Specialization Project

École polytechnique fédérale de Lausanne

Master's Degree in Physics

Applicability of Bayesian Optimization using Gaussian Processes to the reconstruction of the hard x-rays spectrum emitted by runaway electrons in TCV

Author:
Francesca Orsi

Supervisor:
Stefano Coda

Co-Supervisors:
Alessandro Pau
Gino Marceca

EPFL

Contents

1 Abstract

Standard techniques used to measure runaway electrons emission in the hard x-rays range are often inaccurate because they don't account for the generally dominant effect of Compton interactions between the photons and the detector. This effect can lead to wrong reconstructions of the bremsstrahlung spectra, particularly in the high-energy range, which affects our predictions of the runaway electrons energy distribution. However, the exact initial energy of the photon that interacts with Compton scattering with the detector is impossible to retrieve due to the intrinsic probabilistic nature of the process. In this work, we propose a new approach to the reconstruction of the hard x-ray spectrum from the signal measured by the detector, which takes into account the possibility of Compton scattering and pair production. This is based on the reasonable assumption that the bremsstrahlung spectrum can be roughly approximated by an exponential defined by two parameters, that we recover using Bayesian Optimization based on Gaussian Processes. The method relies on a simulation of the detector response at high energies, and in principle can be used also in the case of pile-ups and saturation of the detector. This technique is developed and applied to real measurements of the runaway bremsstrahlung emission measured by the hard x-rays cameras installed at TCV, and the temporal evolution of the electrons mean kinetic energy and the rate of photon emission is estimated.

2 Introduction

Magnetic confinement fusion is one of the most promising alternatives to non-renewable sources for energy production, both for the great availability of the fuel and for the high energy density that we can produce. One of the main challenges in this field is the study of the so-called runaway electrons (RE), which can be accelerated to significantly high energies -of the order of MeV in medium-size tokamaks – and, in the case of loss of confinement, risk to damage in the inner parts of the machine.

2.1 Runaway Production and Emission

The production mechanism of these electrons was first described by Dreicer [4].They are initially produced when the plasma is subject to an intense electric field, stronger than a threshold value called the Dreicer field. In this case, the electrons feel an acceleration that is not compensated by the friction due to electrons-ions collisions and therefore accelerate to very high energies. This mechanism is called *primary runaway generation*, to be distinguished from the *runaway avalanche*, which is a secondary generation that happens after an initial runaway beam is produced. In fact the relativistic electrons can collide with bulk electrons, transferring them a big part of their momentum, but remaining themselves in the relativistic regime, thus causing a growth of the runaway population[10].
Once formed, the runaway electrons slow down in two ways: emitting synchrotron radiation with the gyromotion around the magnetic field lines and through collisions between plasma particles or wall components [21]. The former is emitted mainly in the infra-red range [12], while the latter is bremsstrahlung radiation that is emitted in the range of hard x-rays (HXR). Both these radiations can be used to extract some information about the state of the runaway beam[11].

2.2 Theoretical and Experimental works

One of the main challenges in the field of runaway electron studies is the reconstruction of the 6-D phase space dynamics, which is particularly complicated due to the many physical processes involved. Research in this direction proceeds mainly on two sides: on one hand, kinetic simulations try to reproduce the plasma dynamics from first principles, and on the other hand we try to extract information from measurements of synchrotron radiation emission and bremsstrahlung. At the state of art of the research we still lack a reliable technique which tells us precisely the energy distribution of runaway electrons throughout all the plasma discharge.

Kinetic simulations have been implemented in the codes LUKE [13] and in CODE [20]. These simulations start from direct measurements of plasma parameters, such as plasma current and loop voltage, to reconstruct the dynamics of particles of the shot, in order to predict the runaway energy distribution and their radiation. The simulations are based on solving a steady-state Fokker-Planck equation, where the drift and the diffusion coefficient are calculated based on some models which try to take into account all the relevant physics[20]. This is an extremely complicated task due to the several factors affecting the runaway dynamics: from the large-angle collisions between existing runaways and bulk electrons, to the radial transport of RE, to radiation mechanisms and

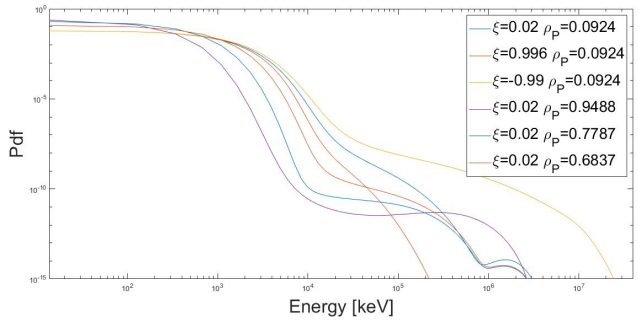


Figure 1: Runaway energy distributions predicted by LUKE for shot #61081 at TCV tokamak at time t=0.8s, as a function of different pitch angles and poloidal radiuses

other characteristics such as magnetic field inhomogeneities. In general the simulations tend to agree that the runaway energy distribution is not monotonically decreasing but it presents a bump-on tail at high energies [20]. The exact position of the tail has been estimated in [3]. Simulations from [19] and from LUKE also show that the energy spectrum of the bremsstrahlung emission has mainly an exponential form, with some deviation from this trend at high energies, as shown in Figure 1.

On the other hand, we can also try to extract information about the runaway beam from experimental measurements of their bremsstrahlung emission. Measurements of hard x-rays are usually performed with either semiconductor detectors, such as the CdTe detectors installed at TCV [8] tokamak, or with scintillator detectors as those installed at JET [26] and AUX[18]. In order to interpret this kind of measurements, however, we have to rely on some assumptions that are not always true. The main problem that we face when measuring high-energy photons is that, above a certain energy threshold, the most probable interaction between photon and detector is Compton scattering. When this happens, the energy deposited on the detector is only a part of the energy of the incoming photon, and it becomes impossible to know the initial photon energy due to the intrinsic probabilistic nature of the process. Most of the measurements that we perform today disregard the problem and consider the height of the signal measured by the detector simply proportional to the energy of the incoming photon, affecting the reliability of the measurements at high energies. This problem could be mitigated by increasing the size of the detector in such a way to rely on multiple Compton scatterings, however this solution would increase the noise caused by the neutrons, and a trade-off must be found [21]. Moreover, these measurements rely also on the hypothesis that there are no pile-up events and they cannot give any information when the incoming energy is high enough to saturate the detector.
In this work, we want to propose a radically different method to reconstruct the incoming energy spectrum from the measurements of the hard x-rays cameras installed at TCV. This method takes into account the possibility of Compton scattering, pair production, pile ups, and is applicable also when the detector saturates. Using a simulation of the detector response for energies up to hundreds of MeV, which models all of these possibilities, we can use a machine learning technique called Bayesian optimization to reconstruct the exponential part of the photon energy spectrum.

3 Experimental Method

3.1 TCV tokamak and Hard X-Rays Cameras

The Tokamak à Configuration Variable (TCV) is a medium-size tokamak which has the characteristic of allowing many different plasma shapes, thanks to the introduction of internal control coils alongside the toroidal and poloidal coils. Their introduction allows greater control on the plasma instabilities when the plasma is elongated, thus permitting to reach a very wide range of plasma shapes, such as the *snowflake* divertor configuration, obtained for the first time at TCV in 2010 [22]. The tokamak has a major radius of 0.88m and an aspect ratio $\epsilon \simeq 3.5$, and can reach a value of toroidal field up to 1.54T and a plasma current $I_p < 1$MA. Plasma heating is carried out with Ohmic heating and with two additional systems: the Electron Cyclotron Resonance Heating (ECRH), which relies on the wave-particles interaction to accelerate the electrons, and a Neutral Beam Injection (NBI), available at TCV since 2015, that injects very energetic neutral particles into the plasma which are then ionized and drive additional current.
In particular, the EC heating system has the effect of generating a population of suprathermal electrons, with energies up to hundreds of keVs, which then emit bremsstrahlung in the hard x-rays range due to collisions with other plasma particles. Suprathermal electrons' hard x-rays emission at TCV tokamak is measured with four hard x-rays semiconductor cameras with different poloidal view of the plasma. Each camera is composed of 25 CdTe detectors of dimensions 2x2x2 mm with different radial orientations, and each detector is shielded by a lead collimator. This setup allows for tomographic reconstruction of suprathermal electrons' dynamics with a time resolution of the order of ms. The cameras are studied to work with photon energies up to 200keV, since below this value the most probable photon-detector interaction is photoelectric interaction [8].

