

RESEARCH ARTICLE | MARCH 15 2021

Plasma parameter profile inference from limited data utilizing second-order derivative priors and physic-based constraints ^{EP}

T. Nishizawa  ; M. Cavedon ; R. Dux; F. Reimold; U. von Toussaint; ASDEX Upgrade Team



Phys. Plasmas 28, 032504 (2021)

<https://doi.org/10.1063/5.0039011>



Physics of Plasmas

Features in Plasma Physics Webinars

Register Today!

Plasma parameter profile inference from limited data utilizing second-order derivative priors and physic-based constraints

Cite as: Phys. Plasmas **28**, 032504 (2021); doi: [10.1063/5.0039011](https://doi.org/10.1063/5.0039011)

Submitted: 1 December 2020 · Accepted: 21 February 2021 ·

Published Online: 15 March 2021



View Online



Export Citation



CrossMark

T. Nishizawa,^{1,a)}  M. Cavedon,¹  R. Dux,¹ F. Reimold,² U. von Toussaint,¹ and ASDEX Upgrade Team^{b)}

AFFILIATIONS

¹Max Planck Institute for Plasma Physics, Boltzmannstr. 2, 85748 Garching, Germany

²Max Planck Institute for Plasma Physics, Wendelsteinstr. 1, 17491 Greifswald, Germany

^{a)}Author to whom correspondence should be addressed: takashi.nishizawa@ipp.mpg.de

^{b)}See Meyer *et al.* (2019) for the ASDEX Upgrade team (<https://doi.org/10.1088/1741-4326/ab18b8>).

ABSTRACT

A Bayesian framework has been used to improve the quality of inferred plasma parameter profiles. An integrated data analysis allows for coherent combinations of different diagnostics, and Gaussian process regression provides a reliable regularization process and systematic uncertainty estimation. In this paper, we propose a new profile inference framework that utilizes our prior knowledge about plasma physics, along with integrated data analysis and a Gaussian process. In order to facilitate the use of the Markov chain Monte Carlo sampling, we use a Gaussian process to define quantities corresponding to the second derivatives of the profiles. We validate the analysis technique by using a synthetic one-dimensional plasma, in which the transport properties are known and demonstrate that the proposed analysis technique can infer plasma parameter profiles from line-integrated measurements only. Furthermore, we can even infer unknown parameters in our physics models when our physics knowledge on the system is incomplete. This analysis framework is applicable to laboratory plasmas and provides a means to investigate plasma parameters, to which standard diagnostics are not directly sensitive.

© 2021 Author(s). All article content, except where otherwise noted, is licensed under a Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>). <https://doi.org/10.1063/5.0039011>

I. INTRODUCTION

Reliable measurements of plasma parameter profiles are essential for the advancement of plasma physics. Since important properties of plasmas such as transport or instability onsets often depend on the derivatives of plasma parameters,^{1,2} even small errors in the profile measurements cannot be tolerated under some circumstances. To this end, many diagnostic techniques have been developed to characterize the spatial variation of plasma parameters. In modern devices, the same plasma parameters are often measured by multiple diagnostics.^{3,4} While this is desirable, different diagnostics do not necessarily provide the same profile measurements. In addition, each result usually depends on an analysis method, such as the choice of a fitting function. Formulating an analysis framework that systematically processes multiple diagnostics is an important topic of research along with diagnostic hardware improvement.

Integrated data analysis (IDA)^{3,5} has been introduced to the profile measurements to combine different diagnostics and provide a coherent result. IDA is based on Bayes' theorem,

$$p(\theta|\mathbf{d}, \mathbf{I}) = \frac{p(\mathbf{d}|\theta, \mathbf{I})p(\theta|\mathbf{I})}{p(\mathbf{d}|\mathbf{I})}, \quad (1)$$

where \mathbf{I} is the prior knowledge, θ is a parameter to be inferred, and \mathbf{d} is the set of measurements. The prior $p(\theta|\mathbf{I})$ defines the probability distribution of θ before measurements are made. $p(\theta|\mathbf{d}, \mathbf{I})$ and $p(\mathbf{d}|\theta, \mathbf{I})$ are called the posterior and the likelihood, respectively. $p(\mathbf{d}|\mathbf{I})$ is the marginal likelihood. Since $p(\theta|\mathbf{d}', \mathbf{d}, \mathbf{I}) \propto p(\mathbf{d}'|\theta, \mathbf{I})p(\theta|\mathbf{d}, \mathbf{I})$ holds for another measurements \mathbf{d}' , we can integrate any number of diagnostics to obtain a more informed posterior. The Bayesian framework also allows for a systematic treatment of uncertainties.^{4,6,7} All possible sources of uncertainties such as calibration errors can be implemented into a forward model.

Another important advance in profile evaluation is made by Gaussian process (GP) regression,^{8–13} which can also be formulated in the Bayesian framework. In conventional profile fitting, uncertainties are estimated either by a covariance matrix or by Monte Carlo sampling.^{14–16} In contrast, Bayes' theorem Eq. (1) defines the probability

distribution of the parameters of interest for a given \mathbf{d} as the product of conditional probabilities. A GP can provide a family of smooth lines that can be used as the prior distribution. This approach allows for systematic uncertainty estimation without introducing a specific fitting function. Recently, profile measurements by using both IDA and a GP have been reported.⁴

In principle, Bayes' theorem states that profiles can be inferred for a given prior distribution even when the likelihood does not provide useful information. It simply means that the choice of a prior will be reflected to a large degree in the inferred profiles and uncertainties in such a case. Measurements do not even need to be local. In fact, line-integrated electron density measured by interferometer is often combined with local electron density measurements in IDA.^{3,6} Signals that depend on multiple plasma parameters can also be useful in the Bayesian framework. For example, Ref. 17 reports two-dimensional profile measurements of electron density, electron temperature, and neutral density using mainly Balmer line intensities.

In this paper, we further expand the analysis framework based on IDA and a GP to cope with cases where our diagnostic capabilities are limited, e.g., no local measurements are available. A GP constructs a prior distribution of lines (or surfaces for two-dimensional cases) that are compatible with the possible properties of profiles such as amplitudes and correlation lengths. In addition to that prior knowledge, we also utilize constraints based on plasma physics to devise an even more informative prior.

Even when local plasma parameters are measured with high spatial resolution, we still need to interpolate between discrete data points. Thus, the number of parameters representing a profile is usually larger than the number of constraints given by measurements. This situation is one example of ill-posed problems. For ill-posed problems, a procedure called "regularization" is required to determine optimum values for all parameters. In fact, profile fitting using a model function and GP regression correspond to different regularization procedures. Regularization is discussed in more detail in Sec. II. Following Sec. II, this paper discusses the new profile inference technique by using a one-dimensional synthetic plasma while the analysis formalism presented herein is general and applicable to laboratory plasmas as well. Section III defines a physics-based prior belief by using a particle conservation condition. In Sec. IV, we construct a prior, second-order-derivative prior, by using a GP to ensure that the prior satisfies the expected properties of the plasma parameter profiles. In Sec. V, we explain techniques to facilitate computation for a high-dimensional probability distribution. We attempt profile inference for electron density, electron temperature, and neutral density by using only line-integrated measurements in Sec. VI. Then, we discuss a case where our physics knowledge is incomplete in Sec. VII. Also, the capabilities of the proposed inference framework when the plasma is well-diagnosed are investigated. Finally, a summary and conclusions are given in Sec. VIII.

We use many symbols in this paper. To avoid confusion, the symbols are summarized in Nomenclature. Bold letters represent vectors, and a plasma parameter written in a bold letter means a profile of that quantity, e.g., \mathbf{n}_e stands for the profile of the electron density n_e .

II. REGULARIZATION USING PRIOR BELIEFS

In this section, we review ill-posed problems and the regularization procedure in the Bayesian context. When the number of

unknown parameters is greater than the number of conditions, a problem is ill posed. As an example of such a problem, we consider a case where we would like to determine three parameters, $\boldsymbol{\theta} = [\theta_1, \theta_2, \theta_3]$ from two data points, $\mathbf{d} = [d_1, d_2]$. The elements of $\boldsymbol{\theta}$ are independent of each other. We know that both elements of \mathbf{d} depend on all elements of $\boldsymbol{\theta}$ in different fashions, and a reliable forward model is available. When the errors σ_x , $x = 1, 2$ in the measurements follow a normal distribution, their likelihood is given by

$$p(\mathbf{d}|\boldsymbol{\theta}) \propto \exp \left(- \sum_{x=1,2} \frac{(d_x - d_x^{\text{model}}(\boldsymbol{\theta}))^2}{2\sigma_x^2} \right). \quad (2)$$

In this case, maximum likelihood estimation (MLE) does not give a single point but a continuous line in the $\boldsymbol{\theta}$ space. Conceptually, $p(\mathbf{d}|\boldsymbol{\theta})$ looks like Fig. 1. Note that the likelihood function shown in Fig. 1 is solely for an illustration purpose, and we are not considering a specific form of Eq. (2). Due to the uncertainties of \mathbf{d} , the probability density spreads out from the line of maximum likelihood. When the measurements become more precise, i.e., σ_x decreases, the probability density will localize nearer the line of maximum likelihood. However, we can never determine a single optimum point in the $\boldsymbol{\theta}$ space due to the insufficient conditions.

We can categorize regularization procedures roughly into two groups. The first group reduces the degrees of freedom by introducing some functional model.^{16,18} In the case of Fig. 1, this procedure corresponds to drawing a line or a surface in the $\boldsymbol{\theta}$ space. If we parametrize the model by one parameter, we draw a line. If we use two parameters, we obtain a surface. All the volume that is not on the line or the surface will be excluded from the posterior distribution, no matter how high their likelihood is, or how close they are to the optimum point given by the functional model in the original $\boldsymbol{\theta}$ space.

