

Bayesian data analysis for Gaussian process tomography

T. Wang^{1,2,3}, D. Mazon², J. Svensson⁵, A. Liu⁶, C. Zhou⁶, L. Xu⁷, L. Hu⁷, Y. Duan⁷ and G. Verdoolaege^{3,4}

¹*Southwestern Institute for Physics, CNNC, C-610200 Chengdu, China*

²*Institute for Magnetic Fusion Research, CEA, F-13115 Saint-Paul-lez-Durance, France*

³*Department of Applied Physics, Ghent University, B-9000 Ghent, Belgium*

⁴*Laboratory for Plasma Physics, Royal Military Academy (LPP-ERM/KMS), B-1000 Brussels, Belgium*

⁵*Max-Planck-Institut für Plasmaphysik, D-17491 Greifswald, Germany*

⁶*CAS Key Laboratory of Basic Plasma Physics and Department of Modern Physics, University of Science and Technology of China, Hefei 230026, People's Republic of China*

⁷*Institute of Plasma Physics, Chinese Academy of Sciences, Hefei 230031, People's Republic of China*

Contact email: tianbo.wang@ugent.be

Abstract

Bayesian inference is used in many scientific areas as a conceptually well-founded data analysis framework. In this paper, we give a brief introduction to Bayesian probability theory and its application to the tomography problem in fusion research by means of a Gaussian process prior. This Gaussian process tomography (GPT) method is used for reconstruction of the local soft X-ray (SXR) emissivity in WEST and EAST based on line-integrated data. By modeling the SXR emissivity field in a poloidal cross-section as a Gaussian process, Bayesian SXR tomography can be carried out in a robust and extremely fast way. Owing to the short execution time of the algorithm, GPT is an important candidate for providing real-time feedback information on impurity transport and for fast MHD control. In addition, the Bayesian formalism allows for uncertainty analysis of the inferred emissivity.

1 Data analysis in fusion experiments

1.1 Introduction

Magnetic confinement fusion experiments generate large quantities of complex data. At a basic level, the data reflects the state of the machine and plasma, enabling safe and reliable operation of the device, i.e. well within the design limits of the machine and compatible with the scientific goals of the experiment. In addition, the data is an essential ingredient to interpreting the experiment at a higher level, deepening the physical understanding of the various plasma processes and their interaction with the device.

In a fusion reactor, most of the data will be used for machine operation. The main criteria will be safety and plasma performance, maximizing the fusion output. Modelling codes produce even more data, which needs to be compared to the experimental results. Depending on the requirements, different analysis techniques are needed to extract as much useful

information as possible from the raw data. This ranges from very basic tools for signal resampling or motion correction in images to more advanced statistical analysis or pattern recognition methods. As the available computing power increases, real-time high-level data analysis also starts to become feasible for plasma control.

The throughput volume and complexity of fusion experimental data demand special attention for the analysis. As an illustration, the JET experiment presently generates ~ 55 GB of data per pulse and ITER will produce up to 1 TB in each discharge. The complexity of the data is essentially due to the complexity of the physics of a fusion plasma and to uncertainty introduced by the diagnostic techniques. Indeed, the variables through which the plasma is characterized usually interact in an intricate, nonlinear way and their measurements can be affected by considerable uncertainty. Furthermore, different diagnostics may provide information on the same plasma quantity, resulting in a certain degree of redundancy that can be exploited for the purpose of joint analysis of multiple diagnostics, but may also entail issues of data consistency (through measurement error and incorrect assumptions regarding the measurement model) and data heterogeneity. Modern information science can play an important role in the improvement of data accuracy and robustness and the unravelling of useful physical quantities or relations from the data.

Though modern tools from probability theory and machine learning have been applied to fusion experiments, they are still relatively uncommon in fusion science. The situation is quite different in other data-intensive disciplines of the physical sciences (e.g. high-energy physics, astronomy and ecology), where advanced techniques have been part of the common practice for many years. Therefore, development and application of state-of-the-art techniques based on probability theory and machine learning needs to be stimulated in fusion science, in order to address complex data analysis problems. The goal should be to raise the part of the scientific process related to data analysis to the same high level as the experimentation, modelling and theory building.

We start by introducing some basic concepts from Bayesian probability theory, with a view to the application to Gaussian process tomography presented in Section 2. We do not strive for comprehensiveness; rather mean to give a flavour of typical method and references to further readings, which we hope will prove to be useful to the beginning practitioner.

1.2 A note about probability

We begin our discussion with a short note about the view we choose to adopt on probability. Over time, various interpretations of probability have been proposed, the most popular by far being the *frequentist* and the *Bayesian* approaches. Both interpretations come with their associated set of methods and tools, although practices vary across disciplines (e.g. social sciences vs. physical sciences). In addition, there exist various flavours of Bayesian thinking, some of which are half frequentist and use frequentist terminology. However, in the physical sciences and fusion in particular, the point of view of [Cox61, Jaynes04] has come to prevail, wherein probability theory is considered as an extension of logic to uncertain propositions. The theory is based on a set of common-sense axioms (or desiderata), including the basic requirement of consistency. In other words, different observers should arrive at the same conclusions when having been provided with the same information. From this point of view, a probability always depends on two pieces of information:

the proposition or hypothesis A , the probability of which we seek, and the information B , conditioning the probability. The standard notation used to denote such a probability is $P(A|B)$ and corresponds to a real number in the interval $[0,1]$ expressing the plausibility of A provided B is true. Put differently, it is a measure of the extent to which B *implies* A . As such, a probability denoted by $P(.|.)$ always has two slots that need to be filled to produce a meaningful numerical output. The conditioning information B can be a problem statement, some measured data or any other available background information. The whole of information on which a probability statement is conditioned is often summarized by the generic symbol I , e.g. $P(.|I)$. Specific additional information conditioning the probability may be stated explicitly, e.g. $P(.|B,I)$.

The classical frequentist definition of probability is based on the frequency of an event in the long run and necessitates identical repeats of experiments or ensembles of the system under study. As such, it contrasts with the Bayesian interpretation, which enables a direct evaluation of the probability of any statement or single event. Examples include the probability of the plasma density in a fusion machine being within certain bounds or the probability that a plasma is in the H-mode. In the Bayesian view, probabilities are useful whenever there is a general lack of certainty, e.g. details of plasma particle orbits causing fluctuations in the macroscopic quantities or the electron thermal motion in a detector introducing noise in the measurement. No appeal is made to any notion of ‘randomness’ and although the term ‘random variable’ occurs in some Bayesian literature, we will avoid it. Neither will we follow the habit of using capital letters to denote random variables since, from our point of view, the information carried by any physical variable is subject to uncertainty. This approach is adopted by several excellent textbooks on Bayesian probability theory geared to physicists; see [Sivia06, von der Linden14, Gregory05, Jaynes04].