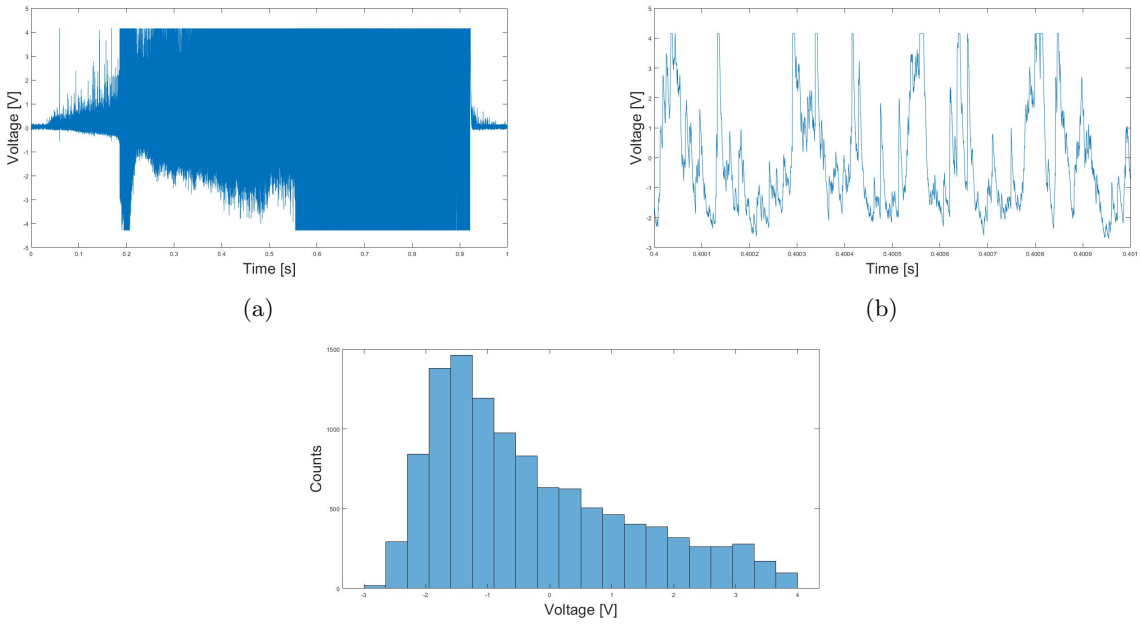


Figure 2: Voltage signal measured during runaway discharge at TCV tokamak for shot #63420 by the detector in channel 10 of camera C4 (a) full time signal (b) zoom between 0.400-0.401s (c) distribution of the non-saturated signal between 0.400-0.401s

3.2 Camera Signal during Runaway Discharge

During runaway discharges, the bremsstrahlung emission reaches energies much higher than 200keV, and standard inversion techniques cannot be applied due to the high probability of Compton scattering, to detector saturation and to the loss of collimation of the incoming photons, which are enough energetic to pass though the lead collimators that shield the detectors. A typical camera signal during runaway discharges in TCV shown in Figure 2a.

Nevertheless, if we zoom in at 0.4s and look at 1ms of signal, we realise that the saturation rate is actually discretely low, and that the signal assumes a certain distribution, from which we could possibly still extract information. These images are shown in Figures 2b and 2c.

In previous works the detector response at high energies was built, where the interaction between photon and detector was modelled with Monte Carlo methods and it was shown that the distribution of voltage signal can give some information on the energy of the incoming photons, even in the case when the signal saturates. **The idea of this work was to find a machine learning technique that could learn the association of a certain voltage signal with some incoming energy spectrum from the simulation of the detector, and then apply this when looking at a real measured signal, returning the shape of the energy spectrum.** This way, we can analyse measured signals where we have pile ups, saturation and Compton scattering, because all of these possibilities are accounted for in the simulation of the detector response. In order to do this, we have to find a way to compare two voltages distribution, to be able to know if they come from the same underlying photon distribution impinging on the detector. Moreover, to simplify the research of energy spectra, we assume that they have an exponential shape of the form

$$f(E) = N \cdot exp(-E/E_0) \tag{1}$$

which is a first approximation of the shapes documented in both theoretical [14][19] works and experimental measurements [17] [25] [7]. The energy spectrum impinging on the detector becomes then a function of two parameters, the rate N and the mean kinetic energy of runaway electrons E_0, and varying these two we

want to find a voltage distribution created from the simulation the most similar possible to the measured voltage distribution. The technique we used to find the values of these two parameters that best reproduce the measured voltage signal is a machine learning technique called Bayesian Optimization based on Gaussian Processes.

3.3 Gaussian Processes Regression and Bayesian Optimization

A Gaussian random variable y is a random variable whose probability density function has a Gaussian distribution with mean μ and variance σ^2:

$$p(y) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(y-\mu)^2}{2\sigma^2}} \tag{2}$$

A Gaussian random variable can be written as $y = N(\mu, \sigma^2)$.
A Gaussian Process (GP) $f(x)$ is a stochastic process where the function at each point of the domain $x \in R^n$ is a Gaussian random variable, with mean and variance depending on the point of the domain: $f(x) = N(\mu(x), \sigma^2(x))$.

Gaussian process regression is a technique to estimate the value of a black-box function f(x) at any point of the domain $x \in R^n$, assuming f(x) to be a Gaussian process. Then, the value of f(x) in all the domain can be calculated given some evaluations of the function, evaluations which can be affected by stochastic white noise.
The problem can be formulated writing the function as

$$f(x) = \phi(x) \cdot \omega \tag{3}$$

where $\phi(x)$ is a basis function vector, for example the functional space of the powers of x $\Phi(x) = (1, x, x^2, x^3, ..., x^n)$, and ω is a vector of constants that we call *weights*. Let's assume to have some evaluations of the function at points $\{x_i^*\}_{i=1..N}$ that we call

$$y^* = f(x^*) = \phi(x^*) \cdot \omega + \epsilon \tag{4}$$

where $\epsilon = N(0, \sigma_n^2)$ is Gaussian white noise. Our goal is to reconstruct the posterior probability of f(x) at any point given the evaluations $\{x^*, y^*\}$.
The technique of Gaussian process regression can be summarized in three main steps: initially, we make an assumption on the probability distribution of the weights, which is called our prior probability. Then, from the noisy evaluations of the function, we use Bayes' theorem to calculate the posterior probability on the weights, which simply results in a multivariate Gaussian distribution thanks to the assumption that f(x) is a Gaussian process. Lastly, we estimate the probability distribution of f(x) at every point of the domain with the predictions that we have made on the weights.
More in detail, from the observations $\{x^*, y^*\}$, we can calculate the values of the weights with Bayes' theorem: the posterior on the weights is given by the likelihood, the prior on the weights and the marginalised probability:

$$p(\omega|y^*, x^*) = \frac{p(y^*|x^*, \omega)p(\omega)}{p(y^*|x^*)} \tag{5}$$

The prior is a probability distribution on the weights that expresses the knowledge that we have before the evaluation of the function: it is often set by default as a multivariate Gaussian distribution with zero mean and covariance function on the weights Σ_p: $\omega = N(0, \Sigma_p)$. The marginal likelihood functions only as a normalisation constant and does not depend on the weights. The likelihood, instead, measures the probability of measuring y^* at point x^*. The peculiar thing about Gaussian Processes is that the likelihood is particularly easy to compute, being the product of Gaussian probabilities, and therefore being itself normally distributed:

$$\begin{aligned} p(y^*|x^*, \omega) &= \Pi_{i=1}^N \frac{1}{\sqrt{2\pi}\sigma_n} e^{-\frac{(y_i^* - \phi(x^*)\omega)^2}{2\sigma_n^2}} = \\ &= N(\phi(x^*) \cdot \omega, \sigma_n^2 \cdot I) \end{aligned} \tag{6}$$