The second group of the regularization procedure uses a penalty term instead of a specific model. For this reason, they are called model-free regularization. This approach maximizes the following function:

$$F(\boldsymbol{\theta}) = - \sum_{x=1,2} \frac{(d_x - d_x^{\text{model}}(\boldsymbol{\theta}))^2}{2\sigma_x^2} - \hat{\lambda} \tilde{N}(\boldsymbol{\theta}), \quad \hat{\lambda} > 0. \quad (3)$$

The first term on the RHS of Eq. (3) is a log-likelihood plus some constant. The second term represents a penalty term. Many types of functions are used for $\tilde{N}(\boldsymbol{\theta})$, e.g., maximum entropy, Fisher information, etc.^{19,20} When $\boldsymbol{\theta}$ represents a plasma parameter profile, it is reasonable to employ $\tilde{N}(\boldsymbol{\theta})$ that reduces penalty for smooth profiles. $\hat{\lambda}$ is called a hyper-parameter and determines relative weights between the likelihood and the penalization. In the Bayesian context, the penalty term is

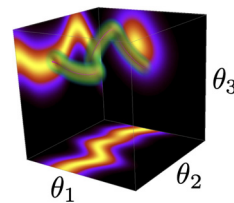


FIG. 1. An example of likelihood function for an ill-posed problem. Marginal probability distributions are projected onto the θ_1 - θ_2 , θ_2 - θ_3 , and θ_3 - θ_1 planes.

replaced by the prior.^{12,17} We obtain an identical form of Eq. (3) by rewriting Bayes' theorem Eq. (1) in the logarithmic form:

$$\ln p(\theta|\mathbf{d}, \mathbf{I}) = \ln p(\mathbf{d}|\theta, \mathbf{I}) + \ln p(\theta|\mathbf{I}) - \ln p(\mathbf{d}|\mathbf{I}). \quad (4)$$

The first term on the RHS of Eq. (3) corresponds to the likelihood, while the second term corresponds to the prior in the Bayesian context. The penalty term in Eq. (3) is identical to a prior $\propto \exp(-\hat{\lambda}\tilde{N}(\theta))$. The marginal likelihood term $\ln p(\mathbf{d}|\mathbf{I})$ is independent of θ . Thus, θ that maximizes $F(\theta)$ has the highest posterior probability, i.e., a maximum a posterior probability estimate (MAP).

Assuming that θ_1 , θ_2 , and θ_3 represent quantities at spatially adjacent points, we impose a smoothness condition by using a prior:

$$p(\theta|\mathbf{I}) \propto \exp\left(-\hat{\lambda} \sum_{i=1,2} (\theta_i - \theta_{i+1})^2\right). \quad (5)$$

Figure 2 shows the probability distribution of Eq. (5) in the θ space. The probability is highest on the line $\theta_1 = \theta_2 = \theta_3$. As $\hat{\lambda}$ increases, the probability density becomes more localized near this line.

In a model-free regularization in the Bayesian context, we carry out regularization by multiplying the likelihood Eq. (2) and the prior Eq. (5). Posterior distributions with different $\hat{\lambda}$ are shown in Fig. 3. Since the prior distribution also contributes to the posterior distribution, a finite number of optimum points can be determined in the θ space. As $\hat{\lambda}$ increases, MAP and the high probability density region move from the line of maximum likelihood to the line of $\theta_1 = \theta_2 = \theta_3$. Note that points on the maximum likelihood line are not necessarily seen as optimum since the reduced chi-square becomes zero there. This is an indication of over-fitting. However, if we increase $\hat{\lambda}$ too much, MAP and the high probability density region become inconsistent with the measurements, i.e., over-penalization. In imaging problems, MAP is often chosen as an optimum point, and uncertainties are sometimes of secondary concern. Many criteria for optimizing hyper-parameters have been proposed with regard to these cases.²¹ However, we calculate the posterior probability for uncertainty evaluation in the proposed inference framework. In order to ensure that the posterior is distributed near maximum likelihood, we evaluate normalized residual $(d_x - d_x^{\text{model}})/\sigma_x$. If the posterior distribution is consistent with the measurements, $(d_x - d_x^{\text{model}})/\sigma_x$ distributes near zero with a characteristic deviation of unity. In Secs. III and IV, we introduce a prior distribution based on several prior beliefs to select an optimum volume in the parameter space. We will see that this prior has some specific factors to consider when determining hyper-parameters.

III. PHYSICS-BASED PRIOR BELIEFS

Here, we incorporate our knowledge of plasma physics into a probability distribution. In profile inference, many parameters are

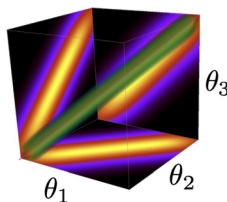


FIG. 2. Prior distribution constructed from the penalty Eq. (5). Marginal probability distributions are projected onto the θ_1 - θ_2 , θ_2 - θ_3 , and θ_3 - θ_1 planes.

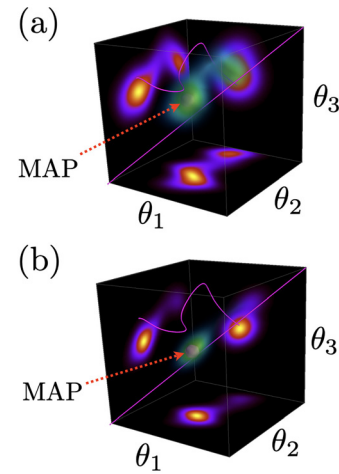


FIG. 3. Posterior distributions with small $\hat{\lambda}$ (a) and with large $\hat{\lambda}$ (b). Red arrows point at a maximum a posterior probability estimate (MAP) shown in a sphere. The line of maximum likelihood and $\theta_1 = \theta_2 = \theta_3$ are shown in purple lines.

required to express spatially varying quantities, and localizing the posterior distribution becomes far more challenging than the three-dimensional case discussed in Sec. II. When diagnostic capabilities are limited, a smoothness condition alone may no longer be sufficient to provide a useful posterior. Fortunately, our available knowledge about the plasma extends beyond the requirement of reasonably smooth profiles. Depending on the time and spatial ranges of interest, there are many equations that describe the plasma behaviors with satisfactory accuracy.

As demonstration and validation, we consider a one-dimensional system composed of a hydrogen plasma and neutrals (hydrogen atoms) for the rest of the paper. We assume that the plasma parameters are symmetric around $L = 0$, where L is the spatial coordinate. Hydrogen molecules and impurities are neglected for simplicity. We further assume that the electron transport is diffusive, and no convective transport exists in this system. Thus, the electron particle conservation equation is

$$\frac{\partial n_e}{\partial t} - \frac{\partial}{\partial L} \left(D_{n_e} \frac{\partial n_e}{\partial L} \right) = S(n_e, n_0, T_e), \quad (6)$$

where n_e , n_0 , and T_e are the electron density, the neutral density, and the electron temperature, respectively. D_{n_e} stands for the diffusion coefficient for n_e , and S is the source. The explicit form of S is

$$S(n_e, n_0, T_e) = n_e n_0 SCD(n_e, T_e) - n_e^2 ACD(n_e, T_e), \quad (7)$$

where SCD and ACD are an effective ionization rate and an effective recombination rate, respectively. Atomic database ADAS²² provides these rates. If we treat the system as quasi-stationary over the time interval of δt ,

$$\frac{\left(\frac{\partial}{\partial L} \left(D_{n_e} \frac{\partial n_e}{\partial L} \right) + S \right) \delta t}{n_e} \ll 1 \quad (8)$$

is satisfied everywhere along L . We implement this condition through the following probability distribution:

$$P(I_{pb}|\mathbf{n}_e, \mathbf{n}_0, T_e) \propto \exp \left(- \int_{L_{\min}}^{L_{\max}} \frac{\left(\frac{\partial}{\partial L} \left(D_{n_e} \frac{\partial}{\partial L} n_e \right) + S \right)^2}{2n_e^2 \hat{\tau}^{-2}} dL \right), \quad (9)$$

where I_{pb} represents the prior knowledge based on physical laws in the system, i.e., how close (or far) a given profile is to physically reasonable profiles. The vector quantities \mathbf{n}_e , \mathbf{n}_0 , and T_e represent the profiles of each plasma parameter. $\hat{\tau}$ and \hat{L}_0 are hyper-parameters. When \hat{L}_0 is the correlation length of the system, $\hat{\tau}$ becomes a characteristic time that corresponds to the standard deviation of the normalized particle conservation condition given by Eq. (8). Note that Eq. (9) is the probability of I_{pb} under the condition $\mathbf{n}_e \cap \mathbf{n}_0 \cap T_e$, not the probability of $\mathbf{n}_e \cap \mathbf{n}_0 \cap T_e$ under the condition I_{pb} . When a prior belief I_1 is independent of another prior belief I_2 , $p(I_1|\theta)p(\theta|I_2) \propto p(\theta|I_1, I_2)$ holds. Thus, a new prior with the physics-based prior belief I_{pb} can be formed by multiplying Eq. (9) by another prior based on other beliefs. As $\hat{\tau}$ becomes larger for a fixed \hat{L}_0 , the relative weight of $P(I_{pb}|\mathbf{n}_e, \mathbf{n}_0, T_e)$ with respect to the likelihood and other probability distributions in the prior increases. In the case of a simple smoothness condition such as Eq. (5), the prior probability becomes maximum for flat profiles, which are not physically reasonable. In contrast, $P(I_{pb}|\mathbf{n}_e, \mathbf{n}_0, T_e)$ peaks near optimum points. Thus, the physics-based prior belief allows us to localize the posterior distribution to a larger degree without causing over-penalization. Even when the transport properties of the system are not exactly known, an approximate form of particle conservation equations can still be used to give a physics-based probability distribution such as Eq. (9), whose peaks are closer to the optimum point than the ones based on simple smoothness conditions. In addition, we can treat a transport coefficient such as D_{n_e} as an additional unknown and perform an inference. An example of this procedure will be discussed in Sec. VII.

IV. SECOND-ORDER DERIVATIVE PRIORS

In this section, we introduce a prior using a GP. Unlike the physics-based prior belief, this prior is intended to remove unphysical volume from the prior distribution while keeping reasonably high probabilities for all sensible profiles. We utilize the following prior beliefs:

- \mathbf{n}_e and T_e increase monotonically from the edges to $L = 0$.
- \mathbf{n}_0 decreases monotonically from the edges to $L = 0$.
- \mathbf{n}_e , \mathbf{n}_0 , and T_e are reasonably smooth.
- \mathbf{n}_e , \mathbf{n}_0 , and T_e are symmetric about $L = 0$.
- The first derivatives of \mathbf{n}_e , \mathbf{n}_0 , and T_e are zero at $L = 0$.

