defining features (e.g. spatial and temporal gradients) of experimental profiles automatically across entire discharges. In past classification exercises to develop large confinement regime databases [10], individual time slices needed to be manually reviewed for the presence of density and/or temperature pedestals to identify windows containing different regimes (e.g. L-, H-, or I-modes). This is a highly time-intensive and arguably subjective route to develop meaningful confinement regime databases since discretely classifying plasma behaviour can miss underlying nuances (e.g. staircase pedestals, profile hollowing). Therefore, applying this multidimensional GP regression method permits automating characterization of profiles based upon certain observed quantitative features of confinement regimes such as gradient scale lengths. Namely, it can handle pertinent dynamics across the edge-pedestal region where properties atop the pedestal are expected to be different from relevant spatiotemporal scales in the relatively colder SOL when crossing the separatrix. Different varieties of stationary confinement regimes may have quite different profile dynamics and struc-

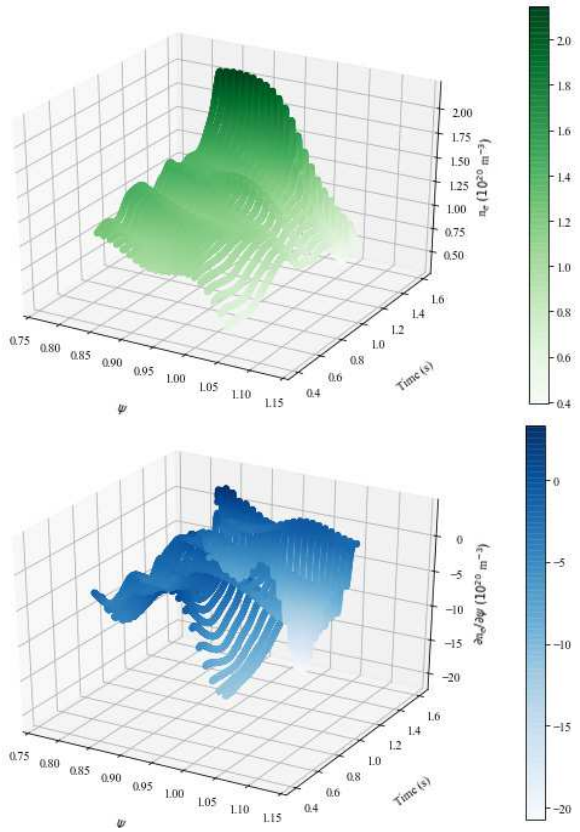

Fig. 7. Electron density and spatial gradients fitted by the adaptive heteroscedastic GP accounting for both spatial and temporal evolution of experimental data across the entire discharge over the edge-pedestal region (i.e. $0.85 < \psi < 1.05$, where ψ is the normalized poloidal flux coordinate and applied for tracking data on both open and closed field lines).

ture for temperature and density necessitating adaptive fitting capability. Additionally, accounting for temporal evolution of plasma profiles helps capture subtleties in types of transitions which are essential to better understand how confinement regimes evolve since no two modes nor transitions are necessarily identical, especially with hysteresis being prevalent across these bifurcation events. Resolving these features helps improve reconstructions of experimental measurements for further large scale analysis whether in stability codes or scientific machine learning or characterizing upstream conditions of background plasmas for divertor heat flux studies [21]. For example, the multidimensional GP can provide key profile information into plasma turbulence simulations (e.g. inputs for global gyrokinetic codes or comparisons with EPED [2]) which may require sampling inputs such as gradient profiles to output sufficient statistics. Overall, this regression tool can be deployed for automatically providing multidimensional density and temperature profiles involving few sharp features for numerical analysis purposes. While the GP does not explicitly find L-, H-, or I-modes, it outputs time-dependent gradient profiles and corresponding uncertainties to systematically scan characteristics of confinement regimes across entire discharges such as identifying steepening in edge temperature gradients and formation of staircase pedestals. Structure imposed on learned edge gradient profiles is minimized by using a data-driven kernel function which can be critical to model plasmas near sensitive instability boundaries. Additionally, the denoised equilibrium plasma dynamics can be useful observational constraints in scientific learning applications [22].