The posterior probability on the weights therefore results in a multivariate Gaussian distribution with updated mean and variance based on the observations.
After the posterior probability distribution on the weights is calculated, the posterior probability of f(x) after the observations is estimated by averaging each value of the weights on their probability [24]:

$$p(f(x)|x, y^*, x^*) = \int d\omega \, p(f(x)|x, \omega) \, p(\omega|y^*, x^*) = N(\bar{f(x)}, \sigma^2_{f(x)}) \tag{7}$$

As we can see, at each point the Gaussian process f(x) is a Gaussian random variable with a mean and variance which depend on the estimated values of the weights and on the noise level. This formulation of the problem is called the *weight space view* and it gives a more mathematical overlook of how the posterior probability is calculated.
The problem can be reformulated in a more familiar way, called the *function space view*, where we introduce a kernel function that contains the information about the weights, the function basis $\Phi(x)$ and the noise. The kernel function is easier to work with because it is defined as the covariance function of f(x):

$$k(x_p, x_q) = cov(f(x_p), f(x_q)) =< f(x_p)f(x_q) > \tag{8}$$

The kernel function depends itself on some parameters, called hyperparameters, namely the lengthscale, the variance of the function and the variance of the noise. Typical shapes of the kernel function are, for example, the squared exponential kernel: $k(x_p, x_q) = \sigma_f^2 e^{-\frac{(x_p - x_q)^2}{l}} + \sigma_n^2$, or the Matérn kernel [24].
An example of Gaussian process regression in one dimension is shown in Figure 3.
Given this information, we can estimate the mean and the variance of the function f(x) in all the domain.
Training a Gaussian Process means to find the best form of kernel and the best values of the hyperparameters in order to do a good estimation of the function.

Bayesian optimization (BO) is a techniqe used to find the global optimum of a black-box function which usually takes very long to evaluate and that can't be written down analytically. Bayesian Optimization based on Gaussian Processes relies on the assumption that this function is a gaussian process and that, given some evaluations, we can apply gaussian process regression to calculate the posterior of the function in all the domain. Since the evaluations are very costly, the goal of this technique is to find the optimum with the least number of evaluations of the function[5]. In order to do this, we introduce a so-called acquisition function, that calculates, based on the posterior, where to evaluate next in order to find the minimum. This is done by trying to minimize the *cumulative regret*, defined as

$$r_N = N f_{min} - \sum_{i=1}^{N} f(x_i) \tag{9}$$

which is minimized if we start to evaluate f_{min} with the least amount of evaluations [9]. There are different acquisition functions that can be used, the most popular being the Expected Improvement (EI) and the Lower Confidence Bound (LCB). These functions too depend on some parameters, which describe for example how flexible this function can be, and the training of the process includes also finding optimal values for these parameters. The library for Bayesian Optimization that was used in this work is Python *GPyOpt* library [2]. Bayesian Optimization is widely employed for parameters research in the field of Gravitational Waves detection, where it is used to minimize a function of up to 15 variables [27].

3.4 Application to Hard X-Rays Signal

In this section it is described how Bayesian optimization was applied to the problem of reconstructing the incoming energy spectrum on the HXR camera starting from the measured signal.

The goal is to find the energy spectrum that gives a signal as similar as possible to the measured one. The way this was done is to assume an incoming exponential energy spectrum depending on two parameters, N

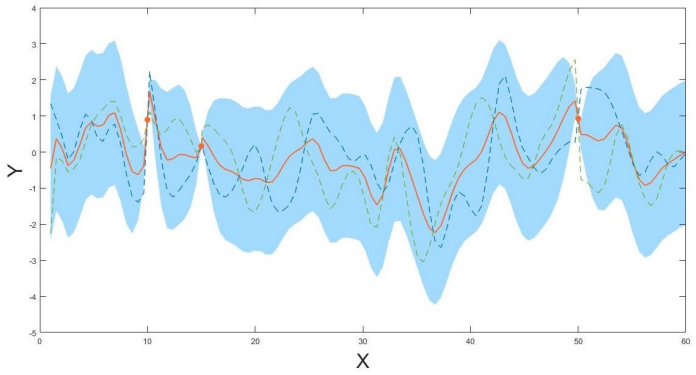


Figure 3: Example of gaussian process regression in 1D: the red dots are noiseless evaluations of the function, created randomly. The red line is the mean of the posterior and the blue area represents the standard deviation at each point. The dashed lines are samples drawn from the posterior (Excercise 2.9.1 from [24]).

and E_0, as described in section 3, and create through the simulation of the detector response the voltage signal that the camera would measure in response to this incoming spectrum. We create many voltage signals varying the two parameters and we choose the set of parameters for which the voltage signal is the closest to the measured one. Therefore, the black-box function to minimize needs to measure the distance between the measured voltage and the voltage output of the simulation, which depends on the parameters N and E_0. Since the simulation of the detector response is quite time-consuming, we want to find the best parameters set performing the least amount of simulations.

The technique that was chosen to compare two voltage signals in a certain time interval was to create the voltage distributions of the time signal and then compare the two distributions obtained. The chosen time interval was usually 1ms, because in runaway discharges the processes are very fast and the voltage distribution can significantly change from 1ms to the other. In Figure 4 we can see an example of how Bayesian Optimization works in one dimension: from an exponential bremsstrahlung emission with parameters $N = 4.5 \cdot 10^5$ and $E_0 = 400keV$, we build the voltage distribution measured by the camera when exposed to this radiation. Our goal is then to retrieve this parameters applying the optimization to the voltage distribution. This is done by creating many voltage signals varying the parameter E_0 and fixing N to the correct value, and comparing with the chi square statistics the initial voltage distribution and the simulated ones. By doing this, we find the red dots evaluations in Figure 4c. Starting from these dots, the program reconstructs the mean of the function, the standard deviation at each point, and plots below in red the acquisition function. Here, the value that found of $E_0^{found} = 430keV$, which is very close to the real one and produces a voltage distribution similar to the original, as we see in Figure 4b.
Many choices of the black-box functions were possible, since there exist many statistics to determine whether two distributions are sampled from the same underlying distribution or not [23]. The statistics that were considered are the χ^2, the mean squared error, the Kolmogrov-Smirinov statistics, the Cramér-von-Mises and the Anderson Darling tests [16] . All these functions give different importance to the mean of the distribution and to the tails, and therefore all of them were tested to see which one performed best in comparing voltage signals.