While there are some exceptions,^{23,24} monotonicity is often a fair assumption for \mathbf{n}_e , T_e , and \mathbf{n}_0 . Without the monotonicity conditions, the profile shape of the prior has too much freedom, and obtaining useful inference results becomes difficult with limited diagnostic capabilities. Equation (9) imposes smoothing constraints on \mathbf{n}_e when S is continuous. But, nothing ensures the smoothness of \mathbf{n}_0 and T_e . We prescribe smoothness conditions for all profiles through the additional prior beliefs. Profiles with well-defined derivatives up to higher orders can be drawn from a GP:¹⁰

$$p(\theta|\Sigma, \mu) \propto \exp \left(-\frac{1}{2} (\theta - \mu)^T \Sigma^{-1} (\theta - \mu) \right), \quad (10)$$

where Σ and μ are a covariance matrix and a mean, respectively. If we use a squared exponential (SE) covariance function

$$\Sigma_{i,i'} = \hat{\sigma}^2 \exp \left(-\frac{|i - i'|^2}{2\hat{l}^2} \right) \quad (11)$$

for Σ , profiles drawn from this GP are infinitely differentiable. In this specific case, we assume that each element of θ is equally spaced in a one-dimensional space, and we use the index i to represent a position. In a general case, i is replaced by a space vector. $\hat{\sigma}$ and \hat{l} are hyper-parameters. $\hat{\sigma}$ sets the variation amplitude, and \hat{l} determines the spatial correlation length. The prior given by Eq. (5) only evaluates the difference between spatially adjacent points and controls the degree of smoothness by the hyper-parameter $\hat{\lambda}$. On the other hand, through Eq. (11), we can prescribe the spatial correlation length \hat{l} directly, for which we often have good estimate based on the characteristic scale of variation in the plasma parameters.

A GP is especially useful when \mathbf{d} is linear with respect to θ since the posterior distribution can be obtained analytically by using conjugate distributions.^{8,12} However, we are now considering general problems where the likelihood or the prior contains non-conjugate distributions so that any diagnostics can be included within the IDA framework, and a physics-based belief such as Eq. (9) can be used. One severe challenge to this approach is the need for numerically calculating a high-dimensional posterior distribution. In the latter part of this paper, we calculate probability distributions in more than 150 dimensional spaces. Even if we use only 10 points for each dimension, we need to calculate the posterior probability at more than 10^{150} points, which is practically impossible. In order to overcome this difficulty, we use a Markov chain Monte Carlo⁵ (MCMC) method instead of trying to cover the whole volume in the high dimensional space. MCMC randomly samples points in the parameter space in such a way that the distribution of samples converges to the probability distribution we are trying to calculate. Thus, MCMC asymptotically traces out the probability distribution of interest as the number of samples increases. However, the performance of MCMC generally degrades as the dimensionality increases, and most of the MCMC algorithms are not scalable to ~ 150 dimensions when the probability distribution is complex. We employ No-U-Turn Sampler (NUTS) algorithm,²⁵ provided by PyMC3.²⁶ NUTS algorithm is based on Hamiltonian Monte Carlo (HMC) algorithm.²⁷ In most of our relevant problems, HMC has a favorable scaling for dimensionality compared with other MCMC algorithms such as random-walk Metropolis.^{25,28,29} The downside of HMC is that its efficiency strongly depends on user-tunable parameters. NUTS automatically optimizes some of those. In addition, PyMC3 provides several initialization methods for optimizing the mass matrix, which also affects the performance of HMC. Even though we utilize these novel algorithms for high-dimensional inference problems, we find that additional care is required when we try to infer profiles from limited data. In the following, we formulate a prior in such a way that it helps NUTS algorithm run smoothly.

Poor performance of MCMC is often due to inappropriate parametrization. The efficiency can significantly improve if the parametrization is chosen to be amenable to the MCMC algorithm. In Ref. 8, the profile of interest is drawn directly from a GP, e.g., setting $\theta = \mathbf{n}_e$. The posterior distribution can be calculated analytically for this parametrization when \mathbf{d} is linear with respect to the plasma parameter,

and its uncertainties follow a normal distribution. However, when MCMC is used under the monotonicity condition, parametrization such as $\theta = n_e$ leads to a poor performance due to a strong correlations between the elements of θ .³⁰ In order to reduce the correlations, we use a GP to define a line that corresponds to the second derivatives and construct a profile by integrating the original parameters twice. As discussed in Ref. 31, second derivatives can have relatively large freedom even when strong constraints are imposed to the original profiles. Since integrals entail ambiguity, we can impose boundary conditions in the integration process. This parametrization facilitates MCMC sampling also by excluding unwanted profiles from the original parameter space. The following mapping $y_f = \phi_{\mu, \Sigma}^{\text{free}}(\theta_f)$ constructs a profile with a free boundary condition at $l=0$ and a fixed boundary condition at $l=N$ from parameter θ_f drawn from a GP with μ and Σ :

$$\Theta_{f,j} = 1 - \frac{\sum_{i=1}^j \theta_{f,i}}{\sum_{i=1}^N \theta_{f,i}}, \quad j = 1, 2, \dots, N, \quad (12)$$

$$y_{f,l} = \begin{cases} 0, & l = 0 \\ \sum_{j=1}^l \Theta_{f,j} / \sum_{j=1}^N \Theta_{f,j}, & l = 1, 2, \dots, N, \end{cases}$$

where $f = n_e, n_0$, and T_e . $y_{f,l}$ changes from 0 to 1 by definition. Similarly, we define a mapping $y_f = \phi_{\mu, \Sigma}^{\text{fix}}(\theta_f)$ that imposes a fixed boundary condition at both ends as follows:

$$\Theta_{f,j} = \begin{cases} 0, & j = 0 \\ \sum_{i=1}^j \theta_{f,i} - \frac{j}{N} \sum_{i=1}^N \theta_{f,i}, & j = 1, 2, \dots, N, \end{cases} \quad (13)$$

$$y_{f,l} = \begin{cases} 0, & l = 0 \\ \sum_{j=0}^{l-1} \Theta_{f,j} / \sum_{j=0}^N \Theta_{f,j}, & l = 1, 2, \dots, N+1. \end{cases}$$

Also, $y_{f,0} = 0$ and $y_{N+1,l} = 1$ by definition. The distributions of $y_{f,l}$ depend on μ and Σ . Technically, $y_{f,l}$ cannot be defined when the denominator in Eq. (12) or in Eq. (13) is zero. In practice, however, we have not encountered θ_f that does not lead to finite y_f when running MCMC for the inference problems in this paper. We set the mean of each normal distribution to zero, i.e., $\mu = \mathbf{0}$. For Σ , we use a non-stationary SE with

$$\hat{\sigma}^2 = \hat{\sigma}_i \hat{\sigma}_{\bar{i}}, \quad (14)$$

where

$$\hat{\sigma}_i = i/N + 0.1. \quad (15)$$

This choice is based on our prior belief that the variation in the second derivative of the profile increases as it approaches the peak. We further assume that spatial correlation length is 15% of the distance from the edge to $L=0$, which gives $\hat{l} = 0.15N$. In principle, optimum values exist for those hyper-parameters. However, the optimization requires the evaluation of the marginal likelihood,^{10,12,32} which is difficult to calculate for high-dimensional problems when the posterior

distribution cannot be derived analytically. In addition, the function form of Σ is also subject to an optimization problem. Many other covariance functions other than SE are also proposed.¹⁰ Note that those hyper-parameters control the prior distribution as $\hat{\lambda}$ does in Fig. 2, and we are not introducing any specific functional models. Unlike $\hat{\lambda}$, SE allows us to choose \hat{l} and $\hat{\sigma}^2$ based on their physical meanings. Optimizing hyper-parameters in a systematic fashion remains future work.

Figure 4 shows original samples taken from the GP, the resulting profiles, and their first derivatives. We used the NUTS sampler to create those plots. Monotonicity is imposed by adding a penalty term; the details of which will be discussed in Sec. V. Due to the mapping, θ_f still relatively freely explores the original parameter space under the monotonicity condition. MCMC samples favorably either from positive or negative values of θ_f depending on the initialization. However, this does not affect the resulting y_f since it is independent of the sign of θ_f due to the normalization in Eqs. (12) and (13). In order to illustrate how the hyper-parameters affect the prior distribution, we show profiles with steep gradients in the edge in Fig. 5(b). The hyper-parameters $\hat{\sigma}$ and \hat{l} used for those samples are given in the Appendix. While the original parameter distribution shown in Fig. 5(a) is largely different from the one shown in Fig. 4(a), they share similar resulting profiles. In general, the posterior distribution is relatively insensitive to small changes in hyper-parameters.

The mapping processes discussed above only offer normalized profile shapes. We define plasma parameter profiles from the edge to the peak as follows:

$$f = (f_{\text{peak}} - f_{\text{edge}})y_f + f_{\text{edge}}\mathbf{1}, \quad f = n_e, n_0, T_e, \quad (16)$$

where f_{peak} and f_{edge} are the values of a plasma parameter at the peak and at the edge, respectively. In the following discussion, we assume that n_e , n_0 , and T_e at the boundary are known. Therefore, we need only one additional parameter f_{peak} to define the whole profile f from y_f . The prior distribution introduced in this section is

$$P(f|I_{\text{mo}}, \phi_{\mu, \Sigma}^{\text{free}} f_{\text{peak}})P(f_{\text{peak}}) \propto P(I_{\text{mo}}|f)P(f|\phi_{\mu, \Sigma}^{\text{free}} f_{\text{peak}})P(f_{\text{peak}}), \quad (17)$$

where $P(I_{\text{mo}}|f)$ penalizes profiles of f , which are not monotonic. We introduce the explicit expression for this probability distribution in Sec. V. Similarly, we define a prior distribution with a fixed boundary condition at the boundary using $\phi_{\mu, \Sigma}^{\text{fix}}$. We choose a uniform distribution for $P(f_{\text{peak}})$. This parametrization removes unphysical negative values from the prior distribution and avoids introducing a bias in amplitude. Since the second derivatives of the resulting profiles and the original parameters share the same relative shape, we name the prior given by the above definition a second-order derivative (SOD) prior.