1.3 Bayesian diagnostic modelling

1.3.1 Introduction

An important application of Bayesian probability theory that has gained considerable use within the fusion community can be found in fusion diagnostics modelling. This practice is motivated by the frequent and substantial uncertainties entering the derivation of physical quantities from raw (measured) data, often a voltage at a sensor. Without proper measures, these uncertainties can enhance and complicate the total uncertainty on the quantity of interest, deteriorating the quality and reliability of the result. In extreme cases the result may become meaningless, which, for instance, may occur in the calculation of profiles from line-integrated measurements through an inversion technique. Not only is this problem heavily underdetermined (multiple solutions fit the data), but the noise on the line-integrated measurements also propagates through the inversion and generally worsens the result. Limiting the number and smoothness of the solutions by regularization techniques is the standard solution, but further improvements are often possible. Bayesian probability theory offers a solid and coherent framework for better profile reconstruction. It consistently handles any kind of uncertainty by assigning probabilities (or probability distributions) and manipulating them using a set of well-defined rules. One of these is *Bayes’ rule* – hence the designation ‘Bayesian probability theory’ – which we will introduce below. The result of a Bayesian calculation is a probability (distribution) for the quantity of interest or a hypothesis, depending on the available information.

In addition to modelling of individual diagnostics, Bayesian theory enables combination of measurements of individual diagnostics, resulting in an increase in accuracy of common physics parameters of interest. By modelling each diagnostic in the Bayesian framework, a probability for a specific plasma quantity can be derived that is consistent with the data from all diagnostics involved in the analysis. Therefore, when multiple diagnostic measurements contain information regarding a certain plasma quantity (e.g. interferometry and Thomson scattering for the electron density of a plasma), it is possible to exploit the advantages of each diagnostic, hence improving the quality of the estimate of the desired plasma quantity. This quality can be reflected in smaller error bars, higher resolution or enhanced robustness, since the analysis involves information from diagnostics that are based on different measurement principles or different hardware components.

1.3.2 Concepts of Bayesian probability theory

In this section, we introduce some basic concepts from Bayesian probability theory. We do this in the context of a simple example involving estimation of the parameters of a normal distribution based on suitable prior information and repeated measurements of a physical quantity.

In fusion, as in many scientific activities, we do not have a direct access to the physical properties of the plasma. The experimentalist must instead devise a measurement technique for a diagnostic, which returns numbers (usually voltages) that are related to the quantity of interest. If the properties of a physical system were known precisely, together with the full details of the measurement process, the corresponding measurement values could be computed straightforwardly. Calculating the measurements would require the measurement process to be encoded in a mathematical model: the *forward model*. However, since the microscopic details of the physical system and the measurement process are unknown, the raw measurement itself is never a truly precise number.¹ We may nevertheless be able to model this uncertainty by assigning a probability for the measured value to lie within certain bounds. Representing the measurement by a real number x , it is convenient to introduce the *probability density function* (PDF) of x .² The PDF is usually denoted $p(x|I)$, where $p(x|I)dx$ can be interpreted as the probability to find the measurement value between x and $x + dx$, given the background information I .³ Strictly speaking, for continuous variables, the term *distribution function* refers to the cumulative distribution function; it is nevertheless used colloquially to denote the PDF. To indicate that a variable x is assumed to be distributed according to (or sampled from) the model with PDF $p(x|I)$, one writes $x \sim p$. One of the most common models is the *normal* or *Gaussian distribution*, the density function of which is the well-known function:

$$p(x|\mu, \sigma, I) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{(x - \mu)^2}{2\sigma^2}\right]. \quad (1.1)$$

This is often written as $x \sim \mathcal{N}(\mu, \sigma^2)$, where μ is the mean (expectation value) of the distribution and σ the standard deviation. Note that, in the Bayesian spirit, we have explicitly written the PDF as a two-slot function $p(x|\mu, \sigma, I)$, where

¹ At a certain point, one also hits the rather philosophical question of whether there is such a thing as an ‘exact value’ of a (microscopic) physical quantity.

² We will treat only continuous variables here. This represents the most common situations in physical sciences.

³ We will follow the common slight abuse of notation of using the symbol p to denote a specific model (e.g. Gaussian) of the probability density of a specific variable or, in general, the probability density of a variable, possibly unknown or unspecified.

the second slot emphasizes that the probability density can only be calculated for known values of μ and σ . Any additional information (e.g. the fact that it is a Gaussian distribution) has been summarized in the symbol I . With the same knowledge, the probability that x is found to lie within two given bounds can be calculated by integration (involving the error function erf). The measurement uncertainties that can be described by probability models such as (1.1) are often referred to as *stochastic* or *statistical uncertainties*. This is opposed to *systematic uncertainties*, which result in a ‘deterministic’ deviation of the result from the ‘correct’ measurement.

An entirely different question concerns the assignment of probabilities to the parameters μ and σ , given a set of measurements for the variable x , denoted by \vec{x} . The solution, known as *Bayes’ rule* or *Bayes’ theorem*, is generally credited to Thomas Bayes (1763) and Pierre-Simon Laplace (1812). This is a direct consequence of the well-known product rule of probability theory, which in this context reads

$$p(\vec{x}, \mu, \sigma | I) = p(\vec{x} | \mu, \sigma, I) p(\mu, \sigma | I). \quad (1.2)$$

Of course, the same rule also yields

$$p(\vec{x}, \mu, \sigma | I) = p(\mu, \sigma | \vec{x}, I) p(\vec{x} | I),$$

which, together with (1.2), results in Bayes’ rule (applied to the current example):

$$\overbrace{p(\mu, \sigma | \vec{x}, I)}^{\text{posterior}} = \frac{\overbrace{p(\vec{x} | \mu, \sigma, I)}^{\text{likelihood}} \overbrace{p(\mu, \sigma | I)}^{\text{prior}}}{\underbrace{p(\vec{x} | I)}_{\text{evidence}}}. \quad (1.3)$$

Thus, the rules of probability theory naturally enable solving the inverse problem by using a forward model. Determining appropriate values for the quantities of interest – μ and σ – from (1.3) is referred to as the task of *parameter estimation*. We will discuss this task in more detail below, still in the context of a normal distribution.

Bayes’ rule can also be interpreted in terms of a learning process. In this sense, information regarding the parameters μ and σ is considered before the data is acquired, using the so-called *prior distribution*, indicated in (1.3). The prior can be obtained from background knowledge about the problem or other experiments. Alternatively, we may choose an *uninformative prior*, which, as the term implies, allows us to adopt a maximally ignorant point of view before performing the experiment. However, one should be careful in keeping the problem well-posed (or *identifiable*), i.e. there should be sufficient information in the data and the prior to estimate the parameters with reasonable accuracy (e.g. the discussion in [Verdoolaege10]). Various criteria and methods exist to assign uninformative prior distributions, including invariance under basic transformations (e.g. translations and rotations), *maximum entropy distributions* in the presence of specific pieces of *testable* information (e.g. knowledge of the mean of a distribution), smoothness-ensuring priors for regularizing

ill-posed problems and others (see [Sivia06, von der Linden14] for practical advice and [Jaynes04] for more in-depth discussions). Whether one chooses an uninformative or informative prior, in the objective Bayesian view the prior distribution is not a subjective judgment by an individual scientist, but rather the result of quantifying the available prior information. In this sense, two individuals, given the same information prior to the experiment, should arrive at the same prior distribution, hence should obtain the same conclusions. Many practitioners use a very pragmatic rule to assign prior distributions, which is related to the analytical tractability of the posterior. This gives rise to the so-called *conjugate priors*, described for instance in [Gelman13]. However, one could object that more objective prior information is to be preferred over a criterion based on computational ease, especially as the likelihood and the experimental setup also depend, to some extent, on ‘subjective’ choices.