IV. Conclusion

An adaptive multidimensional GP routine is outlined and utilized to automate fitting and uncertainty estimation of edge-pedestal measurements from the Thomson scattering diagnostic on the Alcator C-Mod tokamak while trying to be robust to edge-pedestal phenomena. Spatiotemporal evolution of plasma density and temperature is tracked with varying length scales extant in experimental data including the formation of both particle and energy transport barriers. The application is focused on edge measurements of tokamak plasmas with relevance to numerical analysis of the pedestal, but these techniques extend beyond analysis of the edge-pedestal region and can be suitably adapted to novel scenarios exhibiting singular transient events. The GP introduced provides an automated tool to tackle nonlinear multidimensional regression tasks and helps resolve measurements of equilibrium profiles with dynamics spanning a wide range of physical scales.

Acknowledgement

We wish to give special credit and thanks to J.H. Metzen for original development of the code and package extensions created through `gp_extras` (github.com/jmetzen/gp_extras) and `scikit-learn`. This paper's examples can be found on Github (github.com/AbhilashMathews/gp_extras_applications) and the repository will become public upon publication.

Appendix

For completeness, a simple derivation of the GP machinery to perform conditional predictions from observed data, i.e. (1) and (2), based originally upon [23] is reproduced below. As previously remarked, a normally distributed random variable, $\mathbf{y}$, in N-dimensions is modelled by

$$P(\mathbf{y}|\boldsymbol{\mu},\boldsymbol{\Sigma}) = \frac{1}{2\pi^{N/2}|\boldsymbol{\Sigma}|^{1/2}}\exp[-\frac{1}{2}(\mathbf{y}-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\mathbf{y}-\boldsymbol{\mu})], \tag{A.1}$$

where $\boldsymbol{\Sigma}$ is a positive semi-definite covariance matrix, $\boldsymbol{\mu} = [\mu_1, ..., \mu_N]$, $\mathbf{y} = [y_1, ..., y_N]$, and $\boldsymbol{\Sigma}_{i,j} = \mathbb{E}[(y_i - \mu_i)(y_j - \mu_j)]$. In this formalism, the conditional probability of predicting a new point (or set of points), $\mathbf{y}_b$, can be ascertained from an observed point (or set of points), $\mathbf{y}_a$, through the posterior distribution: $\mathbf{y}_b|\mathbf{y}_a \sim \mathcal{N}(\boldsymbol{\mu}_{\mathbf{y}_b|\mathbf{y}_a}, \Sigma_{\mathbf{y}_b|\mathbf{y}_a})$ [12]. In a Bayesian framework, the multivariate normal distribution's conditional mean and variance are:

$$\boldsymbol{\mu}_{\mathbf{y}_b|\mathbf{y}_a} = \boldsymbol{\mu}_b + \Sigma_{b,a}{\Sigma_{a,a}}^{-1}(\boldsymbol{y}_a - \boldsymbol{\mu}_a) \tag{A.2}$$

$$\Sigma_{\mathbf{y}_b|\mathbf{y}_a} = \Sigma_{b,b} - \Sigma_{b,a}{\Sigma_{a,a}}^{-1}\Sigma_{a,b} \tag{A.3}$$

Following the treatment in [23], this can be derived by defining $\mathbf{z} \equiv \mathbf{y}_b + \mathbf{A}\mathbf{y}_a$ where $\mathbf{A} \equiv -\Sigma_{b,a}\Sigma_{a,a}^{-1}$, implying

$$\begin{aligned}\mathrm{cov}(\mathbf{z}, \mathbf{y}_a) &= \mathrm{cov}(\mathbf{y}_b, \mathbf{y}_a) + \mathrm{cov}(\mathbf{A}\mathbf{y}_a, \mathbf{y}_a)\\ &= \Sigma_{b,a} + \mathbf{A}\mathrm{var}(\mathbf{y}_a)\\ &= \Sigma_{b,a} - \Sigma_{b,a}\Sigma_{a,a}^{-1}\Sigma_{a,a}\\ &= 0\end{aligned} \tag{A.4}$$

Consequently, $\mathbf{z}$ and $\mathbf{y}_a$ are uncorrelated and, since they are assumed jointly normal in GP regression, they are independent. It is evident $\mathbb{E}[\mathbf{z}] = \boldsymbol{\mu}_b + \mathbf{A}\boldsymbol{\mu}_a$, and it follows that the conditional mean can be expressed as