V. ASYMPTOTIC POTENTIAL AND MONOTONICITY PENALTY

To ensure robust performance of MCMC sampling, we introduce modified probability functions. We also use this functional form for imposing monotonicity on the profiles. When the original forms of the probabilities are used, Markov chains sometimes get trapped at local maxima of probability, and MCMC fails for the inference problems discussed in Secs. VI and VII. In order to help Markov chains reach a typical set, we simplify the volume of a probability distribution

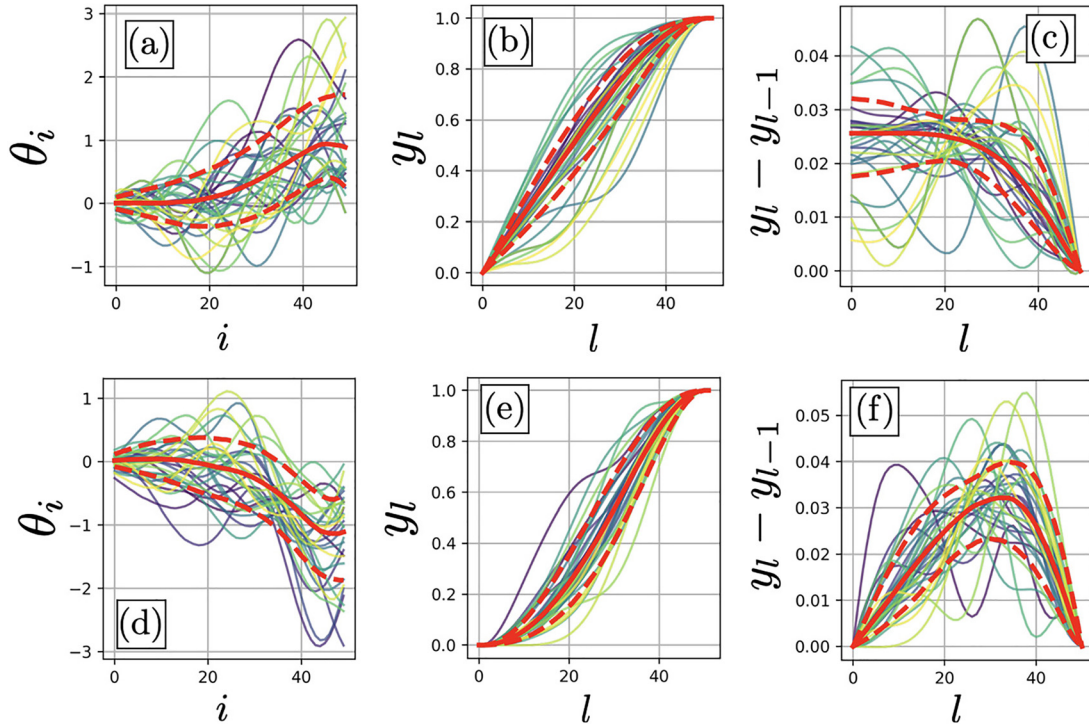


FIG. 4. Original samples (a) and (d), corresponding profiles (b) and (e), and their derivatives (c) and (f) drawn from the Gaussian process. Equation (12) is used for (a)–(c), while Eq. (13) is for (d)–(f). $N = 50$ and $\hat{l} = 7.5$ for both cases. The red-dashed lines represent the 16th and 84th percentiles at each spatial point. The red solid lines represent the 50th percentiles at each spatial point. Lines with the same color correspond to the same sample. The percentiles are calculated from 10,000 samples.

with extremely low probabilities by rewriting a Gaussian-type probability $P(z) \propto \exp(-z^2/2)$ in an asymptotic form:

$$\tilde{P}(z) \propto \exp\left(-\frac{|z|^2}{2} \cdot \frac{1 + uv^2|z|}{1 + v^2|z|^2/2}\right), \quad (18)$$

where z is a quantity to be minimized, e.g., $z = |d_i - d_i^{\text{model}}|/\sigma_i$. Note that z is normalized by the standard deviation. We use $u = 0.5$ and $v = 0.1$ for Eq. (9), and $u = 5$ and $v = 0.5$ for the likelihood since these

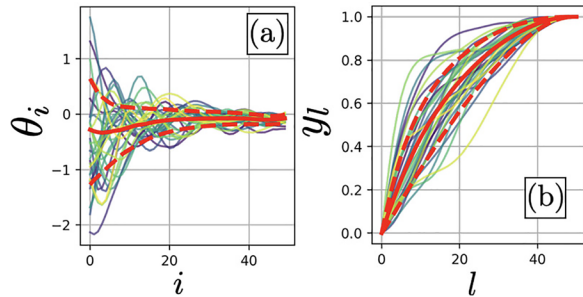


FIG. 5. Original samples (a) and corresponding profiles (b) drawn from the Gaussian process defined in Appendix. Equation (12) is used. The red-dashed lines represent the 16th and 84th percentiles at each spatial point. The red solid lines represent the 50th percentiles at each spatial point. The percentiles are calculated from 10 000 samples.

choices offer stable MCMC sampling for the problems discussed in this paper. These values are found heuristically by trying many combinations of numbers. Figure 6 shows the probability density of \tilde{P} and its log probability for $u = 0.5$ and $v = 0.1$. As $|z|$ increases, $\ln \tilde{P}(z)$ asymptotically approaches $-u|z| + \text{const}$ while $\tilde{P}(z) \approx P(z)$ for $|z| \ll 1/v$. When we evaluate volume outside the typical set where the probability is extremely low, the gradient of the log probability in terms of $|z|$ becomes $-u$ (constant). In contrast, this gradient is $-|z|$ (function of z) for the original Gauss probability. Since HMC calculates the trajectory of a Markov chain by using the gradient of the log probability, the asymptotic form reduces the chance of local trapping. Once a burn-in period completes, MCMC samples from the region where $\tilde{P} \approx P$ holds.

As discussed in Sec. IV, we exclude non-monotonic profiles from the prior belief. We also implement the asymptotic behavior in the penalty term for monotonicity. First, we calculate the following quantity for each profile:

$$w_{f,l} = \frac{0.5 \left(\left| \frac{df_l}{dl} \right| \pm \frac{df_l}{dl} \right)}{\arg\max_{k \in A} \left| \frac{df_k}{dl} \right|}, \quad \text{where } A = [2, 3, \dots, N]. \quad (19)$$

This way, undesired slopes are normalized by the maximum absolute value of the derivative. $w_{f,l}$ is zero if the derivative has the desired sign. We use the minus sign for $f = n_e, T_e$ and the plus sign for $f = n_0$.

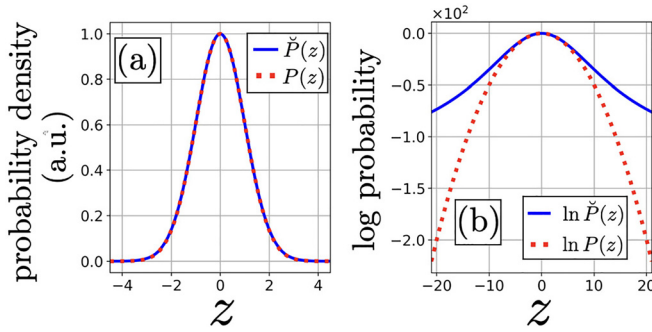


FIG. 6. Comparison between the asymptotic potential $\tilde{P}(z)$ given by Eq. (18) and the original form $P(z)$. (a) Probability densities. (b) Log probabilities. $u=0.5$ and $v=0.1$.

Monotonicity of the profiles is imposed by the following probability distribution:

$$P(I_{mo}|\mathbf{f}) \propto \exp\left(-\sum_{l=2}^N \frac{|w_{f,l}|^2}{2} \cdot \frac{1+uv^2|w_{f,l}|}{1+v^2|w_{f,l}|^2/2}\right), \quad (20)$$

where I_{mo} represents the prior belief about the monotonicity of the profiles. We choose 100 for both u and v in Eq. (20).

The combination of the NUTS sampler, SOD priors, and the asymptotical forms of the probabilities leads to computation time <48 hr for each inference problem discussed in this paper using an Intel Xeon E5-2630 v4 (20 threads, 2.2 GHz).

VI. INFERENCE OF ELECTRON DENSITY, NEUTRAL DENSITY, AND ELECTRON TEMPERATURE PROFILES FROM LINE-INTEGRATED MEASUREMENTS

In this section, we apply the physics-based prior belief and SOD priors to a synthetic one-dimensional system consisting of a hydrogen plasma and neutral (hydrogen atoms) and attempt simultaneous inferences of n_e , n_0 , and T_e from line-integrated measurements only. We use the parameters shown in Table I to define the plasma to be measured. The spatial coordinate L ranges from -10 to $+10$ cm. The input T_e profile is defined by a Gaussian function. Quasi-steady-state profiles of n_e and n_0 are obtained by numerically solving transport equations of n_e and n_0 , assuming that the transport of n_e and n_0 can be described by a diffusive process with the diffusion coefficients given in Table I.

TABLE I. List of parameters used for the synthetic plasma. The units of L and T_e are cm and eV, respectively. $n_{e,edge}$ and $n_{0,edge}$ are the electron density and the neutral density at $L = \pm 10$ cm.

Quantity	Value
$T_e(L)$	$5 \exp\left(-\frac{L^2}{2 \cdot 3^2}\right) + 0.11$ (eV)
$n_{e,edge}$	1×10^{18} (m^{-3})
$n_{0,edge}$	1×10^{18} (m^{-3})
D_{n_e}	$1 + T_e$ (m^2/s)
D_{n_0}	100 (m^2/s)

In addition to the edge boundary conditions shown in Table I, a fixed boundary condition is imposed to both n_e and n_0 at $L=0$. We use 100 discrete points to represent continuous profiles, and thus the spatial resolution is $20 \text{ cm} / 100 = 0.2 \text{ cm}$. With this discretization, the obtained n_e profile satisfies $(\frac{d}{dL} D_{n_e} \frac{d}{dL} n_e + S)/n_e < 1/200 \text{ ms}^{-1}$ at every point along L , indicating that Eq. (8) holds up to $\delta t \sim 200 \text{ ms}$. Since the prior knowledge assumes that the plasma parameters are symmetric around $L=0$, a profile of each plasma parameter has 50 degrees of freedom in our inference problem.