In order to do the training of the Gaussian Process and Bayesian Optimization to see which functions and parameters worked best, we performed some synthetic tests. We start from a phantom incoming energy distribution of which we choose the values of N and E_0 and create the corresponding voltage distribution. Then, starting from this, we apply the Bayesian Optimization to retrieve the values of the two parameters and we verify if the values we found were compatible to the ones we started with. With this technique, we were able to confirm that the spectrum that was found with the parameters found by Bayesian Optimization is in fact compatible within the experimental error to the real one we started from. Then, we chose the function and parameters which worked best and applied the technique to real measurements to find the time

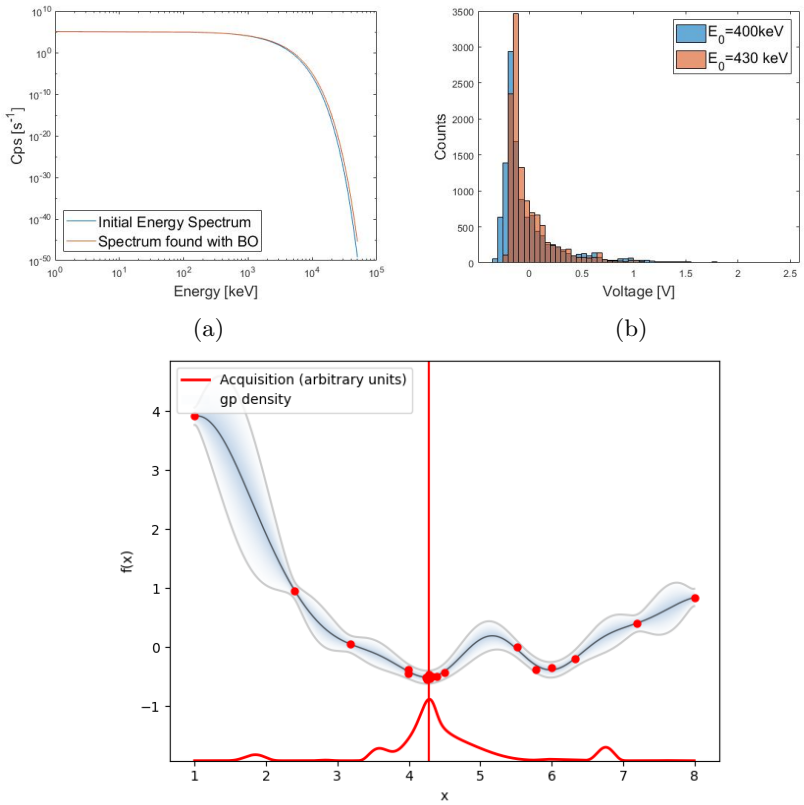
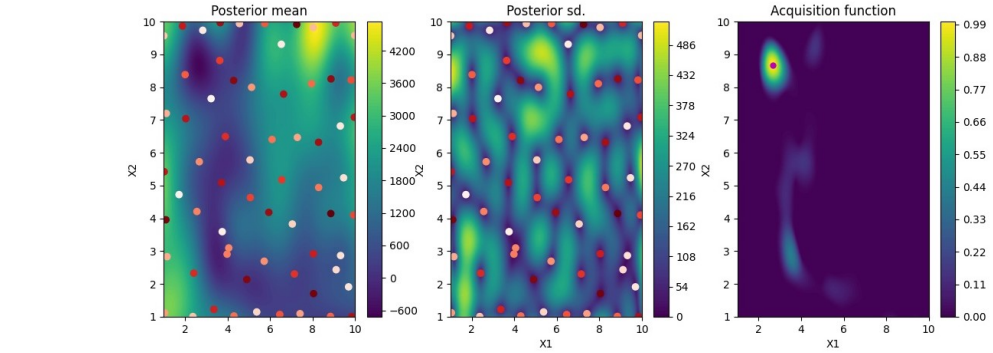


Figure 4: Bayesian Optimization in 1D: starting from the blue distribution in (a), we vary the parameter E_0 to find the voltage distribution that resembles the most to the blue one in (b), using the function chi square to compare the histograms. The evaluations of the function are shown in (c), together with the posterior and the acquisition function. The distribution obtained with the best found value of E_0 is shown in (b) and its proximity to the initial one is shown in (a).

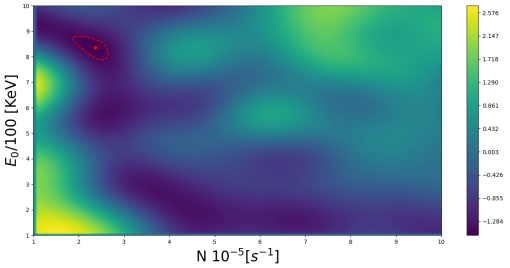
evolution of the bremsstrahlung spectrum emitted by the runaway electrons.

3.5 Estimation of the Error on the Parameters

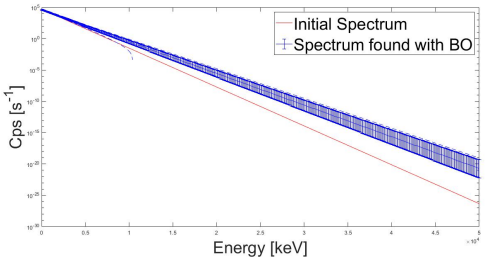
In the previous section it was explained how to use Bayesian Optimization in order to compare voltage signals and find the values of the best parameters for the rate and the mean kinetic energy of the bremsstrahlung emission. Here we want to explain the technique we used to calculate the error on the parameters, which is not directly given by the Bayesian Optimization. The result of this optimization, as we said, it's an estimate of the mean of the black-box function, also called objective function, and the variance at each point. In Figure 5a we report an example of the output of a 2D Bayesian Optimization: we have the posterior of the mean, of the standard deviation and the value of the acquisition function at each point.
The question to be asked when estimating error on parameters is: how far from the minimum do I have to move in parameters space such that the voltage distribution becomes enough different from the measured distribution? Since the value of f(x) in the minimum can oscillate between $f(x_{min}) \pm \sigma(x_{min})$, we can assume that the voltage distribution becomes different enough from the measured one when the mean at a point x equals the maximum value of f at the minimum, therefore $f(x) = f(x_{min}) + \sigma(x_{min})$. This way, the variability of f at the minimum is reflected on the error on the parameters and the width of the Gaussian distribution around the minimum mean value of f(x) provides us with an estimate of the resolution for our method.



(a)



(b)



(c)

Figure 5: Example of application of 2D Bayesian Optimization in a synthetic test. (a) Evaluations, mean, standard deviation and acquisition function of the Anderson-Darling statistics as a function of the parameters N and E_0. (b) Best value -red dot- and one-sigma error of parameters (c) Comparison between initial distribution and the one constructed with parameters found by Bayesian Optimization.

An example of error estimation is given in Figure 5b: here we represent the mean of the posterior distribution, its minimum and the red line corresponds to the values of f where $f(x) = f(x_{min}) + \sigma(x_{min})$. This line determines then our error on the parameters.

4 Results

In this section we report the results of the training of the Bayesian Optimization and its application to real measurements of TCV.
In the first part, we look for the best functions to use, in particular the objective function, the kernel and the acquisition function, as well as the best hyperparameters. The choice is made by starting to a phantom energy spectrum with known parameters N and E_0, and looking for which choice of functions and hyperparameters best retrieves the initial parameters, by comparing the voltage signals, as explained in Section 3.4.
In the second part we apply this technique to real measurements of the HXR emission of runaway at TCV, registered by the cameras described in Section 3, and we find the evolution of the bremsstrahlung emission in time during four different shots.