$$\begin{aligned}\mathbb{E}[\mathbf{y}_b|\mathbf{y}_a] &= \mathbb{E}[]\mathbf{z} - \mathbf{A}\mathbf{y}_a|\mathbf{y}_b]\\ &= \mathbb{E}[\mathbf{z}|\mathbf{y}_a] - \mathbb{E}[\mathbf{A}\mathbf{y}_a|\mathbf{y}_a]\\ &= \mathbb{E}[\mathbf{z}] - \mathbf{A}\mathbf{y}_a\\ &= \boldsymbol{\mu}_b + \mathbf{A}(\boldsymbol{\mu}_a - \mathbf{y}_a)\\ &= \boldsymbol{\mu}_b + \Sigma_{b,a}\Sigma_{a,a}^{-1}(\mathbf{y}_a - \boldsymbol{\mu}_a)\end{aligned} \tag{A.5}$$

which proves the conditional mean, and

$$\begin{aligned}\mathrm{var}(\mathbf{x} - \mathbf{D}\mathbf{y}) \equiv \mathrm{var}(\mathbf{x}) &+ \mathbf{D}\mathrm{var}(\mathbf{y})\mathbf{D}^T\\ &- \mathrm{cov}(\mathbf{x}, \mathbf{y})\mathbf{D}^T - \mathbf{D}\mathrm{cov}(\mathbf{y}, \mathbf{x})\end{aligned} \tag{A.6}$$

implies

$$\begin{aligned}\mathrm{var}(\mathbf{y}_b|\mathbf{y}_a) &= \mathrm{var}(\mathbf{z} - \mathbf{A}\mathbf{y}_a|\mathbf{y}_a)\\ &= \mathrm{var}(\mathbf{z}|\mathbf{y}_a) + \mathrm{var}(\mathbf{A}\mathbf{y}_a|\mathbf{y}_a)\\ &- \mathbf{A}\mathrm{cov}(\mathbf{y}_a, \mathbf{z}) - \mathrm{cov}(\mathbf{z}, \mathbf{y}_a)\mathbf{A}^T\\ &= \mathrm{var}(\mathbf{z}|\mathbf{y}_a) = \mathrm{var}(\mathbf{z})\end{aligned} \tag{A.7}$$

Plugging the above result into the conditional variance,

$$\begin{aligned}\text{var}(\mathbf{y}_b|\mathbf{y}_a) &= \text{var}(\mathbf{y}_b + \mathbf{A}\mathbf{y}_a) \\ &= \text{var}(\mathbf{y}_b) + \mathbf{A}\text{var}(\mathbf{y}_a)\mathbf{A}^T + \\ &\mathbf{A}\text{cov}(\mathbf{y}_b, \mathbf{y}_a) + \text{cov}(\mathbf{y}_a, \mathbf{y}_b)\mathbf{A}^T \\ &= \Sigma_{b,b} + \Sigma_{a,b}\Sigma_{a,a}^{-1}\Sigma_{a,a}\Sigma_{a,a}^{-1}\Sigma_{a,b} \\ &\quad - 2\Sigma_{b,a}\Sigma_{a,a}^{-1}\Sigma_{a,b} \\ &= \Sigma_{b,b} + \Sigma_{b,a}\Sigma_{a,a}^{-1}\Sigma_{a,b} \\ &- \quad 2\Sigma_{b,a}\Sigma_{a,a}^{-1}\Sigma_{a,b} \\ &= \Sigma_{b,b} - \Sigma_{b,a}\Sigma_{a,a}^{-1}\Sigma_{a,b} \end{aligned} \tag{A.8}$$

Therefore, all that is required for the GP machinery are priors over $\mathbf{y}_a$ and $\mathbf{y}_b$ to obtain $\boldsymbol{\mu}_a$ and $\boldsymbol{\mu}_b$, respectively, and a covariance function which is defined in this paper by a heteroscedastic uncertainty component, ϵ, along with an adaptive kernel function, $k(\mathbf{x}_i, \mathbf{x}_j)$. The resulting covariance, $\Sigma_{i,j} = k(\mathbf{x}_i, \mathbf{x}_j) + \epsilon(\mathbf{x}_i)\delta_{i,j}$, describes similarity between data points and accounts for spatiotemporal correlations in the data since $\mathbf{x}$ represents the independent variables (e.g. ψ and t). The hyperparameters of the prescribed adaptive kernel function are optimized through maximum *a posteriori* estimation on each individual discharge's observed data. Practically, given an experimental set of noisy measurements at arbitrary positions and times, this formalism allows us to infer expected values of these measurements across our spatiotemporal domain.