We assume that available diagnostics are line-integrated absolute intensities of four Balmer lines $H_\epsilon(n=7 \rightarrow 2)$, $H_\delta(n=6 \rightarrow 2)$, $H_\gamma(n=5 \rightarrow 2)$, and $H_\beta(n=4 \rightarrow 2)$, and a line-integrated electron density $\int n_e dL$. A local intensity of Balmer line $x = \epsilon, \delta, \beta$, or γ is

$$i_x = i_x^{ex} + i_x^{re}, \quad (21)$$

where

$$i_x^{ex} = n_0 n_e \mathcal{P} \mathcal{E} \mathcal{C}_x^{ex}(n_e, T_e), \quad (22)$$

$$i_x^{re} = n_e^2 \mathcal{P} \mathcal{E} \mathcal{C}_x^{re}(n_e, T_e). \quad (23)$$

$\mathcal{P} \mathcal{E} \mathcal{C}$ stands for a photon emissivity coefficient. These are also provided by ADAS.²² i_x^{ex} and i_x^{re} represent excitation emission and recombination emission, respectively. The intensity of one emission process with respect to another varies for each Balmer line. As the principal quantum number of the upper state in the transition decreases, the total emission becomes more dominated by the excitation emission. The profiles of i_x^{ex} and i_x^{re} for the input n_e , n_0 , and T_e are shown in Fig. 7. Due to the different sensitivities to excitation and recombination, multiple Balmer lines are often measured simultaneously to characterize plasma parameters,^{33–35} while probing local quantities through this diagnostic is usually difficult. We neglect other radiation processes for simplicity. We create synthetic input data for the Balmer line intensities by integrating the local emission intensity profiles:

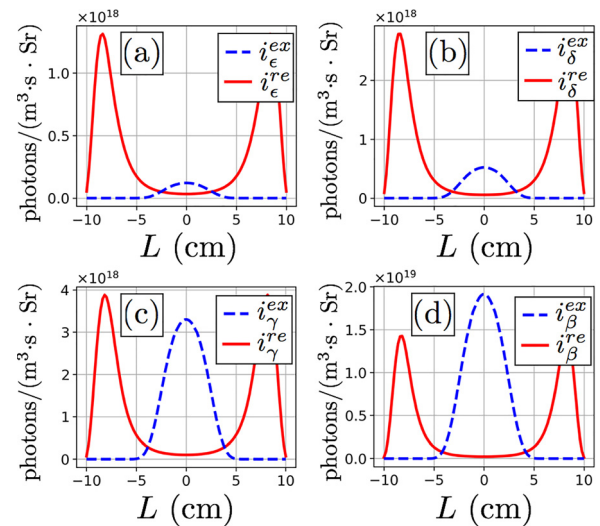


FIG. 7. Profiles of Balmer line emission. (a) H_ϵ , (b) H_δ , (c) H_γ , and (d) H_β . The blue-dashed lines and red solid lines represent excitation emission and recombination emission, respectively.

$$d_{\text{Bal},x} = (1 + \tilde{\epsilon}) \int dL (i_x^{\text{ex}} + i_x^{\text{re}}). \quad (24)$$

We set the uncertainties of this measurement to 5% for all Balmer lines. This is simulated by randomly drawing $\tilde{\epsilon}$ from a normal distribution with $\mu = 0$ and $\sigma = 0.05$. This measurement gives the likelihood as follows:

$$P(d_{\text{Bal}} | \mathbf{n}_e, \mathbf{n}_0, T_e) \propto \exp \left(- \sum_{x=\epsilon, \delta, \gamma, \beta} \frac{(d_{\text{Bal},x} - d_{\text{Bal},x}^{\text{model}})^2}{2\sigma_{\text{Bal},x}^2} \right), \quad (25)$$

where $\sigma_{\text{Bal},x} = 0.05 \times d_{\text{Bal},x}^{\text{model}}$. We calculate the forward-modeled data $d_{\text{Bal},x}^{\text{model}}$ for given \mathbf{n}_e , \mathbf{n}_0 , and T_e by using Eq. (24) with $\tilde{\epsilon} = 0$. Similarly, the likelihood of the line-integrated n_e measurement is given by $P(d_{\text{FIR}} | \mathbf{n}_e) \propto \exp[-(d_{\text{FIR}} - d_{\text{FIR}}^{\text{model}})^2 / 2\sigma_{\text{FIR}}^2]$. The subscript *FIR* represents an infrared interferometer,³⁶ which is a common diagnostic to measure a line-integrated n_e . d_{FIR} is calculated by integrating the input \mathbf{n}_e along L and by adding 5% noise, as is done for the Balmer line intensities. The forward-modeled data $d_{\text{FIR}}^{\text{model}}$ are an integral of \mathbf{n}_e and $\sigma_{\text{FIR}} = 0.05 \times d_{\text{FIR}}^{\text{model}}$. Since we are formulating this analysis technique within the Bayesian framework, different diagnostics can be combined to provide a further informative posterior.³ The posterior is

$$\begin{aligned} P(\mathbf{n}_e, \mathbf{n}_0, T_e | d_{\text{Bal}}, d_{\text{FIR}}, I_{\text{all}}) \\ \propto P(d_{\text{Bal}} | \mathbf{n}_e, \mathbf{n}_0, T_e) P(d_{\text{FIR}} | \mathbf{n}_e) \\ \times P(I_{\text{pb}} | \mathbf{n}_e, \mathbf{n}_0, T_e) \times P(T_e | I_{\text{mo}}, \phi_{\mu,\Sigma}^{\text{fix}}, T_{e,\text{peak}}) P(T_{e,\text{peak}}) \\ \times P(\mathbf{n}_0 | I_{\text{mo}}, \phi_{\mu,\Sigma}^{\text{free}}, n_{0,\text{peak}}) P(n_{0,\text{peak}}) \\ \times P(\mathbf{n}_e | I_{\text{mo}}, \phi_{\mu,\Sigma}^{\text{free}}, n_{e,\text{peak}}) P(n_{e,\text{peak}}), \end{aligned} \quad (26)$$

where I_{all} represents all prior beliefs used in the inference. We use a free boundary condition at the edges for \mathbf{n}_e and \mathbf{n}_0 , whereas we choose a fixed edge boundary condition for T_e . When a free boundary condition is imposed on T_e at the edges, MCMC behaves poorly, and the typical set is not explored. When this inference framework is applied to plasmas in reality, appropriate boundary conditions need to be chosen based on prior knowledge or by trying out multiple combinations of the boundary conditions. In principle, we can also incorporate the neutral particle conservation equation into the prior as is done for n_e in Sec. III since the diffusive transport is assumed for n_0 when calculating the input profiles. However, we utilize the electron particle conservation only for this problem. While a fluid-neutral model is proposed,³⁷ neutral transport in laboratory plasmas is a complex process, which is typically modeled by Monte Carlo codes.³⁸ Defining a probability distribution such as Eq. (9) is difficult for n_0 .

The results of the three different inferences are shown in Fig. 8. The hyper-parameter $\hat{\tau}$ defined in Eq. (9) is set to 1 ms for Figs. 8(a), 8(b), and 8(c), whereas $\hat{\tau} = 10$ ms for Figs. 8(d), 8(e), and 8(f). The other hyper-parameter \hat{L}_0 in Eq. (9) is set equal to \hat{L} ($= 1.5$ cm). When we compare the posterior distributions of the n_e profile shown in Figs. 8(a) and 8(d), the larger $\hat{\tau}$ leads to smaller uncertainties and a better agreement with the input profile. However, Figs. 8(b), 8(c), 8(e), and 8(f) show that the input n_0 and T_e profiles are not within the 16th and 84th percentiles in the case of $\hat{\tau} = 10$ ms, while the inference with $\hat{\tau} = 1$ ms provides reasonable agreements within the uncertainties. In this inference problem, $\hat{\tau} = 10$ ms causes over-penalization. The

optimum values for the hyper-parameters in Eq. (9) depend on the balance between the prior and the likelihood. In Sec. VII, we will see that $\hat{\tau} = 10$ ms successfully localizes the posterior distribution without over-penalization when more diagnostics are available. Unlike the over-penalization due to the smoothing condition shown in Fig. 3(b), the inferred profiles can still be consistent with the measurements when too much weight is given to the physics-based prior belief. Figure 9 shows the distributions of the residual $(d_{\text{Bal},x} - d_{\text{Bal},x}^{\text{model}}) / \sigma_{\text{Bal},x}$ and $(d_{\text{FIR}} - d_{\text{FIR}}^{\text{model}}) / \sigma_{\text{FIR}}$ for the posterior distributions with $\hat{\tau} = 1$ ms and $\hat{\tau} = 10$ ms. The residuals are distributed within the ranges expected of the measurement uncertainties. Note that both probabilities and volume elements in the parameter space contribute to marginal posterior distributions.³⁹ Even though the posterior probability is high for the input profiles, the peaks of the marginal posterior distributions can be away from the inputs if the volume elements around them are small. As discussed in Sec. IV, systematically optimizing hyper-parameters is not within the scope of this paper. In this inference framework, we can only suggest heuristic approaches by considering the balance between the likelihood and the prior.

Figures 8(g), 8(h), and 8(i) show the posterior distribution of each plasma parameter when $P(I_{\text{pb}} | \mathbf{n}_e, \mathbf{n}_0, T_e)$ is not used. The inference leads to wider posterior distributions in this case, indicating that our physics-based prior belief helps constrain the posterior distributions.

VII. SIMULTANEOUS INFERENCE OF ELECTRON DENSITY, NEUTRAL DENSITY, ELECTRON TEMPERATURE, AND DIFFUSION COEFFICIENT PROFILES

Here, we will demonstrate that the physics-based prior belief improves the plasma parameter inference even when the transport properties are partially available. In addition, if we have ample diagnostic capabilities for n_e and T_e , well-localized posterior distributions of \mathbf{n}_0 and unknowns in our transport model can be obtained.

We assume that the transport of n_e is still known to be diffusive, and the profile of the electron particle diffusion coefficient is given by

$$D_{n_e} = a_{D_{n_e}} \mathbf{I} + b_{D_{n_e}} T_e. \quad (27)$$

However, we do not know the exact values of the coefficients $a_{D_{n_e}}$ and $b_{D_{n_e}}$. We treat them as additional parameters of interest. Thus, the posterior distribution is given by

$$\begin{aligned} P(\mathbf{n}_e, \mathbf{n}_0, T_e, D_{n_e} | d_{\text{FIR}}, d_{\text{Bal}}, I_{\text{all}}) \\ \propto P(d_{\text{FIR}} | \mathbf{n}_e) P(d_{\text{Bal}} | \mathbf{n}_e, \mathbf{n}_0, T_e) \\ \times P(I_{\text{pb}} | \mathbf{n}_e, \mathbf{n}_0, T_e, D_{n_e}) P(a_{D_{n_e}}) P(b_{D_{n_e}}) \\ \times P(T_e | I_{\text{mo}}, \phi_{\mu,\Sigma}^{\text{fix}}, T_{e,\text{peak}}) P(T_{e,\text{peak}}) \\ \times P(\mathbf{n}_0 | I_{\text{mo}}, \phi_{\mu,\Sigma}^{\text{free}}, n_{0,\text{peak}}) P(n_{0,\text{peak}}) \\ \times P(\mathbf{n}_e | I_{\text{mo}}, \phi_{\mu,\Sigma}^{\text{free}}, n_{e,\text{peak}}) P(n_{e,\text{peak}}). \end{aligned} \quad (28)$$

$P(a_{D_{n_e}})$ and $P(b_{D_{n_e}})$ are uniform distributions ranging from 0.5 to 2. The units are m^2/s and $\text{m}^2/(\text{s} \cdot \text{eV})$, respectively. We further assume that those upper and lower limits are known. The hyper-parameter $\hat{\tau}$ is set to 1 ms. We again assume that Balmer line intensities and line-integrated n_e measurements are available with 5% uncertainties.