Continuing with our example of inferring the parameters of a normal distribution, we could next perform a series of n measurements, here described by \vec{x} . We will assume, given a common μ and σ , that the measurements are performed independently from each other. Fixed model parameters μ and σ imply stationary conditions in the system of interest and the measurement process. Under these assumptions, the distribution of \vec{x} , conditioned on μ and σ , can be factorized into a product of marginal distributions for each of the x_i components of \vec{x} ($i = 1 \dots n$):

$$p(\vec{x}|\mu, \sigma, I) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{(x_i - \mu)^2}{2\sigma^2}\right] = \frac{1}{(2\pi)^{n/2}\sigma^n} \exp\left[-\sum_{i=1}^n \frac{(x_i - \mu)^2}{2\sigma^2}\right]. \quad (1.4)$$

This is the *likelihood distribution*. According to (1.3), the product of the likelihood and the prior is proportional to the *posterior distribution* for the quantities of interest. The posterior quantifies our knowledge regarding the parameters of interest μ and σ , based on the prior information and the data. In this respect, an important observation is that, in estimating μ and σ , we do not need to care about the normalization of the posterior. Only the posterior’s dependence on μ and σ matters, since it determines the shape of their joint distribution. For this reason, the denominator in (1.3) is irrelevant in the context of parameter estimation and may be absorbed in a proportionality constant. The denominator is referred to as the *evidence* (or marginal likelihood) since it gives the probability of the data in the light of the background information, which includes knowledge about the model that we use for the physical and measurement systems. This term provides evidence for our model and becomes important in comparing different models. In the Bayesian approach, the task of *model selection* is also performed according to Bayes’ rule; we will not go deeper into this topic here, but the reader may refer to [von der Linden14].

It is possible to show that the uniform distribution is an adequate uninformative prior for μ , while Jeffrey’s scale prior ($\propto 1/\sigma$) is suitable for σ [von der Linden14]. Under these assumptions, the joint posterior distribution for μ and σ obeys the following proportionality:

$$p(\mu, \sigma | \vec{x}, I) \propto \frac{1}{(2\pi)^{n/2} \sigma^{n+1}} \exp \left[- \sum_{i=1}^n \frac{(x_i - \mu)^2}{2\sigma^2} \right]. \quad (1.5)$$

From this, the marginal distribution for μ can be obtained by integrating over σ – an operation that is ubiquitous in Bayesian probability theory and that is referred to as *marginalization*. For μ , this leads to Student’s t-distribution, while for σ one obtains a χ^2 marginal distribution [von der Linden14].

2 Gaussian process tomography

In this section, we give an example application of Bayesian inference in fusion research, using the relatively new technique of Gaussian process tomography.

2.1 Tomography in fusion devices

Reconstruction of a space-resolved image or profile of a plasma quantity using tomography has been considered on many fusion devices [Mlynar15, Mazon12, Jardin16]. The tomographic problem essentially involves prediction of high-dimensional physics parameters by an inversion of a limited number of measurements. This is an ill-posed problem, as the number of measurements is always lower than the number of unknowns.

In our present example we consider tomography for a soft X-ray (SXR) diagnostic at WEST (Tungsten Environment in Steady-State Tokamak) using synthetic data. Then, we provide another example using real data from EAST (Experimental Advanced Superconducting Tokamak). SXR spectroscopy can give valuable information about the plasma in a two-dimensional poloidal cross-section, such as magnetohydrodynamic activity and impurity transport. In the past, various tomographic reconstruction techniques have been applied to SXR, such as the Cormack method [Cormack64], the maximum entropy method, the minimum Fisher information method [Ertl98], etc. Particularly the minimum Fisher information technique has been widely adopted in the fusion community [Mlynar15, Mazon12, Jardin16, Anton 96]. This reconstruction method involves χ^2 optimization, regularized by the Fisher information. Intuitively, the goal is to find the least complex solution that is compatible with the data. The method is often implemented, e.g. on Tore Supra and WEST, using additional information concerning the location of the equilibrium magnetic flux surfaces, obtained from magnetic measurements. It provides a good trade-off between the initial magnetic flux surface information and the SXR data.

2.2 Bayesian tomographic model

The purpose of SXR tomography in magnetic fusion devices is to reveal the spatial distribution of SXR emissivity in a poloidal cross-section, by the inversion of a number of noisy line-integrated emissivity measurements. A common and simple approach to discretize the emissivity field in a poloidal cross-section uses a square grid. We here impose a 100×100 grid comprising square cells with a dimension of $16\text{mm} \times 16\text{mm}$. The SXR emissivity within each pixel can reasonably be assumed to be constant, so the SXR line-integrated emissivity \bar{d}_m along m viewing chords can be written in the following matrix form:

$$\bar{d}_m = \bar{R}_{m \times n} \cdot \bar{E}_n + \bar{\varepsilon}. \quad (2.1)$$

Here, \bar{E}_n is the unknown vector of local emissivities in $n = 10^4$ cells, while \bar{R} is the geometry matrix, whose elements R_{ij} represent the physical length of chord i through cell j . $\bar{\varepsilon}$ denotes an error term to account for measurement uncertainty, which is usually limited to statistical errors only.

From the Bayesian perspective, (2.1) is the basic forward model of the tomography problem. The classic optimization approach uses a criterion that minimizes the difference between the measured line integrals and the prediction by the model. Because of the ill-posedness, the optimization has to be combined with some regularization technique, e.g. assuming a spline model for the local emissivity field, or by optimizing at the same time some information measure like the Shannon entropy or the Fisher information. In this paper we choose a probabilistic methodology, which provides a probability distribution $p(\bar{E}_n)$ of the emissivity in all cells rather than a single solution. In Bayesian inference, one starts from the prior probability distribution of the emissivity field, which can be used to encode the regularization. This is then updated through Bayes' theorem as data become available (the background information I is implicit):

$$p(\bar{E}_n | \bar{d}_m) = \frac{p(\bar{d}_m | \bar{E}_n) p(\bar{E}_n)}{p(\bar{d}_m)} \sim p(\bar{d}_m | \bar{E}_n) p(\bar{E}_n). \quad (2.2)$$

$$p(\bar{d}_m) = \int p(\bar{d}_m, \bar{E}_n) d\bar{E}_n = \int p(\bar{d}_m | \bar{E}_n) p(\bar{E}_n) d\bar{E}_n. \quad (2.3)$$

\bar{E}_n Vector of emissivity values in all n pixels at a particular time t

\bar{d}_m Vector of m line-integrated GEM array measurements at time t

In Eq. (2.2), the likelihood term $p(\bar{d}_m | \bar{E}_n)$ assesses the mismatch between the measured line integrals \bar{d}_m and their predictions by the forward model, under the assumption of some emissivity field \bar{E}_n . In Eq. (2.3), we have marginalized over the emissivity field. The resulting evidence (marginal likelihood) $p(\bar{d}_m)$ depends on the particular forward measurement model, which we will assume to be fixed. Therefore, again it can be considered as a normalization factor, independent of the emissivity. The posterior probability distribution $p(\bar{E}_n | \bar{d}_m)$ quantifies our uncertainty on the estimated emissivity field, given our model, prior knowledge and the measured data. Thus, Bayesian inference yields probabilities for all possible results consistent with our model. In principle, systematic uncertainties can also be estimated, provided some knowledge is available about them from other sources of information, such as other experiments.