4.1 Training

The first part of the work is dedicated to the training of the Bayesian Optimization method. The first step that was done was to compare the different objective functions: our goal is to define a measure that quantifies the difference between two histograms. Many types of statistics exist that can perform this task: starting from the chi square, that compares the difference of each bin, to the men squared error, that gives more importance to the bins with many counts, to other methods that do not require a binning of the voltage

signal, such as the Kolmogrov-Smirinov method, the Cramér von Mises and the Anderson-Darling test, that compare the cumulative distribution functions of the two voltage distributions. Lastly, we considered also a statistics composed by the mean square difference of the momenta of the distribution,

$$f(V_s(N, E_0)) = \sqrt{\sum_{i=2}^{4} \left(\frac{E[(V_m - \bar{V_m})^i]}{\sigma_m^i} - \frac{E[(V_s - \bar{V_s})^i]}{\sigma_s^i}\right)^2} \tag{10}$$

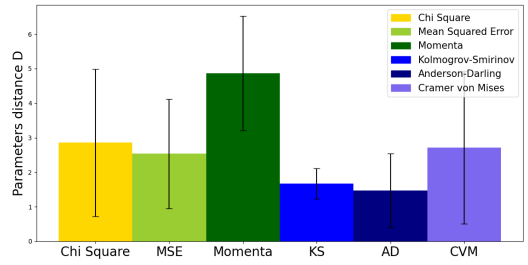
where V_m is the measured voltage distribution and V_s is the synthetic one created from given values of the parameters.
The comparison was done starting from some voltage signals created with known parameters of N and E_0 and then try to retrieve the value of the two with the different methods. The evaluation of the performance could be done in two ways: either to compare the distance between the found parameters and the real parameters, or to compare the distance between the energy spectra obtained with the found parameters and the original one. The first was done by measuring the Euclidean distance between the rescaled parameters:

$$D = \sqrt{\left(\frac{E^0_{real} - E^0_{found}}{100keV}\right)^2 + \left(\frac{N_{real} - N_{found}}{10^5 s^{-1}}\right)^2} \tag{11}$$

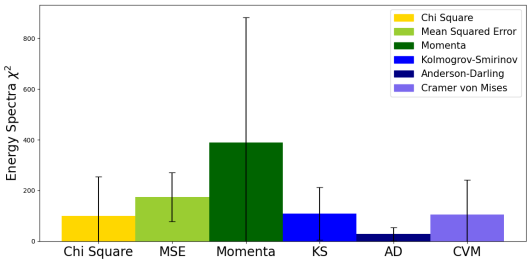
While the second way was simply done by measuring the chi square between the found energy distribution and the real energy distribution.

The results of this research for the best black-box function are shown in Figure 6: the synthetic research for parameters shows that the Anderson-Darling statistics finds results that are the closest to the real parameters (Fig. 6a), where the distance D is calculated like in Eq.11, and the energy distribution with the parameters found is by far the closest to the real distribution (Fig. 6b). The Anderson-Darling statistics was therefore chosen as the best objective function to use.

The following step is then to determine which kernel function works best with the Anderson-Darling statistics: in order to do this, we considered the three most used kernels, which are the Squared Exponential (SE), the Matérn 3/2 and the Matérn 5/2 kernels [24]. Looking at the results in Fig. 7a, the SE kernel was chosen.
The same was done with the acquisition function, where we considered the Expected Improvement (EI), the Lower Confidence Bound (LCB) and the Maximum Probability of Improvement (MPI) [2]. Since as we can see in Fig. 7b there is not a big difference in results among the three functions, he



(a)



(b)

Figure 6: Results of synthetic test to determine which objective function to use to compare the voltage distribution in Bayesian Optimization: (a) distance D defined in equation 11 to measure the distance between real and found parameters (b) χ^2 value between the initial energy spectrum and the one reconstructed with found parameters. The mean is calculated over different parameters sets.

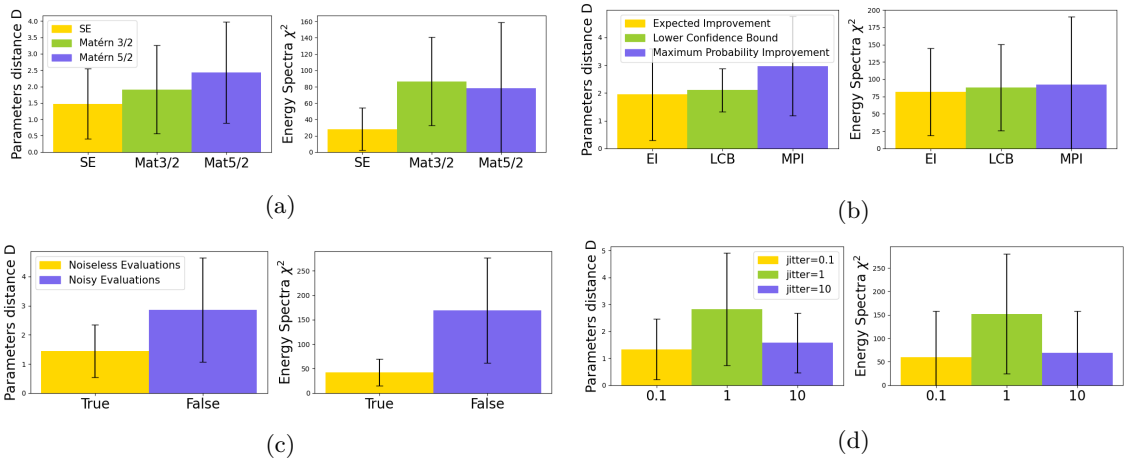


Figure 7: Different steps of the training of the Bayesian Optimization: different parameters are changed and the distance between initial and found parameters and spectra is reported. The different choices are (a) kernel (b) acquisition function (c) noise on evaluations (d) jitter

first the was chosen.

Then we passed to examinate other two parameters: the noise on the evaluations of the function and the parameter called *jitter*, which expresses the degree of exploration of the phase space. The first parameter goes in the GPyOpt library [2] under the name *exact function evaluation* ($exact\_feval$), and it can be true or false: this tells the optimization whether the evaluations contain noise or not. It turned out that turning off the noise lead to much better results, as shown in Fig. 7c.
Then, the parameter called jitter determines the degree of exploration of the parameters space: the bigger it is, the more distanced will be successive evaluations. The greater it is, the lower probability to be stuck in local minima, but at the same time it will give a less precise result on the parameters [9]. In our case it seems equally good to take a small one and a big one, therefore it was chosen equal to ten to escape local minima.

Lastly, we considered setting an initial value for the lengthscales in the two dimensions and the variance. However, since these values strongly depended on the value of the real parameters N and E_0, we decided to leave them free and let the optimization find the best values.

4.2 Precision of the method

Let's evaluate now the reliability of this method: how well can we effectively reconstruct the parameters and the initial energy spectrum. In order to do this, the procedure explained in Section 3.5 was applied to many parameters sets and three quantities were computed:

$$v_1 = \frac{N_{real} - N_{found}}{\sigma_N} \tag{12}$$

$$v_2 = \frac{E^0_{real} - E^0_{found}}{\sigma_{E^0}} \tag{13}$$

$$v_3 = \frac{1}{N_{bin}} \sum_{i=1}^{N_{bin}} \frac{h_{real}(i) - h_{found}(i)}{\sigma(i)} \tag{14}$$

where the first two quantities measure the distance between the initial parameters and those found with BO, while the last one measures the distance between initial and found spectra, where the error on each bin $\sigma(i)$ is derived from the error on the parameters N and E_0. The results are shown in Table 1. They show that the reconstructed spectrum always includes the real one within the 2σ error, while the effective minimum values of the parameters are often imprecise, especially the rate. The possible causes of this are analysed in Section 5.1. Even if this the best results of this technique don't give very precise results, we still decided to try to apply it to real camera measurements in order to estimate the order of magnitude of the parameters.

v_1	2.54±1.22
v_2	1.755±1.18
v_3	1.139±0.173

Table 1

4.3 Application to real measurements

Once the technique was established, it was applied to three different runaway shots of TCV: numbers 61081,63424,62420. We looked at the camera C4, which was oriented horizontally in the first shot and vertically in the others.

The shots were done during different experiments: in number 61081, whose reconstruction is shown in Figure 8, ECRH waves were injected in the plasma between $0.7 - 0.9s$. An example of the result of the reconstruction and the comparison between measured signal and signal constructed with the best parameters is given in Fig. 10.