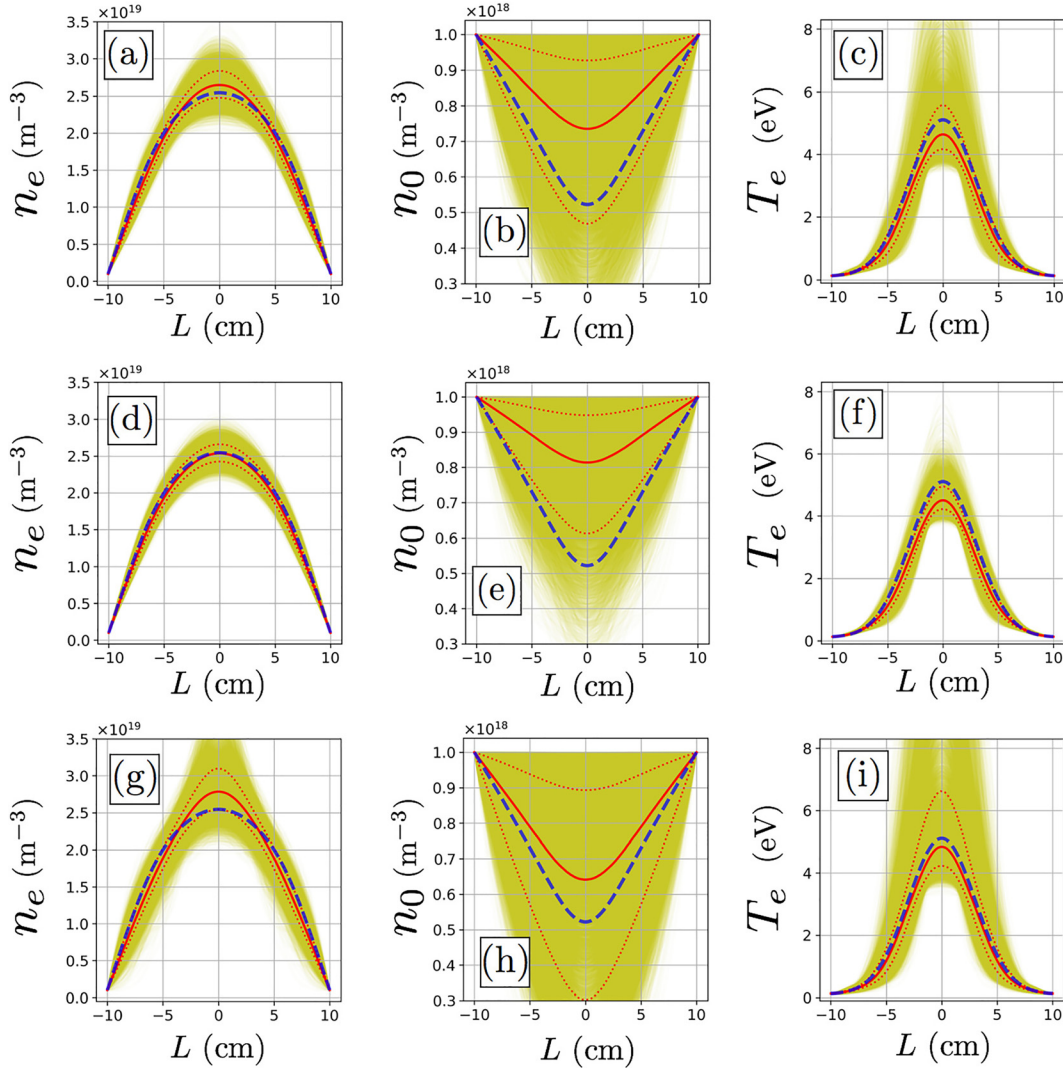


FIG. 8. Results of n_e , n_0 , and T_e inferences from line-integrated Balmer line intensities and line-integrated electron density. $\hat{\tau}$ is set to 1 ms for (a), (b), and (c), while $\hat{\tau} = 10$ ms used for (d), (e), and (f). $P(I_{pb}|n_e, n_0, T_e)$ is removed for (g), (h), and (i). The input profiles are shown in the blue-dashed lines. The red-dotted lines represent the 16th and 84th percentiles at each spatial point. The red solid lines represent the 50th percentiles at each spatial point. The original samples are shown in the yellow transparent lines. The number of samples and autocorrelation are 18,000 and < 20 samples, respectively, for all the three inferences.

We use the same input profiles as the Sec. VI and calculate the posterior distribution given by Eq. (28). The results of the n_e , n_0 , and T_e inferences are shown in Fig. 10. The posterior distributions look similar to those shown in Figs. 8(a)–8(c) even though Eq. (28) has two additional parameters $a_{D_{ne}}$ and $b_{D_{ne}}$. In reality, our transport models for laboratory plasmas almost certainly entail some degree of incompleteness. Even under such circumstances, we can still utilize the available part of the transport properties and improve the profile inference. Marginal posterior distributions of the unknowns in the electron particle diffusion coefficient are shown in Fig. 11. The distributions reach the upper and lower limits given by the prior. If wider ranges had been chosen for the $a_{D_{ne}}$ and $b_{D_{ne}}$, different profiles would have been inferred. Thus, in order to properly regularize the posterior distribution, we need to determine the upper and lower limits that are reasonable in terms of plasma physics.

Next, we consider a case where we have more diagnostic capabilities. So far, our inferences have been based on the line-integrated measurements. However, local measurements can also be included to construct even more informative posterior distributions using IDA. We assume that we install a Thomson scattering (TS) system⁴⁰ that measures n_e and T_e at 10 equally spaced spatial points from $L = -9$ to 0 cm. The likelihood for TS is

$$P(\mathbf{d}_{TS}|\mathbf{n}_e, \mathbf{T}_e) \propto \exp \left(-\sum_{x=1}^{10} \frac{(d_{TS,x}^{n_e} - d_{TS,x}^{n_e,model})^2}{2(\sigma_{TS,x}^{n_e})^2} \right) \times \exp \left(-\sum_{x=1}^{10} \frac{(d_{TS,x}^{T_e} - d_{TS,x}^{T_e,model})^2}{2(\sigma_{TS,x}^{T_e})^2} \right), \quad (29)$$

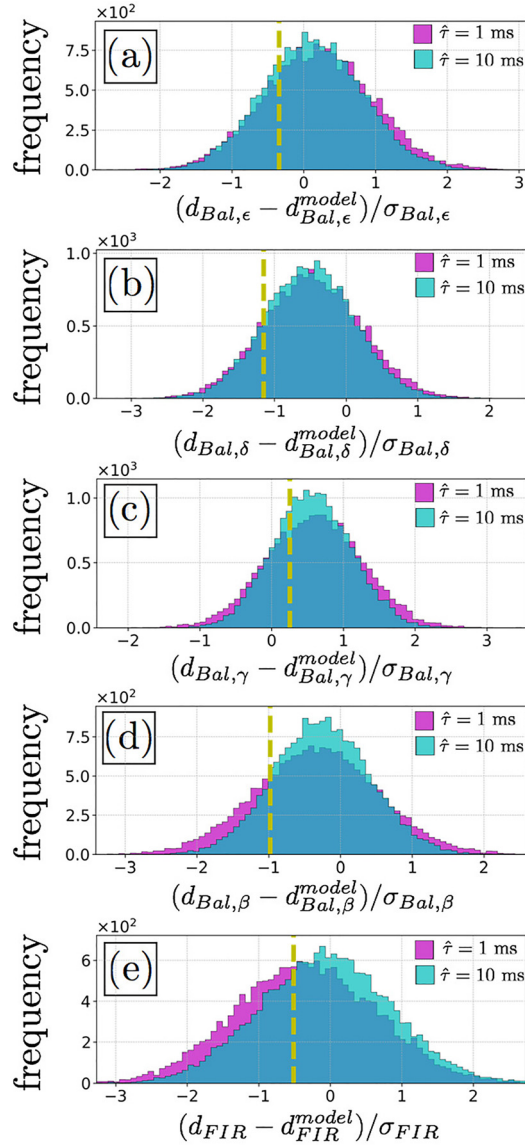


FIG. 9. Distributions of residual $(d_{Bal,\epsilon} - d_{Bal,\epsilon}^{model})/\sigma_{Bal,\epsilon}$ (a), $(d_{Bal,\delta} - d_{Bal,\delta}^{model})/\sigma_{Bal,\delta}$ (b), $(d_{Bal,\gamma} - d_{Bal,\gamma}^{model})/\sigma_{Bal,\gamma}$ (c), $(d_{Bal,\beta} - d_{Bal,\beta}^{model})/\sigma_{Bal,\beta}$ (d), and $(d_{FIR} - d_{FIR}^{model})/\sigma_{FIR}$ (e). The magenta histograms represent the residual distributions for the posterior distribution with $\hat{\tau} = 1$ ms shown in Figs. 8(a)–8(c), while the cyan histograms are the residual distributions for the posterior distribution with $\hat{\tau} = 10$ ms shown in Figs. 8(d)–8(f). The residuals for the data before random 5% noise is added are shown by the yellow-dashed lines. The input profiles correspond to these lines.

where $d_{TS,x}^{n_e}$ and $d_{TS,x}^{T_e}$ are local values of n_e and T_e , respectively. A 5% noise is again added by drawing samples from the corresponding normal distribution. $d_{TS,x}^{n_e,model}$ and $d_{TS,x}^{T_e,model}$ are the local values of n_e and T_e given by the prior, respectively. The subscript x represents the measurement locations of TS. $\sigma_{TS,x}^{n_e} = 0.05 \times d_{TS,x}^{n_e,model}$ and $\sigma_{TS,x}^{T_e} = 0.05 \times d_{TS,x}^{T_e,model}$. In real TS measurements, sophisticated forward models are

often used to take account of various sources of uncertainties.⁴ However, the scope of this paper is to provide a new analysis formalism, and no specific diagnostic system in reality is considered. Thus, we assume that the uncertainties of $d_{TS,x}^{n_e}$ and $d_{TS,x}^{T_e}$ follow a normal distribution for simplicity. The updated posterior distribution is given by

$$P(\mathbf{n}_e, \mathbf{n}_0, T_e, \mathbf{D}_{n_e} | \mathbf{d}_{TS}, \mathbf{d}_{FIR}, \mathbf{d}_{Bal}, \mathbf{I}_{all}) \propto P(\mathbf{d}_{TS} | \mathbf{n}_e, T_e) P(\mathbf{n}_e, \mathbf{n}_0, T_e, \mathbf{D}_{n_e} | \mathbf{d}_{FIR}, \mathbf{d}_{Bal}, \mathbf{I}_{all}). \quad (30)$$