2.3 Gaussian process prior and posterior

Gaussian process tomography (GPT) is a new technique whereby the prior distribution regularizes the tomographic reconstruction process [Svensson11], by imposing a smoothness level dictated by the correlation between pixels. Briefly, a

Gaussian process (GP) is a generalization of the multivariate normal (Gaussian) distribution to a function space. It is described by a mean function $\bar{\mu}(\vec{r})$ and a covariance function $\bar{\Sigma}(\vec{r}, \vec{r}')$, where $GP \sim \mathcal{N}(\bar{\mu}, \bar{\Sigma})$. The distribution of a Gaussian process is the joint distribution of infinitely many normally distributed random variables, and as such, it is a distribution over functions with a continuous domain, e.g. time or space. GPT is related to Gaussian process regression (or “*kriging*” in geoscience), a nonparametric regression technique widely used in machine learning. Being nonparametric, Gaussian process regression does not assume any functional form for the regression function, hence leaving a lot of flexibility. Instead, the regression surface is regularized through the covariance matrix of the Gaussian process. Likewise, GPT assumes that the prior joint distribution of the emissivity in the n cells with coordinates \vec{r}_i is multivariate Gaussian (Fig. 2.1) with covariance matrix $\bar{\Sigma}_E$ given by:

$$\bar{\Sigma}_E = \begin{pmatrix} k(\vec{r}_1, \vec{r}_1) & \cdots & k(\vec{r}_1, \vec{r}_n) \\ \vdots & \ddots & \vdots \\ k(\vec{r}_n, \vec{r}_1) & \cdots & k(\vec{r}_n, \vec{r}_n) \end{pmatrix}. \quad (2.4)$$

Here, $k(\vec{r}_i, \vec{r}_j) = \text{cov}[E(\vec{r}_i), E(\vec{r}_j)]$, with $E(\vec{r}_i) = E_i$ the emissivity in pixel i , is the *covariance kernel function*, for which we choose the common squared-exponential form:

$$k_{SE} = \sigma_f^2 \exp\left(-\frac{d^2}{2\sigma_l^2}\right), \quad d = \|\vec{r}_i - \vec{r}_j\|. \quad (2.5)$$

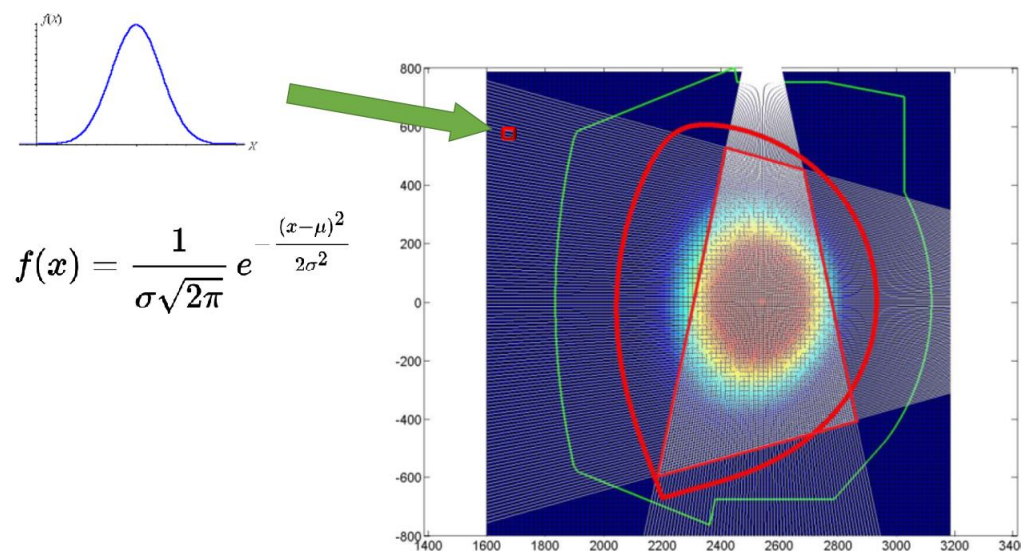


Fig. 2.1. Tomographic capabilities of the WEST SXR system based on GEM detectors. The horizontal camera views along 128 lines-of-sight. The vertical camera is inside the vertical port and is coupled to 75 sight lines. In the Gaussian process framework, the emissivity in each cell follows a Gaussian distribution, while the joint distribution of every subset of pixels is multivariate normal. This imposes structure on the emissivity field, avoiding wildly fluctuating emissivity in neighboring cells.

In turn, the kernel function depends on two parameters σ_f and σ_l , referred to as the signal standard deviation and length scale, respectively. In Bayesian terminology, any parameters of the prior distribution are called *hyperparameters*. Here, the

key hyperparameter is σ_l , as it determines the smoothness of the prior emissivity field. On the other hand, σ_f sets the overall scale of the prior covariance matrix, and as such it determines the weight of the smoothing prior in the posterior covariance. Summarizing σ_f and σ_l by $\bar{\theta}$, the total inference problem can be written as

$$p(\bar{E}_n|\bar{d}_m, \bar{\theta}) = \frac{p(\bar{d}_m|\bar{E}_n, \bar{\theta}) \cdot p(\bar{E}_n|\bar{\theta})}{p(\bar{d}_m|\bar{\theta})} \sim p(\bar{d}_m|\bar{E}_n, \bar{\theta}) \cdot p(\bar{E}_n|\bar{\theta}), \quad (2.6)$$

where the prior is given by

$$p(\bar{E}_n|\bar{\theta}) = \frac{1}{(2\pi)^{\frac{n}{2}}|\bar{\Sigma}_E|^{\frac{1}{2}}} \exp\left[-\frac{1}{2}(\bar{E}_n - \bar{\mu}_E)^T \bar{\Sigma}_E^{-1} (\bar{E}_n - \bar{\mu}_E)\right]. \quad (2.7)$$

Here, $\bar{\mu}_E$ is the prior mean, which will be fixed at 0, or it may be chosen on the basis of earlier experiments or prior knowledge of the involved domain experts. In principle, the hyperparameters can be marginalized from the problem (i.e. integrated out), but this would greatly increase the computational complexity of the method. Instead, we will employ a common approximation wherein a fixed set of hyperparameters is determined by maximizing the evidence $p(\bar{d}_m|\bar{\theta})$, and plugging those estimates into Eq. (2.6). This procedure is motivated in the appendix.