REFERENCES

[1] R. Groebner, C. Chang, J. Hughes, R. Maingi, P. Snyder, X. Xu, J. Boedo, D. Boyle, J. Callen, J. Canik, I. Cziegler, E. Davis, A. Diallo, P. Diamond, J. Elder, D. Eldon, D. Ernst, D. Fulton, M. Landreman, A. Leonard, J. Lore, T. Osborne, A. Pankin, S. Parker, T. Rhodes, S. Smith, A. Sontag, W. Stacey, J. Walk, W. Wan, E.-J. Wang, J. Watkins, A. White, D. Whyte, Z. Yan, E. Belli, B. Bray, J. Candy, R. Churchill, T. Deterly, E. Doyle, M. Fenstermacher, N. Ferraro, A. Hubbard, I. Joseph, J. Kinsey, B. LaBombard, C. Lasnier, Z. Lin, B. Lipschultz, C. Liu, Y. Ma, G. McKee, D. Ponce, J. Rost, L. Schmitz, G. Staebler, L. Sugiyama, J. Terry, M. Umansky, R. Waltz, S. Wolfe, L. Zeng, and S. Zweben, "Improved understanding of physics processes in pedestal structure, leading to improved predictive capability for ITER," *Nuclear Fusion*, vol. 53, no. 9, p. 093024, Aug 2013.

[2] P. Snyder, N. Aiba, M. Beurskens, R. Groebner, L. Horton, A. Hubbard, J. Hughes, G. Huysmans, Y. Kamada, A. Kirk, C. Konz, A. Leonard, J. Lannroth, C. Maggi, R. Maingi, T. Osborne, N. Oyama, A. Pankin, S. Saarelma, G. Saibene, J. Terry, H. Urano, and H. Wilson, "Pedestal stability comparison and ITER pedestal prediction," *Nuclear Fusion*, vol. 49, no. 8, p. 085035, Jul 2009.

[3] J. Hughes, P. Snyder, M. Reinke, B. LaBombard, S. Mordijck, S. Scott, E. Tolman, S. Baek, T. Golfinopoulos, R. Granetz, M. Greenwald, A. Hubbard, E. Marmar, J. Rice, A. White, D. Whyte, T. Wilks, and S. Wolfe, "Access to pedestal pressure relevant to burning plasmas on the high magnetic field tokamak Alcator C-Mod," *Nuclear Fusion*, vol. 58, no. 11, p. 112003, Sep 2018.

[4] S. Mordijck, R. Groebner, J. Hughes, A. Diallo, F. Scotti, and V. A. Soukhanovskii, "Role of transport versus fueling upon pedestal density," *DOE Fusion Energy Sciences FY2019 Joint Research Target Report*, 9 2019.

[5] R. Groebner, D. Baker, K. Burrell, T. Carlstrom, J. Ferron, P. Gohil, L. Lao, T. Osborne, D. Thomas, W. West, J. Boedo, R. Moyer, G. McKee, R. Deranian, E. Doyle, C. Rettig, T. Rhodes, and J. Rost, "Progress in quantifying the edge physics of the H mode regime in diii-d," *Nuclear Fusion*, vol. 41, no. 12, pp. 1789–1802, Dec 2001.

[6] A. Diallo, R. Maingi, S. Kubota, A. Sontag, T. Osborne, M. Podestà, R. Bell, B. LeBlanc, J. Menard, and S. Sabbagh, "Dynamical evolution of pedestal parameters in ELMy H-mode in the national spherical torus experiment," *Nuclear Fusion*, vol. 51, no. 10, p. 103031, Sep 2011.

[7] R. Fischer, C. J. Fuchs, B. Kurzan, W. Suttrop, and E. Wolfrum, "Integrated data analysis of profile diagnostics at ASDEX Upgrade," *Fusion Science and Technology*, vol. 58, no. 2, pp. 675–684, 2010.

[8] M. Chilenski, M. Greenwald, Y. Marzouk, N. Howard, A. White, J. Rice, and J. Walk, "Improved profile fitting and quantification of uncertainty in experimental measurements of impurity transport coefficients using Gaussian process regression," *Nuclear Fusion*, vol. 55, no. 2, p. 023012, Jan 2015.

[9] E. Doyle, W. Houlberg, Y. Kamada, V. Mukhovatov, T. Osborne, A. Polevoi, G. Bateman, J. Connor, J. Cordey, T. Fujita, X. Garbet, T. Hahm, L. Horton, A. Hubbard, F. Imbeaux, F. Jenko, J. Kinsey, Y. Kishimoto, J. Li, T. Luce, Y. Martin, M. Ossipenko, V. Parail, A. Peeters, T. Rhodes, J. Rice, C. Roach, V. Rozhansky, F. Ryter, G. Saibene, R. Sartori, A. Sips, J. Snipes, M. Sugihara, E. Synakowski, H. Takenaga, T. Takizuka, K. Thomsen, M. Wade, and H. Wilson, "Chapter 2: Plasma confinement and transport," *Nuclear Fusion*, vol. 47, no. 6, pp. S18–S127, Jun 2007.