We use $\hat{\tau} = 10$ ms, which is ten times larger than the case without TS. Figure 12 shows the posterior distribution of \mathbf{n}_e , \mathbf{n}_0 , and T_e given by Eq. (30). The local data points provided by TS greatly localize the posterior distributions of \mathbf{n}_e and T_e . Furthermore, we can see a significant improvement in the inference of \mathbf{n}_0 compared with Fig. 10(b) even though TS itself is not sensitive to n_0 . Unlike Figs. 8(h) and 8(i), no over-penalization is observed even with $\hat{\tau} = 10$ ms due to the constraints provided by TS. Figure 13 shows the marginal posterior distributions of $a_{D_{n_e}}$ and $b_{D_{n_e}}$ when TS is included. The distributions are well-localized, and the input values are within the ranges expected of the uncertainties. We have a combination of $a_{D_{n_e}}$, $b_{D_{n_e}}$, and T_e for each sample obtained by MCMC. Thus, by applying Eq. (27) to each sample, the posterior distribution of \mathbf{D}_{n_e} shown in Fig. 14 can be calculated. Since the probability distributions of $a_{D_{n_e}}$ and $b_{D_{n_e}}$ decay to sufficiently small values at the upper and lower limits given by the prior, similar results will be obtained for the \mathbf{D}_{n_e} inference when wider prior ranges are used for $a_{D_{n_e}}$ and $b_{D_{n_e}}$. Note that we applied model-based regularization to \mathbf{D}_{n_e} by using a specific functional form Eq. (27). In principle, model-free profile inference of a transport coefficient profile such as \mathbf{D}_{n_e} can also be attempted. When the likelihood is informative, more general model transport models, e.g., with convection, may be used. Investigating such a case is our future work.

VIII. SUMMARY AND CONCLUSIONS

In this paper, we have introduced an analysis framework to infer plasma parameter profiles from limited data within a Bayesian framework. In order to strongly constrain the posterior distribution without causing over-penalization, we utilize the physics-based knowledge (electron particle conservation) as prior information. Other desired characteristics of profiles such as smoothness, differentiability, monotonicity, and boundary conditions are imposed through the second-order derivative (SOD) prior. The parametrization of the SOD prior is chosen such that it facilitates the use of the Markov Chain Monte Carlo method. The hyper-parameters in the SOD prior can be tuned based on the prior knowledge on the profile shapes such as a characteristic scale length of a plasma parameter. The use of these prior beliefs does not reduce the degrees of freedom by assuming a specific functional model. Thus, flexible and unbiased parameter inference becomes possible.

We applied the profile inference technique using the physics-based prior belief and SOD priors to a synthetic one-dimensional plasma consisting of electrons, hydrogen ions, and hydrogen atoms, where the electron particle transport is described by a diffusive process. We attempted the simultaneous profile inference of electron density, neutral density, and electron temperature by using only line-integrated data. The inference results showed that line-integrated quantities are sufficient to provide reasonable estimates on the electron density and temperature profiles. The inferred neutral density profile has relatively

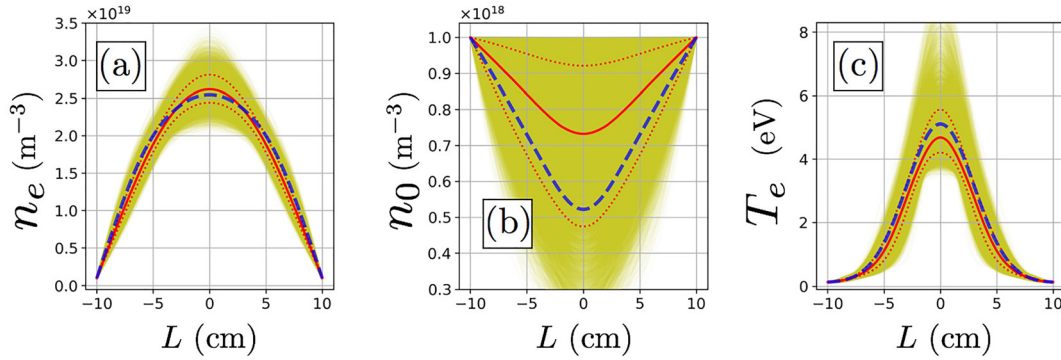


FIG. 10. Results of n_e (a), n_0 (b), and T_e (c) inference from Balmer-line intensities and interferometry when the physics knowledge is only partially available. The input profiles are shown in the blue-dashed lines. The red-dotted lines represent the 16th and 84th percentiles at each spatial point. The red solid lines represent the 50th percentiles at each spatial point. The original samples are shown in the yellow transparent lines. The number of samples and autocorrelation are 18 000 and <20 samples, respectively.

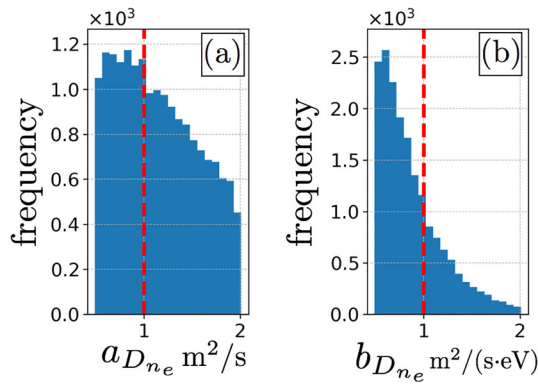


FIG. 11. Marginal posterior distributions of $a_{D_{ne}}$ (a) and $b_{D_{ne}}$ (b) for the posterior distribution given by Eq. (28). The input values are shown in the red-dashed lines.

large uncertainties but still offers useful information. While the SOD prior alone is still able to localize the posterior distribution to some degree, the physics-based prior belief significantly improves the quality of inference.

Finally, we investigated a case where the knowledge on the electron particle transport is only partially available, which is more relevant when this inference framework is applied to laboratory plasmas. We showed that informative inference of the electron density and temperature profiles can still be conducted under such circumstances. When local electron density and temperature measurements are available, we can even localize the posterior distribution for the neutral profile. In addition, the profile of the electron particle diffusion coefficient, to which none of the diagnostics considered in this paper is directly sensitive, is obtained.

In the proposed inference formalism, we infer the system as a whole, which consists of spatially varying multiple plasma parameters, by combining different diagnostics and our physics-based prior knowledge. Information about one plasma parameter helps localize the posterior distributions of others. Also, a quantity that is not

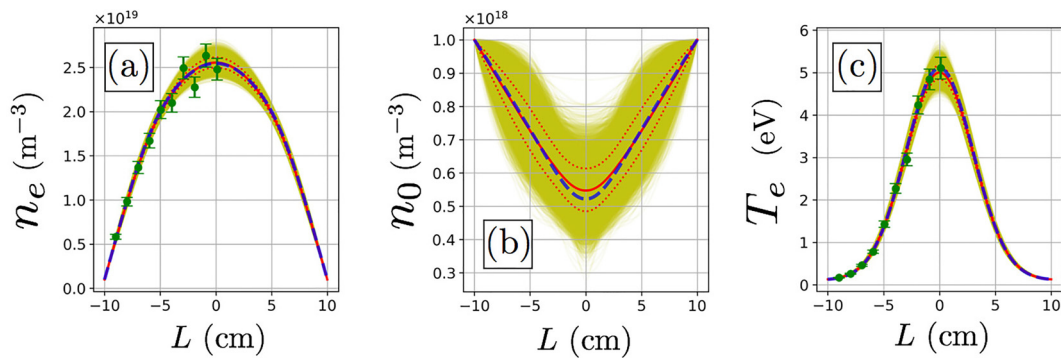


FIG. 12. Results of n_e (a), n_0 (b), and T_e (c) inference from Balmer-line intensities, interferometry, and TS when the physics knowledge is only partially available. The input profiles are shown in the blue-dashed lines. The red-dotted lines represent the 16th and 84th percentiles at each spatial point. The red solid lines represent the 50th percentiles at each spatial point. The original samples are shown in the yellow transparent lines. The green solid circles are the TS data points $d_{TS,x}^{n_e}$ and $d_{TS,x}^{T_e}$. The number of samples and autocorrelation are 18 000 and <20 samples, respectively.

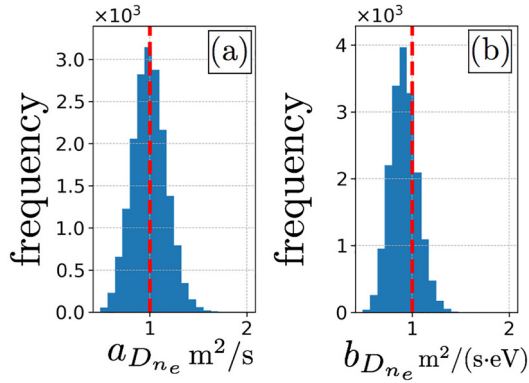


FIG. 13. Marginal posterior distributions of $a_{D_{n_e}}$ (a) and $b_{D_{n_e}}$ (b) for the posterior distribution given by Eq. (30). The input values are shown in the red-dashed lines.

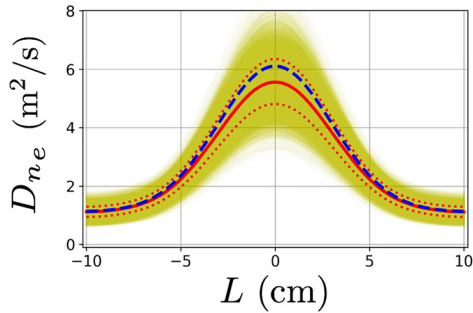


FIG. 14. Results of D_{n_e} inference from Balmer-line intensities, interferometry, and TS when the functional form of D_{n_e} is given by Eq. (27). The input profile is shown in the blue-dashed lines. The red-dotted lines represent the 16th and 84th percentiles at each spatial point. The red solid line represents the 50th percentiles at each spatial point. The original samples are shown in the yellow transparent lines.

directly accessible through standard diagnostics can be inferred. We are currently applying this inference technique to the characterization of divertor plasmas³⁵ and the profile measurements of transport coefficients and neutrals.⁴¹

ACKNOWLEDGMENTS

The authors would like to thank Dr. C. Bownman, Dr. P. David, Dr. S. Kwak, Dr. M. Sasaki, Dr. J. Boguski, and Dr. P. Molina Cabrera for valuable discussions. This work has been carried out within the framework of the EUROfusion Consortium and has received funding from the Euratom research and training program 2014–2018 and 2019–2020 under Grant Agreement No. 633053. The views and opinions expressed herein do not necessarily reflect those of the European Commission.