The next step in the inference process consists of choosing a likelihood function $p(\bar{d}_m|\bar{E}_n, \bar{\theta})$, containing the forward model. Under the reasonable assumption of a normal distribution of the measurement uncertainty on the emissivity line integrals, described by the variable $\bar{\epsilon}$ in Eq. (2.1), the likelihood can be written as

$$p(\bar{d}_m|\bar{E}_n, \bar{\theta}) = \frac{1}{(2\pi)^{\frac{m}{2}}|\bar{\Sigma}_d|^{\frac{1}{2}}} \exp\left[-\frac{1}{2}(\bar{R} \cdot \bar{E}_n - \bar{d}_m)^T \bar{\Sigma}_d^{-1} (\bar{R} \cdot \bar{E}_n - \bar{d}_m)\right]. \quad (2.8)$$

Here, $\bar{\Sigma}_d$ is the covariance of the measured data, describing measurement uncertainty and correlation pertaining to the vector \bar{d}_m of measured line integrals. We will assume that the various line-integrated measurements are uncorrelated and choose a 5% noise level, based on previous experience at Tore Supra. Therefore,

$$\bar{\Sigma}_d = \begin{pmatrix} 0.05 \cdot d_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & 0.05 \cdot d_m \end{pmatrix}. \quad (2.9)$$

Finally, the posterior distribution reads up to a constant factor,

$$p(\bar{E}_n|\bar{d}_m, \bar{\theta}) \sim p(\bar{d}_m|\bar{E}_n, \bar{\theta}) \cdot p(\bar{E}_n|\bar{\theta})$$

$$\sim \exp \left[-\frac{1}{2} (\bar{\bar{R}} \cdot \bar{E}_n - \bar{d}_m)^T \bar{\bar{\Sigma}}_d^{-1} (\bar{\bar{R}} \cdot \bar{E}_n - \bar{d}_m) \right] \cdot \exp \left[-\frac{1}{2} (\bar{E}_n - \bar{\mu}_E)^T \bar{\bar{\Sigma}}_E^{-1} (\bar{E}_n - \bar{\mu}_E) \right]. \quad (2.10)$$

The major advantage of normal distributions and a linear forward model now becomes clear. Indeed, it follows from standard probability calculus that the product of two normal distributions is also Gaussian, with mean vector and covariance matrix given by

$$\bar{\mu}_E^{post} = \bar{\mu}_E^{prior} + \left(\bar{\bar{R}}^T \bar{\bar{\Sigma}}_d \bar{\bar{R}} + \bar{\bar{\Sigma}}_E^{-1} \right)^{-1} \bar{\bar{R}}^T \bar{\bar{\Sigma}}_d^{-1} (\bar{d}_m - \bar{\bar{R}} \cdot \bar{\mu}_E). \quad (2.11)$$

$$\bar{\bar{\Sigma}}_E^{post} = \left(\bar{\bar{R}}^T \bar{\bar{\Sigma}}_d \bar{\bar{R}} + \bar{\bar{\Sigma}}_E^{-1} \right)^{-1}. \quad (2.12)$$

The posterior mean is thus available in a closed form, representing a computationally light-weight estimate of the emissivity field. In addition, the diagonal elements of the posterior covariance matrix $\bar{\bar{\Sigma}}_E^{post}$ quantify the uncertainty on the inference result. It should be noted that the GPT method implemented here only uses the SXR line integral measurements and no assumptions whatsoever are made regarding the magnetic equilibrium. This renders the method very flexible, potentially allowing detection of structures in the emissivity field (e.g. local impurity concentrations) that do not show up in the equilibrium reconstruction. It also prevents misguided information to enter the SXR reconstruction process in case the equilibrium information is incorrect, e.g. during sawtooth activity.

2.4 GPT phantom test on WEST

2.4.1. WEST phantom data

Our implementation of the GPT algorithm has been validated using phantom data, i.e. a set of synthetic SXR emissivity fields characterized by various emissivity patterns, some of which create a challenging tomography problem. Line integrals with added noise were calculated, followed by reconstruction of the emissivity field by GPT and comparison with the original emissivity phantom.

Four different shapes were used for the phantom tests, corresponding to various situations that are expected to be relevant for WEST SXR emission: Gaussian shape, hollow shape, left-right kidney shape and up-down kidney shape, as shown in [Fig. 2.2](#). The kidney shapes correspond to asymmetric impurity distributions, e.g. due to toroidal plasma rotation.

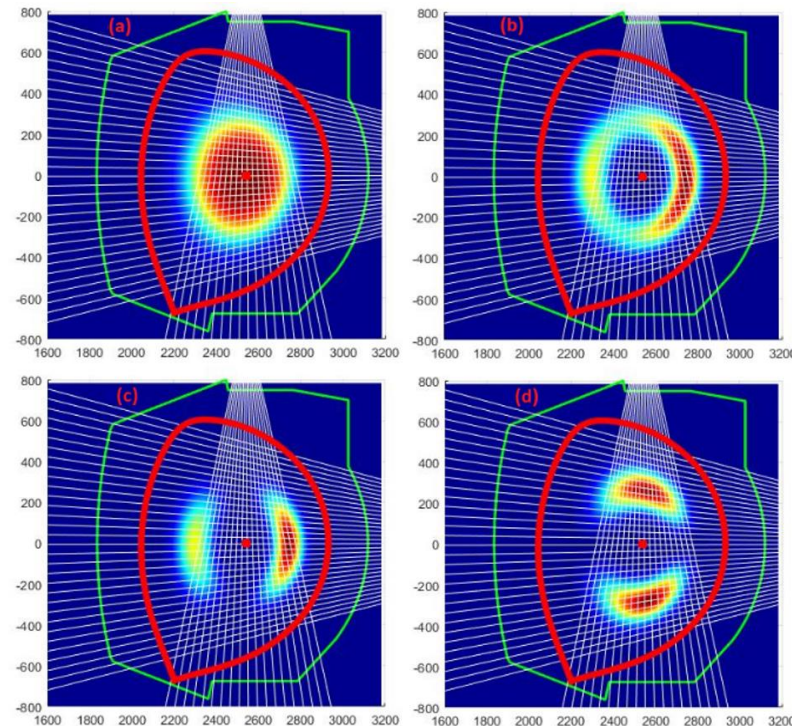


Fig. 2.2. Four phantom emissivity fields are used in our test: (a) Gaussian shape, (b) hollow shape, (c) left-right kidney shape, (d) up-down kidney shape. The green curve in the figure represents the vacuum vessel and the red curve is the last-closed flux surface.

2.4.2 Tomography results

The reconstructed emissivity fields based on line integrals with a 5% noise level are shown in Fig. 2.3. The quality of the reconstructions can be quantified through a relative error map, showing the difference between the phantom and reconstructed field, normalized by the maximum phantom emissivity:

$$\xi_i = \frac{|E_{n,i}^{(rec)} - E_{n,i}|}{\max\{\bar{E}_n\}}. \quad (2.13)$$

In case of the Gaussian shape, the maximum relative error is around 6.9%, 15% for the hollow shape, 12% for the left-right kidney shape and 15% for the up-down kidney shape. In general, more asymmetric emissivity fields are more difficult to reconstruct, the uncertainty level depending greatly on the coverage and field of view of the optical system. Nevertheless, in all cases the characteristic shape of the phantom is recovered relatively well by GPT. In addition, one can compare the line integrals obtained from the original phantom, with those calculated from the reconstructed emissivity field. As shown in Fig. 2.3, good agreement is achieved in all cases. The calculation is also fast: on a typical PC environment with Matlab, each time slice takes about 100 ms computing time.