[10] A. Mathews, J. Hughes, A. Hubbard, D. Whyte, S. Wolfe, T. Golfinopoulos, D. Brunner, R. Granetz, C. Rea, and A. White, "Confinement regime identification on Alcator C-Mod using supervised machine learning methods," *MIT-PSFC Internal Report*, pp. RR–19–6, Apr 2019.

[11] D. Whyte, A. Hubbard, J. Hughes, B. Lipschultz, J. Rice, E. Marmar, M. Greenwald, I. Cziegler, A. Dominguez, T. Golfinopoulos, N. Howard, L. Lin, R. McDermott, M. Porkolab, M. Reinke, J. Terry, N. Tsujii, S. Wolfe, S. Wukitch, and Y. L. and, "I-mode: an H-mode energy confinement regime with L-mode particle transport in Alcator C-Mod," *Nuclear Fusion*, vol. 50, no. 10, p. 105005, Aug 2010.

[12] C. E. Rasmussen and C. K. I. Williams, *Gaussian Processes for Machine Learning (Adaptive Computation and Machine Learning)*. The MIT Press, 2005.

[13] C. Wang and R. M. Neal, "Gaussian process regression with heteroscedastic or non-Gaussian residuals," 2012.

[14] D. K. Duvenaud, "Automatic model construction with Gaussian processes," Ph.D. dissertation, University of Cambridge, 2014.

[15] M. G. Genton, "Classes of kernels for machine learning: A statistics perspective," *J. Mach. Learn. Res.*, vol. 2, p. 299–312, 2002.

[16] M. Heinonen, H. Mannerström, J. Rousu, S. Kaski, and H. Lähdesmäki, "Non-stationary Gaussian process regression with hamiltonian monte carlo," in *Proceedings of the 19th International Conference on Artificial Intelligence and Statistics*, ser. Proceedings of Machine Learning Research, A. Gretton and C. C. Robert, Eds., vol. 51. Cadiz, Spain: PMLR, May 2016, pp. 732–740.

[17] D. Li, J. Svensson, H. Thomsen, F. Medina, A. Werner, and R. Wolf, "Bayesian soft x-ray tomography using non-stationary Gaussian processes," *Review of Scientific Instruments*, vol. 84, no. 8, p. 083506, 2013.

[18] C. J. Paciorek and M. J. Schervish, "Spatial modelling using a new class of nonstationary covariance functions," *Environmetrics*, vol. 17, no. 5, pp. 483–506, 2006.

[19] C. Plagemann, K. Kersting, and W. Burgard, "Nonstationary Gaussian process regression using point estimates of local smoothness," in *Proceedings of the 2008th European Conference on Machine Learning and Knowledge Discovery in Databases - Volume Part II*, ser. ECMLP-KDD'08. Berlin, Heidelberg: Springer-Verlag, 2008, p. 204–219.

[20] J. W. Hughes, D. Mossessian, K. Zhurovich, M. DeMaria, K. Jensen, and A. Hubbard, "Thomson scattering upgrades on Alcator C-Mod," *Review of Scientific Instruments*, vol. 74, no. 3, pp. 1667–1670, 2003.

[21] D. Silvagni, T. Eich, M. Faitsch, T. Happel, B. Sieglin, P. David, D. Nille, L. Gil, and U. Stroth, "Scrape-off layer (sol) power width scaling and correlation between sol and pedestal gradients across L, I and H-mode plasmas at ASDEX Upgrade," *Plasma Physics and Controlled Fusion*, vol. 62, no. 4, p. 045015, Feb 2020.

[22] A. Mathews, M. Francisquez, J. Hughes, and D. Hatch, "Uncovering edge plasma dynamics via deep learning from partial observations," 2020, arXiv:2009.05005 [physics.plasm-ph].

[23] M. Tso, "Multivariate statistical methods," 2010, University of Manchester.

Abhilash Mathews is currently a Ph.D. candidate in the Department of Nuclear Science and Engineering, Massachusetts Institute of Technology, Cambridge, MA, USA.

Jerry W. Hughes is a Principal Research Scientist at the MIT Plasma Science and Fusion Center in Cambridge, MA, USA.