The last two shots instead were carried out during an experiment aimed at trying to restart the plasma after a runaway discharge. In number 63420 Neon was injected between $0.55 - 0.95s$ and in 63424 Kripton at $0.55-0.95s$. In Figure 9 we report the evolution in time of the plasma current, the loop voltage and the parameters of the bremsstrahlung spectra obtained with the technique described above. An interesting feature we can immediately see is that the signal in the different chords is very similar when the camera is oriented horizontally, while it is significantly different when the orientation is vertical. Examples of the signals for the different chords are shown in Fig. 11, together with the proximity of each chord to the magnetic axis.

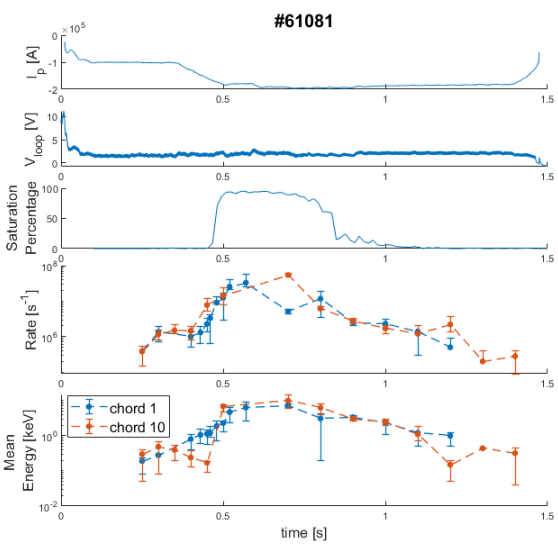
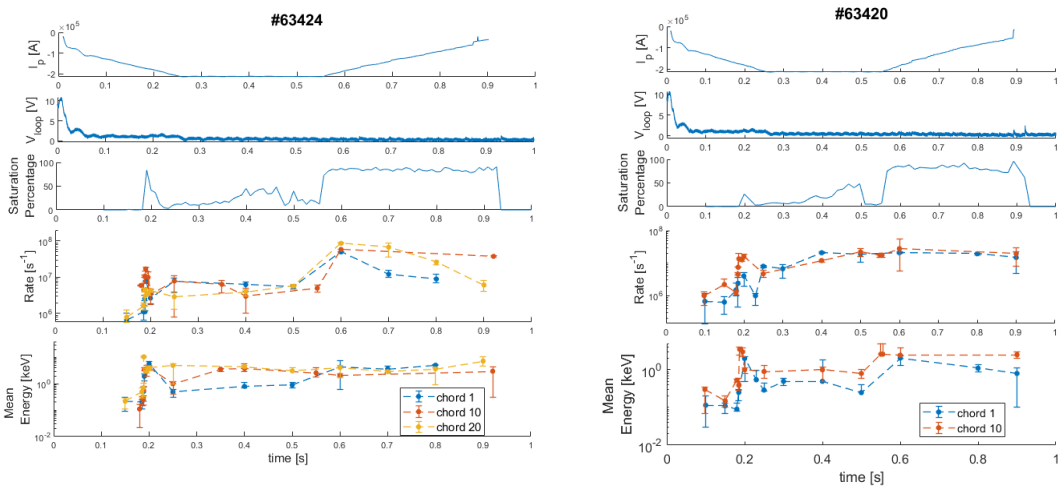
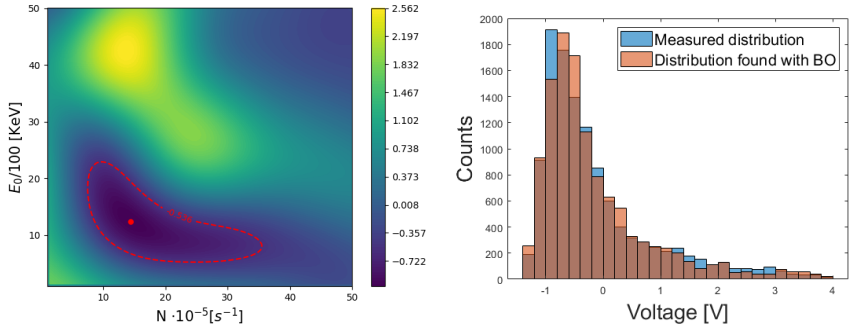


Figure 8: Evolution in time of plasma current, loop voltage, as well as rate and mean electrons energy found with BO for runaway shots #61081



(a) (b)

Figure 9: Evolution in time of plasma current, loop voltage, as well as rate and mean electrons energy found with BO for runaway shots #63424 (a) and #63420 (b).



(a) (b)

Figure 10: Reconstruction of shot #61081 at time $t \in [1.100, 1.101]s$. In (a), the minimum in parameters space found by Bayesian Optimization with the error on the parameters, in (b) the comparison between the measured voltage distribution and the voltage distribution constructed from the results of the best parameters.

5 Discussion

Bayesian Optimization research was used initially to find the best parameters from synthetic voltage distributions obtained by some initial known parameters and then applied to real measurements during runaway discharges.

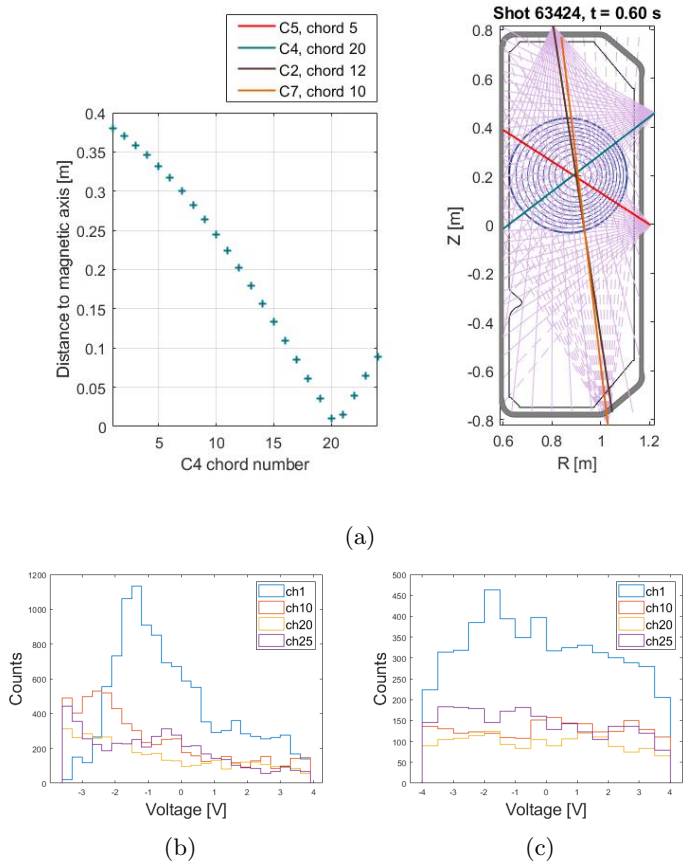
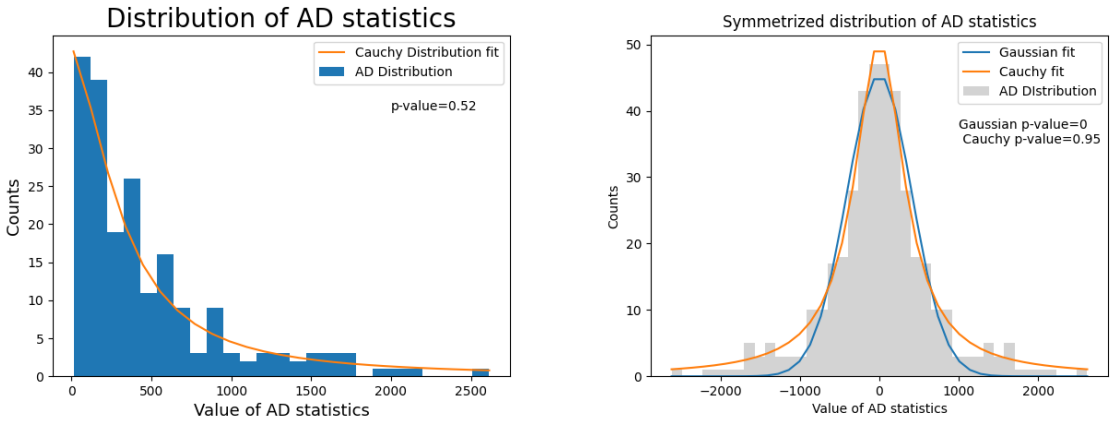