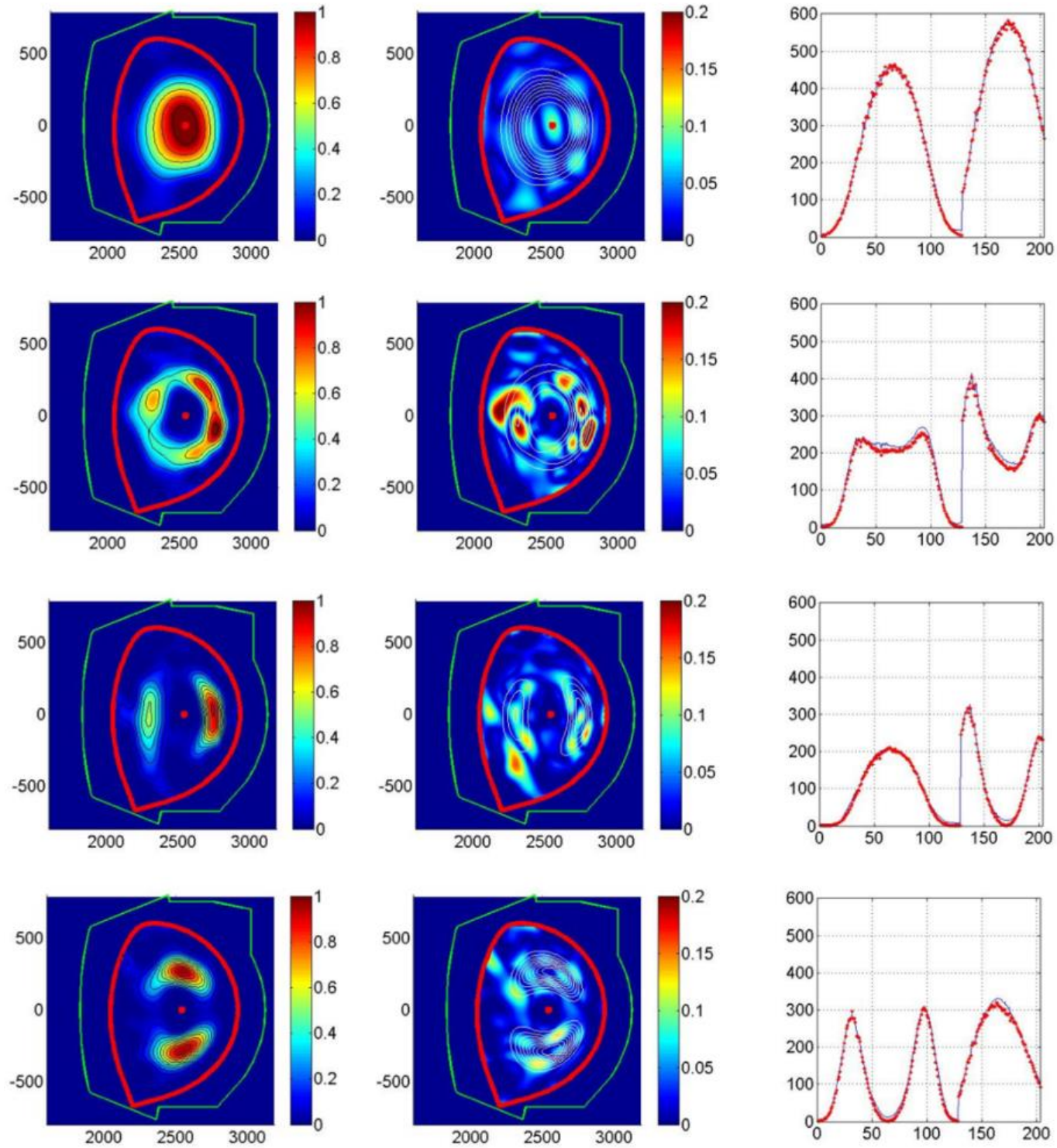


Fig. 2.3. GPT phantom test with 5% noise level. From left to right, the first column contains the reconstructions, the second column shows the relative error maps according to Eq. (2.13) (the white contours represent the original phantom), and the third column gives the comparison between the line integrals obtained from the phantom (red dots) and from the reconstructed emissivity fields (blue curves). Note that the phantom emissivity has been normalized for the benefit of numerical computation.

2.4.3 Gaussian process tomography uncertainty analysis

A valuable advantage of GPT is that it provides uncertainty estimates on the reconstructed emissivity field through the posterior covariance matrix; Eq. (2.12). Indeed, one expects the deviation of the reconstruction from the original phantom to be largest in those pixels where the posterior variance is large. This is confirmed by comparing the posterior variance map with the relative error field, as shown in Fig. 2.4. Naturally, the relative error field will not be available when performing tomography on real WEST data, but the posterior variance can still be calculated.

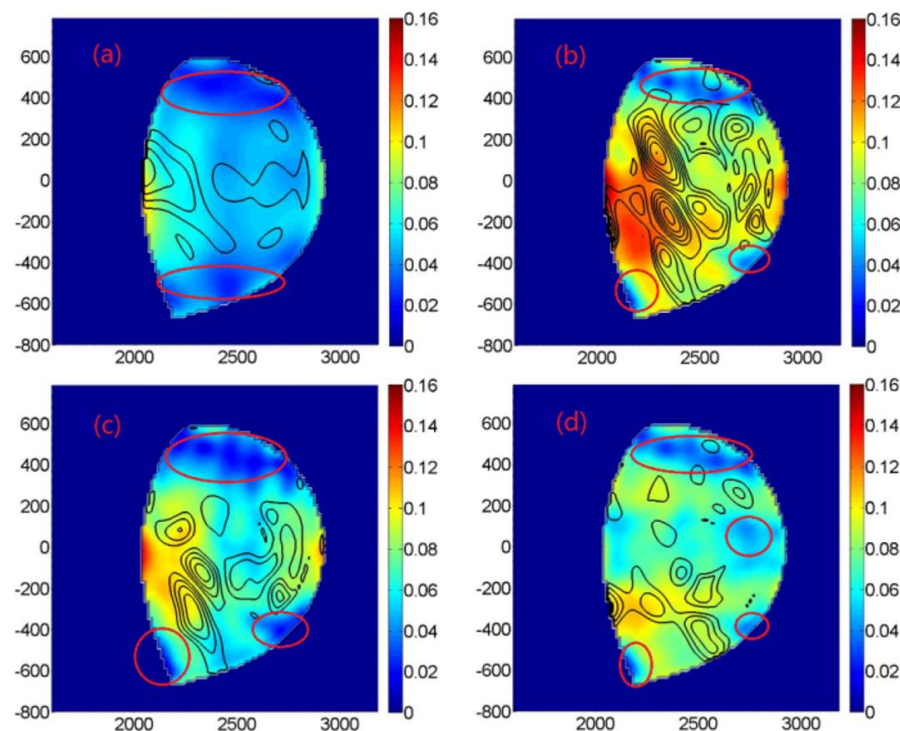


Fig. 2.4. Examples of a comparison between the posterior variance map (color map) and relative error map (black contours) on a 5% noise level: (a) Gaussian shape, (b) hollow shape, (c) left-right kidney shape, (d) up-down kidney shape. In the areas marked by red ellipses, both the posterior variance and relative error are low. (The units are arbitrary.)

2.5 GPT application to EAST SXR

A quantitative comparison between the minimum Fisher information (MFI) tomographic technique and GPT has been demonstrated for SXR data from Tore Supra in [Wang18a]. Here, we present a first application of GPT to SXR data obtained at EAST. A sawtooth-triggered kink mode has been observed, as shown in the reconstructed emissivity fields in Fig. 2.5. Note that this reconstruction incorporates information regarding the magnetic equilibrium, as described in [Wang18b]. The redistribution of plasma due to the sawtooth and the kink mode is clearly visible in the reconstructions.