NOMENCLATURE

ACD	Effective recombination rate
d	Data
d^{model}	Forward-modeled data

D_{n_e}	Discretized profile of D_{n_e}
D_{n_e}	Electron diffusion coefficient
D_{n_0}	Neutral diffusion coefficient
I	Prior knowledge
i	Index for θ
I_{mo}	Monotonicity prior belief
I_{pb}	Physics-based prior belief
L	Spatial coordinate of the plasma
l	Index for spatial coordinate
\hat{l}	Hyper-parameter of SE
\hat{L}_0	Hyper-parameter for I_{pb}
n_e	Discretized profile of n_e
n_e	Electron density
n_0	Discretized profile of n_0
n_0	Neutral density
$\tilde{N}(\theta)$	Penalty term
\mathcal{PEC}	Photon emissivity coefficient
r	Spatial coordinate
S	Source of n_e
SCD	effective ionization rate
t	Time
T_e	Discretized profile of T_e
T_e	Electron temperature
x	Index for d and d^{model}
θ	Parameter
$\hat{\lambda}$	Hyper parameter for $\tilde{N}(\theta)$
μ	Mean of a GP
$\hat{\sigma}$	Hyper-parameter of SE
$\hat{\tau}$	Hyper-parameter for I_{pb}
$\phi_{\mu,\Sigma}^{fix}$	Mapping with a fixed boundary
$\phi_{\mu,\Sigma}^{free}$	Mapping with a free boundary
Σ	Covariance function of a GP

APPENDIX: HYPER-PARAMETERS FOR FIG. 5

We used a SE to make samples shown in Fig. 5. The hyper-parameters are $\mu = 0$, and

$$\hat{\sigma}^2 = \hat{\sigma}_i \hat{\sigma}_{i'}, \quad (A1)$$

where

$$\hat{\sigma}_i = \exp(-5i/N) + 0.1. \quad (A2)$$

We calculated profiles by using the mapping $\phi_{\mu,\Sigma}^{free}$ given by Eq. (12).

DATA AVAILABILITY

The data that support the findings of this study are available from the corresponding author upon reasonable request.

REFERENCES

- ¹P. W. Terry, *Rev. Mod. Phys.* **72**, 109 (2000).
- ²P. Snyder, W. Solomon, K. Burrell, A. Garofalo, B. Grierson, R. Groebner, A. Leonard, R. Nazikian, T. Osborne, E. Belli, J. Candy, and H. Wilson, *Nucl. Fusion* **55**, 083026 (2015).
- ³R. Fischer, C. J. Fuchs, B. Kurzan, W. Suttrop, and E. Wolfrum, and A. U. Team, *Fusion Sci. Technol.* **58**, 675 (2010).
- ⁴S. Kwak, J. Svensson, S. Bozhnikov, J. Flanagan, M. Kempenaars, A. Boboc, and Y.-C. Ghim, *Nucl. Fusion* **60**, 046009 (2020).

- ⁵U. von Toussaint, *Rev. Mod. Phys.* **83**, 943 (2011).
- ⁶S. Kwak, J. Svensson, M. Brix, and Y.-C. Ghim, *Nucl. Fusion* **57**, 036017 (2017).
- ⁷R. Fischer, C. Wendland, A. Dinklage, S. Gori, and V. Dose, and the W7-AS Team, *Plasma Phys. Controlled Fusion* **44**, 1501 (2002).
- ⁸M. Chilenski, M. Greenwald, Y. Marzouk, N. Howard, A. White, J. Rice, and J. Walk, *Nucl. Fusion* **55**, 023012 (2015).
- ⁹A. Ho, J. Citrin, F. Auriemma, C. Bourdelle, F. Casson, H.-T. Kim, P. Manas, G. Szepesi, and H. Weiser, *Nucl. Fusion* **59**, 056007 (2019).
- ¹⁰C. E. Rasmussen, *Summer School on Machine Learning* (Springer, 2003), pp. 63–71.
- ¹¹C. K. I. Williams, *Gaussian Processes Formachine Learning* (Taylor & Francis Group, 2006).
- ¹²D. Li, J. Svensson, H. Thomsen, F. Medina, A. Werner, and R. Wolf, *Rev. Sci. Instrum.* **84**, 083506 (2013).
- ¹³G. T. von Nessi, M. J. Hole, J. Svensson, and L. Appel, *Phys. Plasmas* **19**, 012506 (2012).
- ¹⁴K. Verhaegh, B. Lipschultz, B. P. Duval, A. Fil, M. Wensing, C. Bowman, and D. Gahle, *Plasma Phys. Controlled Fusion* **61**, 125018 (2019).
- ¹⁵B. Lomanowski, A. Meigs, R. Sharples, M. Stamp, and C. Guillemaut, *Nucl. Fusion* **55**, 123028 (2015).
- ¹⁶D. J. D. Hartog, A. F. Almagri, S. C. Prager, and R. J. Fonck, *Rev. Sci. Instrum.* **66**, 444 (1995).
- ¹⁷C. Bowman, J. R. Harrison, B. Lipschultz, S. Orchard, K. J. Gibson, M. Carr, K. Verhaegh, and O. Myatra, *Plasma Phys. Controlled Fusion* **62**, 045014 (2020).
- ¹⁸Y. Nagayama, *J. Appl. Phys.* **62**, 2702 (1987).
- ¹⁹K. Ertl, W. v d Linden, V. Dose, and A. Weller, *Nucl. Fusion* **36**, 1477 (1996).
- ²⁰M. Anton, H. Weisen, M. J. Dutch, W. v der Linden, F. Buhlmann, R. Chavan, B. Marletaz, P. Marmillod, and P. Paris, *Plasma Phys. Controlled Fusion* **38**, 1849 (1996).
- ²¹N. Iwama, M. Yamaguchi, K. Hattori, and M. Hayakawa, *J. Electromagnetic Waves Appl.* **9**, 757 (1995).
- ²²H. P. Summers, *The ADAS User Manual*, version 2.6 ed. (The ADAS Project <http://www.adas.ac.uk>, 2004).
- ²³G. M. D. Hogewei, A. A. M. Oomens, C. J. Barth, M. N. A. Beurskens, C. C. Chu, J. F. M. van Gelder, J. Lok, N. J. Lopes Cardozo, F. J. Pijper, R. W. Polman, and J. H. Rommers, *Phys. Rev. Lett.* **76**, 632 (1996).
- ²⁴H. Takahashi, K. Nagaoka, K. Mukai, M. Yokoyama, S. Murakami, S. Ohdachi, T. Bando, Y. Narushima, H. Nakano, M. Osakabe, K. Ida, M. Yoshinuma, R. Seki, H. Yamaguchi, K. Tanaka, M. Nakata, F. Warmer, T. Oishi, M. Goto, S. Morita, T. Tsujimura, S. Kubo, T. Kobayashi, I. Yamada, C. Suzuki, M. Emoto, T. Ido, A. Shimizu, T. Tokuzawa, K. Nagasaki, T. Morisaki, and Y. Takeri, *Nucl. Fusion* **58**, 106028 (2018).
- ²⁵M. D. Hoffman and A. Gelman, *J. Mach. Learn. Res.* **15**, 1593 (2014).
- ²⁶J. Salvatier, T. V. Wiecki, and C. Fonnesbeck, *Peer J. Comput. Sci.* **2**, e55 (2016).
- ²⁷R. M. Neal, *Handbook of Markov Chain Monte Carlo* (2011), Vol. 2, p. 2.
- ²⁸A. Beskos, N. Pillai, G. Roberts, J.-M. Sanz-Serna, and A. Stuart, *Bernoulli* **19**, 1501 (2013).
- ²⁹M. Creutz, *Phys. Rev. D* **38**, 1228 (1988).
- ³⁰D. Van Ravenzwaaij, P. Cassey, and S. D. Brown, *Psychonomic Bull. Rev.* **25**, 143 (2018).
- ³¹M. Chilenski, M. Greenwald, A. Hubbard, J. Hughes, J. Lee, Y. Marzouk, J. Rice, and A. White, *Nucl. Fusion* **57**, 126013 (2017).
- ³²F. Sciortino, N. Howard, E. Marmar, T. Odstrcil, N. Cao, R. Dux, A. Hubbard, J. Hughes, J. Irby, Y. Marzouk, L. Milanese, M. Reinke, J. Rice, and P. Rodriguez-Fernandez, *Nucl. Fusion* **60**, 126014 (2020).
- ³³K. Verhaegh, B. Lipschultz, B. Duval, O. Février, A. Fil, C. Theiler, M. Wensing, C. Bowman, D. Gahle, J. Harrison, B. Labit, C. Marini, R. Maurizio, H. de Oliveira, H. Reimerdes, U. Sheikh, C. Tsui, N. Vianello, and W. Vijvers, *Nucl. Fusion* **59**, 126038 (2019).
- ³⁴S. Potzel, M. Wischmeier, M. Bernert, R. Dux, H. Müller, and A. Scarabosio, *Nucl. Fusion* **54**, 013001 (2014).
- ³⁵T. Nishizawa, M. Cavedon, F. Reimold, R. Dux, D. Brida, and H. Wu, *Plasma Phys. Controlled Fusion* **62**, 085005 (2020).
- ³⁶W. X. Ding, D. L. Brower, B. H. Deng, and T. Yates, *Rev. Sci. Instrum.* **77**, 10F105 (2006).
- ³⁷N. Horsten, W. Dekeyser, G. Samaey, and M. Baelmans, in *Proceedings of the 22nd International Conference on Plasma Surface Interactions* (2016) [Nucl. Mater. Energy **12**, 869 (2017)].
- ³⁸D. Reiter, M. Baelmans, and P. Börner, *Fusion Sci. Technol.* **47**, 172 (2005).
- ³⁹M. Betancourt, [arXiv:1701.02434](https://arxiv.org/abs/1701.02434) (2017).
- ⁴⁰B. Kurzan and H. D. Murmann, *Rev. Sci. Instrum.* **82**, 103501 (2011).
- ⁴¹R. Dux, M. Cavedon, A. Kallenbach, R. McDermott, G. Vogel, and the ASDEX Upgrade team, *Nucl. Fusion* **60**, 126039 (2020).