Figure 11: Details of the different chords of camera C4 during shot #63424: on top, the distance of every chord from the magnetic axis, on the right the line of sight of the closest chord. Below, the measured voltage distributions at different times at (b) $t = 0.400s$ and (c) $t = 0.900s$

5.1 Bayesian Optimization

In the first part, it was shown how the Anderson-Darling objective function was the best choice to compare two voltages distribution, probably for its characteristic of giving more importance to the tails, and the other parameters were optimized after that.
We proved that, after the optimization, the spectrum reconstructed with the found parameter is always compatible with the initial one, while the value of the parameters themselves sometimes differ significantly from the real ones. The reason why the results of Bayesian Optimization are not always good is probably due to the fact that one of the initial hypothesis is not exactly true: we hypothesize that the AD value at each point has a Gaussian shape. To verify this, we plotted the AD value distribution between multiple voltages obtained from the same parameters and the result is shown in Figure 12. The distribution is more similar to a heavy-tailed distribution than to a Gaussian, which means that we can easily obtain some values that are in the tails, and these unexpected values can confuse the optimization process.

On top of the fact of having imprecise results, this technique remains also very rigid about the shape of the energy spectrum, that must be assumed a priori. Some trials were made allow the spectrum to be more flexible, adding the possibility to have a Gaussian bump on the high energy tail, which increased to five the number of parameters, but the results were not satisfying. Also, each time interval of 1ms requires a different reconstruction, making the process of application to real data still very time-consuming.
Nevertheless, the BO technique for this purpose and for this specific camera system was probably bought to the maximum of its possibilities, and it is limited by the physics behind the Compton scattering and by the



(a) (b)

Figure 12: Distribution of Anderson-Darling statistics for voltage distributions created from the same parameters: (a) real measurements and fit with heavy-tailed distribution (b) symmetrized distribution and fit with heavy-tailed and Gaussian distribution.

saturation of the camera. Application of the same technique to different detectors and proposals different techniques are discussed in Section 5.3.

5.2 Results of real measurements

By looking at Figures 8 and 9 we can notice some difference between the two shots and the third one: in number 61081 both the rate and the energy have a smooth growth and degrowth. The mean energy evolution found in this shot is similar to the one found in [25].In the other two shots instead, there is an initial rate and energy spike around $0.2s$, that could be interpreted as a primary runaway generation and then the two remain high, the rate can increase and there is an abrupt decrease around $0.9s$.
It is interesting to notice how in shot 63424, at the MGI at $0.55s$ corresponds an increase in bremsstrahlung emission rate: this could be compatible with the increased number of collisions between the electrons and the injected gas. In fact, the power emitted by bremsstrahlung emission in a certain frequency range is proportional to the atomic number of the target atom [15]. The difference in atomic number between Ne and Kr could explain why in show 63424 the growth in rate after MGI is much more evident than in number 63420.
On the other hand, it seems that the mean energy of electrons remains constant. The energy and rate evolution in chords 10 and 20 is very similar, compatibly with the similarity in voltage distributions that we see in Fig. 11, while the chord 1 gives more different results. In general, we would expect to find the same mean electron energy in every chord, while the rate depends on the direction of emission. We know in fact that runaways emit bremsstrahlung mainly in a cone centered around their velocity vector, with half angle 1/Y [7] where

$$Y = \frac{E_0}{0.512} \tag{15}$$

The difference in voltage can therefore be explained by the anisotropic emission of radiation, while the difference in mean energy that we find between chord 1 and the others could either be a mistake in the Bayesian Optimization or a result of more complex runaway dynamics that we don't account for.

5.3 Further Developments

The evolution of voltage and rate shown in Section 5.2 are a first effort to develop an alternative technique to interpret the measurements of a detector when the energy is high enough to have Compton scattering events.

In general, we can outline the main limits of this method: the first one is the computational cost of each optimization. For every millisecond that we want to examine it takes 50 minutes of time to find the best parameters. Moreover, the hypothesis of this method that the objective function at each point is a Gaussian process proved not to be exact.
Secondly, due to Compton scattering there is not a univocal correspondence between an energy spectrum and a voltage distribution. This problem is emphasized even more due to the fact that the detector saturates, and we are only able to see a small part of the voltage distribution.
To avoid the saturation effect, the same technique could be applied to detectors that can measure up to energy of the order of MeV, like the scintillator detectors used in [26] and [18]. The better measurements of the high-energy signal could allow for a more flexible shape of the incoming energy spectrum adding some parameters in the research.
Lastly, further developments can be done using more complex machine learning techniques: conditional variational autoencoders have been used for example in Gravitational Waves research to speed up the research in a 9-dimensional parameters space [6]. In this technique, they train two neural networks, an encoder and a decoder, to build posterior distributions starting from some observations of a gravitational wave signal. The posterior can be then marginalised to find the estimation of each parameter. The training is done only once, and the technique is able to speed up the process of parameters searching up to six times faster than the normal methods based on Bayesian inference.

A second technique that could be considered to carry out the parameters research is the Model-Assisted GAN (Generative Adversarial Network). Here, four neural networks work together to learn to imitate the output of a physical simulation, given some input parameters. This can then be used for fast simulations of the outputs. This technique is fairly new, but some initial case studies reveal that it might have a good potential [1].
More powerful techniques like those described could make it easier to increase the number of parameters and therefore the flexibility of the shape of the bremsstrahlung emission.

6 Conclusion

The problem of finding the incoming energy spectrum on a detector becomes ill-posed when the energies are high enough to interact with Compton scattering. Here,we proposed an alternative method of interpreting the signal which takes into account all the ways that the photons can interact with the detector with their probabilities, as well as the possibility of pile-ups and saturation of the detector.
We made use of a detector response simulation at high energies previously built based on Monte Carlo methods and we hypothesised an incoming exponential bremsstrahlung energy spectrum. Of this spectrum we want to find two parameters: the rate N and the mean kinetic electron Energy E_0, given by the slope of the exponential. To find these parameters, we simulate the voltage signal that the camera would measure for a fixed energy spectrum and we compare it to the measured one, trying to find the parameters set which give the most similar voltage signal.
Since the research in the parameters space is computationally costly, we introduced a machine learning technique called Bayesian Optimization based on Gaussian Processes, which is used to find the minimum of a function with the least number of evaluations. We tried to minimize an objective function that measures the similarity between the measured voltage signal and the simulated signal, which is a function of the two parameters. The hypothesis on this function is that at each point it is a Gaussian process, which means that it assumes a random value drawn from a Gaussian distribution with defined mean and variance.
The method had to be trained, which means that we needed to find the best forms for the objective function, the kernel, the acquisition function and the best value for the hyperparameters. The best function to compare two voltage distribution turned out to be the Anderson-Darling statistics, which compares the cumulative distribution functions of the voltage signals giving particular importance to the tails. The error on the parameters was estimated by looking at the distance between the measured and the simulated voltage and setting a maximum value for the difference between them. This estimation on the error on the parameters resulted in an energy spectrum that was always compatible with the initial one, while the single values of the mean energy and the rate had sometimes more than 2σ difference from the initial ones. Such results can be due to the fact that the Anderson-Darling function at one point is not exactly a Gaussian process, because the tails are more probable than in a Gaussian. The physical reason of this is that, given an energy spectrum incoming on the detector, the voltage signal that it can create is very variable because of the different ways that the photons can interact with the detector. The Bayesian Optimisation technique was brought at its best performance, but the results were limited by the assumptions that were not perfectly satisfied.
Nevertheless, we thought it would be interesting to try to apply the method a application of the technique to real measurements to test how the rate and the mean energy evolve in time during a runaway discharge. We applied the technique for three different shots, for which we found different dynamics, in particular between the two shots with MGI and the one without. During the MGI we see an increase in the rates measured by all the chords of the camera, maybe due to the fact that the electrons scatter against the gas atoms and emit more bremsstrahlung. Another interesting feature that we noticed is that, when the camera is in the vertical position and the chords have a different distance from the magnetic axis, the signal they see is different. Since the energies are too high to explain this result with collimation, we hypotised that it could be due to the anisotropic emission of the bremsstrahlung bu the electrons, that emit in a cone centred around their velocity vector. According to some results, it happens that the cameras sometimes see a different mean energy. At the moment we do not have enough data to know if it is due to an error in the research of the minimum or it is a physical result.
This work was a first study that wants to propose a different way to interpret the signal measured by the detectors in the presence of Compton scattering: however, to add more precision and flexibility on the shape of the incoming spectrum, more complex neural networks would be needed, such as conditional variational autoencoder, which already proved very effective in the research of parameters for gravitational waves signals, or model-assisted GANs, that have recently been investigated. These different methods would be even more effective if applied to a detector that saturates at higher energies. In conclusion, many possibilities remain open in order to implement more reliable methods for interpreting the measurements at high energy, and, ultimately, to reconstruct more precisely the runaway energy distribution.