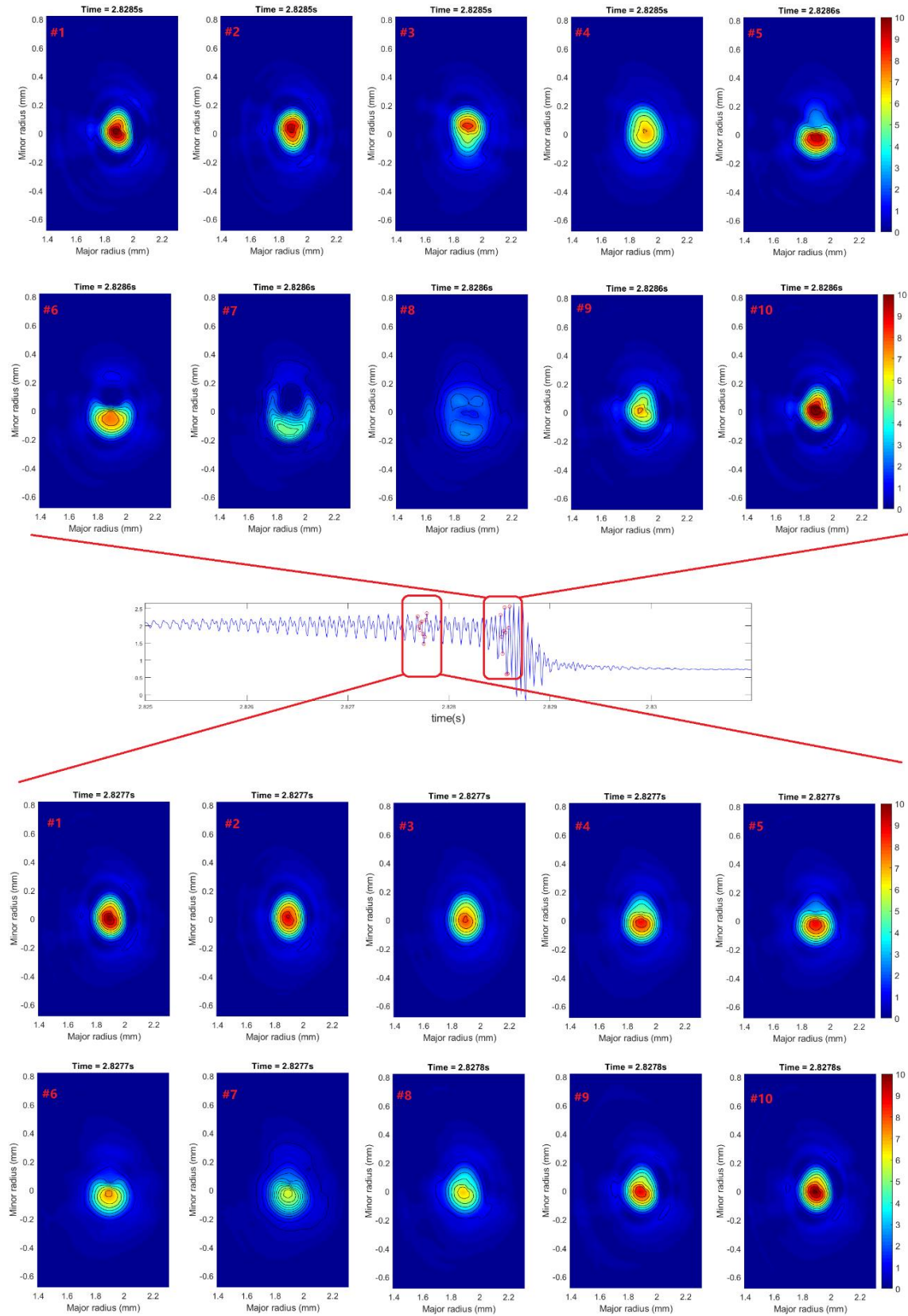


Fig. 2.5. Examples of GPT using data from SXR spectroscopy at EAST during a sawtooth-triggered kink mode. The time trace in the center panel corresponds to one of the SXR channels (units a.u.). The panels below present reconstructed emissivity fields during the sawtooth evolution, while the upper panels show the evolution of the kink mode, which is triggered by the preceding sawtooth.

3 Summary

In this article, some elementary notions from Bayesian data analysis have been presented and applications of SXR tomography on WEST and EAST have been shown. Bayesian analysis is well suited for solving tomographic problems on the basis of line-integrated data from fusion diagnostics. Using a flexible non-parametric Gaussian process as a prior distribution, Gaussian process tomography regularizes the reconstruction without the need for additional information regarding the magnetic equilibrium. Moreover, the Bayesian formalism provides uncertainty estimates on the reconstructed SXR emissivity field, through the posterior distribution. This brings into reach the possibility of an automated self-check of the reconstruction performance, which could be very useful in future devices. Owing to its reliance on Gaussian distributions, a further important advantage of the GPT method is its computational efficiency, rendering the tomography sufficiently fast for real-time applications. Thus, in view of its many advantages over traditional tomographic reconstruction techniques, Gaussian process tomography in the Bayesian framework has excellent potential in fusion science.

Appendix

The choice of suitable hyperparameters is a key issue for the GPT method, as they determine the degree of smoothness of the reconstructed emissivity field. A full Bayesian analysis would determine the hyperparameters together with the emissivity parameters, but this would not be feasible in real time. As a workaround, we determine the hyperparameters $\bar{\theta}$ from the data by maximizing the evidence. The rationale is that the marginal posterior for the hyperparameters (i.e. with the parameters \bar{E}_n marginalized) can be written as

$$p(\bar{\theta}|\bar{d}_m) \sim p(\bar{d}_m|\bar{\theta}) \cdot p(\bar{\theta}). \quad (\text{A. 1})$$

$\bar{\theta}$ *hyperparameters*

\bar{d}_m *GEM measurements*

Now, assuming a non-informative uniform hyperprior distribution $p(\bar{\theta})$ we see that the posterior for the hyperparameters is proportional to the evidence $p(\bar{d}_m|\bar{\theta})$, which also occurs in Eq. (2.6). Hence, by maximizing the evidence w.r.t. $\bar{\theta}$, we find the maximum *a posteriori* estimates of the hyperparameters. Using these values in the posterior Eq. (2.10) is usually a good approximation to the full Bayesian solution. From Eq. (2.3), the marginal posterior for the hyperparameters is given by

$$p(\bar{\theta}|\bar{d}_m) \sim p(\bar{d}_m|\bar{\theta}) = \int p(\bar{d}_m|\bar{E}_n, \bar{\theta}) p(\bar{E}_n|\bar{\theta}) d\bar{E}_n. \quad (\text{A. 2})$$

This results in the following expression:

$$\log(p(\bar{\theta}|\bar{d}_m)) = -\frac{1}{2}\left\{m \log(2\pi) + \log\|\bar{\bar{\Sigma}}_d + \bar{\bar{R}}^T \bar{\bar{\Sigma}}_E \bar{\bar{R}}\| + \bar{d}_m^T (\bar{\bar{\Sigma}}_d + \bar{\bar{R}}^T \bar{\bar{\Sigma}}_E \bar{\bar{R}})^{-1} \bar{d}_m\right\}. \quad (\text{A. 3})$$

The hyperparameters are contained in $\bar{\bar{\Sigma}}_E$, see Eq. (2.4). An example of the evidence as a function of the two hyperparameters is given in Fig. A.1. The data were obtained from a hollow shape phantom test and the optimization results in a length scale $\sigma_l = 144$ mm and signal standard deviation value $\sigma_f = 0.2427$.

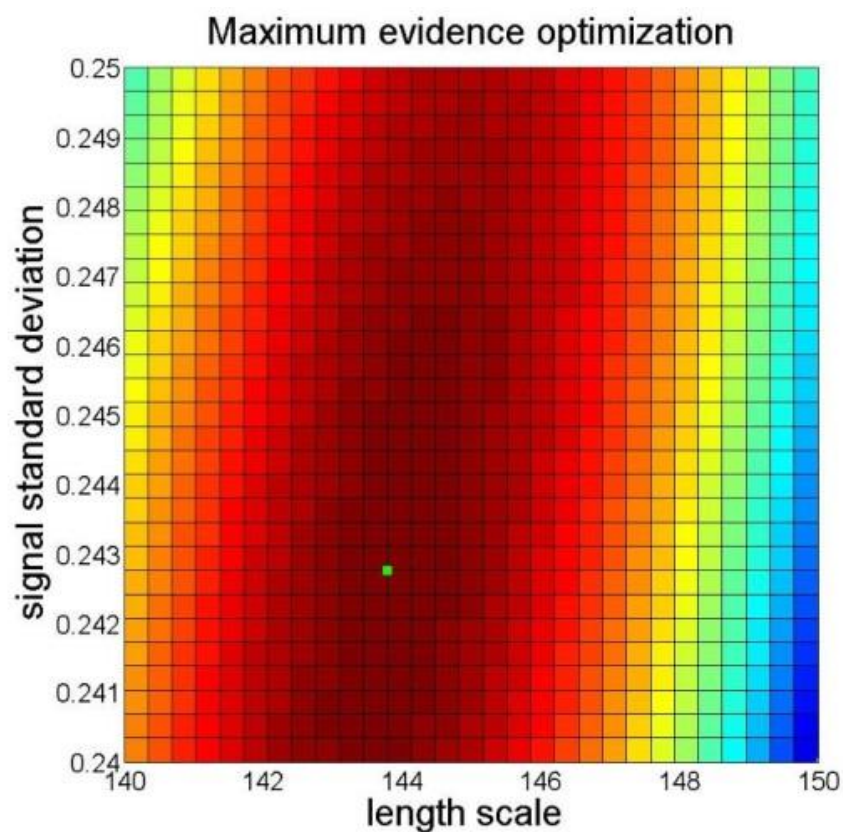


Fig. A.1. An example of $\log(\text{evidence})$ maximization for a 5% noise level using a hollow shape phantom test. The maximum is obtained at the green point.

References

- [Anton96] M. Anton et al, X-ray tomography on the TCV tokamak, 1996 Plasma Phys. Control. Fusion 38 1849.
- [Cormack64] A. CORMACK, “Representation of a Function by Its Line Integrals, with Some Radiological Applications. II,” J. Appl. Phys., 35, 2908 (1964).
- [Cox61] R.T. Cox, *The Algebra of Probable Inference*, Johns Hopkins Press, Baltimore, MD, 1961.

- [Ertl98] K. Ertl, W. von der Linden, V. Dose and A. Weller, *Maximum entropy based reconstruction of soft X-ray emissivity profiles in W7-AS*, Nuclear Fusion, Volume 36, Number 11, 1998.
- [Fischer03] R. Fischer *et al.*, “Bayesian modelling of fusion diagnostics,” *Plasma Phys. Control. Fusion*, vol. 45, pp. 1095-1111, 2003.
- [Fischer04] R. Fischer *et al.*, “Integrated data analysis of fusion diagnostics by means of the Bayesian probability theory,” *Rev. Sci. Instrum.*, vol. 75, pp. 4237-4239, 2004.
- [Fischer10] R. Fischer *et al.*, “Integrated data analysis of profile diagnostics at ASDEX Upgrade,” *Fusion Sci. Technol.*, vol. 58, pp. 675-684, 2010.
- [Gelman13] A. Gelman, J.B. Carlin, H.S. Stern, D.B. Dunson, A. Vehtari and D.B. Rubin, “Bayesian Data Analysis,” 3rd Edition, Chapman & Hall/CRC, Boca Raton, FL, 2013.
- [Gilks96] W.R. Gilks, S. Richardson and D.J. Spiegelhalter, eds., *Markov Chain Monte Carlo in Practice*, Chapman & Hall/CRC, Boca Raton, FL, 1996.
- [Gregory05] P. Gregory, *Bayesian Logical Data Analysis for the Physical Sciences: a Comparative Approach with Mathematica Support*, Cambridge University Press, Cambridge, 2005.
- [Jardin16] A. Jardin, D. Mazon *et al.* “Tomographic capabilities of the new GEM based SXR diagnostic of WEST” Volume 11, 2016
- [Jaynes04] E.T. Jaynes, *Probability Theory. The Logic of Science*, Cambridge University Press, Cambridge, 2004.
- [Mazon12] D. Mazon *et al.* “Soft x-ray tomography for real-time applications: present status at Tore Supra and possible future developments”, *Review of Scientific Instruments* 83, 063505 (2012)
- [Mlynar15] J. Mlynar *et al.*, “Soft X-ray tomographic reconstruction of JET ILW plasmas with tungsten impurity and different spectral response of detectors”, *Fusion Engi. And Design*, Volumes 96–97, October 2015, Pages 869-872
- [Rathgeber10] S. Rathgeber *et al.*, “Estimation of profiles of the effective ion charge at ASDEX Upgrade with Integrated Data Analysis,” *Plasma Phys. Control. Fusion*, vol. 52, art. no. 095008 (16 pp.), 2010.
- [Sivia06] D.S. Sivia and J. Skilling, *Data Analysis: a Bayesian Tutorial*, second edition, Oxford University Press, Cambridge, 2006.
- [Svensson11] J. Svensson, Non-parametric tomography using Gaussian processes. JET Internal Report EFDA-JET-PR(11)24 (2011).
- [Verdoolaege10] G. Verdoolaege, *et al.*, “Potential of a Bayesian Integrated Determination of the Ion Effective Charge via Bremsstrahlung and Charge Exchange Spectroscopy in Tokamak Plasmas,” *IEEE Trans. Plasma Phys.*, vol. 38, pp. 3168-3196, 2010.
- [von der Linden14] W. von der Linden, V. Dose and U. von Toussaint, *Bayesian Probability Theory. Applications in the Physical Sciences*, Cambridge University Press, Cambridge, 2014.
- [Wang18a] T. Wang *et al.*, *Gaussian process tomography for soft-X ray spectroscopy at WEST without equilibrium information*, *Review of Scientific Instruments* **89**, 063505 (2018).
- [Wang18b] T. Wang *et al.*, Incorporating magnetic equilibrium information in Gaussian process tomography for soft X-ray spectroscopy at WEST, submitted to *Review of Scientific Instruments* (2018).