References

[1] Alonso-Monsalve, S., and Whitehead, L. H. Image-based model parameter optimization using model-assisted generative adversarial networks. *IEEE Transactions on Neural Networks and Learning Systems 31*, 12 (Dec 2020), 5645–5650.

[2] authors, T. G. GPyOpt: A bayesian optimization framework in python. http://github.com/SheffieldML/GPyOpt, 2016.

[3] Decker, J., Hirvijoki, E., Embreus, O., Peysson, Y., Stahl, A., Pusztai, I., and Fülöp, T. Numerical characterization of bump formation in the runaway electron tail. *Plasma Physics and Controlled Fusion 58*, 2 (jan 2016), 025016.

[4] Dreicer, H. Electron and ion runaway in a fully ionized gas. i. *Phys. Rev. 115* (Jul 1959), 238–249.

[5] Frazier, P. I. A tutorial on bayesian optimization, 2018.

[6] Gabbard, H., Messenger, C., Heng, I. S., Tonolini, F., and Murray-Smith, R. Bayesian parameter estimation using conditional variational autoencoders for gravitational-wave astronomy, 2020.

[7] Ghanbari, M. R., Ghoranneviss, M., Elahi, A. S., and Mohamadi, S. Measurement of runaway electrons energy by hard x-ray spectroscopy in a small circular cross-section tokamak. *Radiation Effects and Defects in Solids 166*, 10 (2011), 789–794.

[8] Gnesin, S. Electron cyclotron heating and suprathermal electron dynamics in the tcv tokamak. *EPFL* (2011), 201.

[9] Gonzalez, J. Introduction to bayesian optimization with gpyopt. https://nbviewer.jupyter.org/github/SheffieldML/GPyOpt/blob/master/manual/GPyOpt_reference_manual.ipynb. Accessed: 2021-01-30.

[10] Helander, P., Eriksson, L.-G., and Andersson, F. Runaway acceleration during magnetic reconnection in tokamaks. *Plasma Physics and Controlled Fusion 44*, 12B (nov 2002), B247–B262.

[11] Hoppe, M., Embréus, O., Paz-Soldan, C., Moyer, R., and Fülöp, T. Interpretation of runaway electron synchrotron and bremsstrahlung images. *Nuclear Fusion 58* (12 2017).

[12] Hoppe, M., Embréus, O., Tinguely, R., Granetz, R., Stahl, A., and Fülöp, T. Soft: A synthetic synchrotron diagnostic for runaway electrons. *Nuclear Fusion 58* (01 2018), 026032.

[13] J., D., and Y., P. Dke: a fast numerical solver for the 3-d relativistic bounce-averaged electron drift kinetic equation. *Computer Physics Communications 185* (2004).

[14] Knoll, G. F. *Radiation detection and measurement; 4th ed.* Wiley, New York, NY, 2010.

[15] L., H. J. Radiation processes in plasmas. g. bekefi. wiley, new york, 1966. 391 pp., illus. 15.75. *Science 157*, 3796 (1967), 1544–1545.

[16] Loreti, M. *Teoria degli Errori e Fondamenti di Statistica.* Università degli Studi di Padova, 2006.

[17] Lyneis, C., Leitner, D., Todd, D., Virostek, S., Loew, T., Heinen, A., and Tarvainen, O. Measurements of bremsstrahlung production and x-ray cryostat heating in venus. *Review of Scientific Instruments 77*, 3 (2006), 03A342.

[18] Nocente, M., Molin, A. D., Eidietis, N., Giacomelli, L., Gorini, G., Kazakov, Y., Khilkevitch, E., Kiptily, V., Iliasova, M., Lvovskiy, A., Mantsinen, M., Mariani, A., Panontin, E., Papp, G., Pautasso, G., Paz-Soldan, C., Rigamonti, D., Salewski, M., Shevelev, A., Tardocchi, M., , and and. MeV range particle physics studies in tokamak plasmas using gamma-ray spectroscopy. *Plasma Physics and Controlled Fusion 62*, 1 (nov 2019), 014015.

[19] Paz-Soldan, C., Cooper, C., Aleynikov, P., Pace, D., Eidietis, N., Brennan, D., Granetz, R., Hollmann, E., Liu, C., Lvovskiy, A., Moyer, R., and Shiraki, D. Spatiotemporal evolution of runaway electron momentum distributions in tokamaks. *Physical Review Letters 118* (06 2017).

[20] Peysson, Y., Anastassiou, G., Artaud, J.-F., Budai, A., Brizard, A., and Decker, J., Eds. *Modeling runaway electrons dynamics in tokamak plasmas: progresses and challenges* (2019), 27th IAEA Fusion Energy Conference.

[21] Peysson, Y., and Decker, J. Fast electron bremsstrahlung in axisymmetric magnetic configuration. *Physics of Plasmas 15* (September 2008), 14.

[22] Piras, F., Coda, S., Duval, B. P., Labit, B., Marki, J., Medvedev, S. Y., Moret, J.-M., Pitzschke, A., and Sauter, O. Snowflake h mode in a tokamak plasma. *Phys. Rev. Lett. 105* (Oct 2010), 155003.

[23] Porter, F. C. Testing consistency of two histograms, 2008.

[24] Rasmussen, C. E., and Williams, C. K. I. *Gaussian processes for machine learning.* Adaptive computation and machine learning. MIT Press, 2006.

[25] Rasouli, C., Iraji, D., Farahbod, A. H., Akhtari, K., Rasouli, H., Modarresi, H., and Lamehi, M. Runaway electron energy measurement using hard x-ray spectroscopy in "damavand" tokamak. *Review of Scientific Instruments 80*, 1 (2009), 013503.

[26] Shevelev, A., Khilkevitch, E., Kiptily, V., Chugunov, I., Gin, D., Doinikov, D., Naidenov, V., Litvinov, A., and and, I. P. Reconstruction of distribution functions of fast ions and runaway electrons in fusion plasmas using gamma-ray spectrometry with applications to ITER. *Nuclear Fusion 53*, 12 (nov 2013), 123004.

[27] Veitch, J., Raymond, V., Farr, B., Farr, W., Graff, P., Vitale, S., Aylott, B., Blackburn, K., Christensen, N., Coughlin, M., and et al. Parameter estimation for compact binaries with ground-based gravitational-wave observations using the lalinference software library. *Physical Review D 91*, 4 (Feb 2015).