



CZECH TECHNICAL UNIVERSITY IN PRAGUE

FACULTY OF NUCLEAR SCIENCES AND  
PHYSICAL ENGINEERING

DEPARTMENT OF PHYSICS

Optimization of the data processing from the  
Thomson scattering diagnostics on the COMPASS  
tokamak using Bayesian methods

Optimalizace zpracování dat z diagnostiky  
Thomsonova rozptylu na tokamaku COMPASS  
s využitím Bayesovských metod

RESEARCH TASK

Author:	Bc. Jan Hečko
Supervisor:	Ing. Miroslav Šos
Specialization:	Physics and Technology of Thermonuclear Fusion
Academic year:	2019/2020





**Katedra:** fyziky

**Akademický rok:**

2019/2020

## VÝZKUMNÝ ÚKOL

**Student:** Bc. Jan Hečko

**Studijní program:** Aplikace přírodních věd

**Obor:** Fyzika a technika termojaderné fúze

**Vedoucí úkolu:** Ing. Miroslav Šos, Ústav fyziky plazmatu AV ČR, v.v.i.

### **Název úkolu (česky/anglicky):**

Optimalizace zpracování dat z diagnostiky Thomsonova rozptylu na tokamaku COMPASS s využitím Bayesovských metod

Optimization of the data processing from the Thomson scattering diagnostics on the COMPASS tokamak using Bayesian methods

### **Pokyny pro vypracování:**

Pomocí diagnostiky Thomsonova rozptylu (TS – z ang. *Thomson scattering*) jsou na řadě světově významných zařízení zabývajících se fyzikou plazmatu měřeny základní veličiny popisující chování a stav plazmatu, a sice elektronová teplota a hustota elektronů přítomných v plazmatu. Jedinečnou schopností této diagnostiky je měření prostorového profilu zmíněných veličin, aniž by došlo k narušení stavu zkoumaného plazmatu.

Bayesovské metody představují silný nástroj pro podrobnější analýzu diagnostických metod nejenom v oblasti fyziky plazmatu [3, 4, 5]. Jednou z nezpochybnitelných výhod Bayesovského přístupu je korektní práce s chybami měření a predikce správných, relevantních chyb výsledných fyzikálních veličin.

Zadání práce je shrnuto v následujících několika bodech:

1. Sestavte dopředný model diagnostiky Thomsonova rozptylu.
2. Využijte existující modely výbojů v tokamaku COMPASS za účelem generování syntetických dat profilů elektronové teploty a hustoty.
3. Pokuste se využít dopředný model pro zpracování syntetických dat z předchozího úkolu pomocí Bayesovského přístupu
4. Pokuste se do modelu zahrnout další diagnostiky elektronové hustoty (interferometrie)

**Literatura:**

- [1] P. Bilkova, P. Bohm, M. Aftanas, M. Sos, A. Havranek, D. Sestak, V. Weinzettl, M. Hron, R. Panek, and the COMPASS team. High resolution Thomson scattering on the COMPASS tokamak—Extending edge plasma view and increasing repetition rate. *Journal of Instrumentation*, 13(01):C01024, 2018.
- [2] S. L. Prunty. A primer on the theory of Thomson scattering for high-temperature fusion plasmas. *Physica Scripta*, 89(12):128001, 2014.
- [3] A. B. Downey. *Think Bayes*. Green Tea Press, Needham, USA, 2012.
- [4] G. T. von Nessi and M. J. Hole. Using Bayesian analysis and Gaussian processes to infer electron temperature and density profiles on the Mega-Ampere Spherical Tokamak experiment. *Review of scientific instruments*, 84(6):063505, 2013.
- [5] R. Fischer, A. Dinklage and E. Pasch. Bayesian modelling of fusion diagnostics. *IOP Publishing*, 45(7):1095-1111, 2003.

**Datum zadání:** 25.10.2019

.....

**Datum odevzdání:** 26.06.2020

**vedoucí katedry**

*Název práce:*

**Optimalizace zpracování dat z diagnostiky Thomsonova rozptylu na tokamaku COMPASS s využitím Bayesovských metod**

*Autor:* Bc. Jan Hečko

*Obor:* Fyzika a technika termojaderné fúze

*Druh práce:* Výzkumný úkol

*Vedoucí práce:* Ing. Miroslav Šos

Ústav fyziky plazmatu AV ČR, v.v.i.

*Abstrakt:* Diagnostika plazmatu pomocí Thomsonova rozptylu se používá na mnoha světově známých fúzních zařízeních. Tato diagnostika je unikátní ve svém oboru, neboť je s ní možné měřit prostorové profily elektronové teploty a hustoty plazmatu bez narušení jeho aktuálního stavu. Jednou z hlavních součástí každého diagnostického systému je software pro zpracování a analýzu dat. Diagnostika plazmatu pomocí Thomsonova rozptylu na tokamaku COMPASS využívá k výpočtu profilů teploty a hustoty software, který je založený na fitovacích algoritmech. Součástí této výzkumné práce je představení nového způsobu výpočtu profilů, který vychází z metod Bayesovské statistiky. Jsou představeny dva modely, které jsou následně otestovány na umělých datech a srovnány s původním zpracováním dat na tokamaku COMPASS.

*Klíčová slova:* Thomsonův rozptyl, Bayesovská statistika, tokamak

*Title:*

**Optimization of the data processing from the Thomson scattering diagnostics on the COMPASS tokamak using Bayesian methods**

*Author:* Bc. Jan Hečko

*Abstract:* The Thomson scattering (TS) is an important plasma diagnostics that is used on many fusion devices worldwide. It measures spatial profiles of electron temperature and density of plasma without influencing its state and properties, which makes it unique in its field. The main component of each diagnostic system is a data processing software. To calculate the profiles, the TS diagnostic system on the COMPASS tokamak uses a software based on curve fitting algorithms. In this thesis a different approach which utilizes the methods of Bayesian inference is introduced. Two inference models are implemented, tested on synthetic data and compared with the original COMPASS software.

*Keywords:* Thomson scattering, Bayesian inference, tokamak

# Contents

<b>Introduction</b>	<b>1</b>
<b>1 Thomson scattering on the COMPASS tokamak</b>	<b>2</b>
1.1 Fundamentals of TS . . . . .	2
1.2 Layout of TS . . . . .	3
1.3 TS data processing . . . . .	4
<b>2 Bayesian inference</b>	<b>6</b>
2.1 Motivation for Bayesian inference in experimental physics . . . . .	6
2.2 Bayes' theorem and terminology . . . . .	7
2.3 Calculating the posterior distribution . . . . .	8
2.4 Introduction to the PyMC3 library for Python . . . . .	8
2.4.1 Constructing the inference model . . . . .	9
2.4.2 Sampling the posterior . . . . .	9
2.4.3 Using variational methods to approximate the posterior . . . . .	10
2.5 Gaussian processes . . . . .	10
2.5.1 Kernels . . . . .	10
2.5.2 Gaussian processes with positive domain . . . . .	12
<b>3 Synthetic TS data</b>	<b>15</b>
3.1 Source of synthetic profiles . . . . .	15
3.1.1 Model for TS profile fitting . . . . .	15
3.2 Simulating profile fluctuations . . . . .	18
3.3 Simulating measurement noise . . . . .	19
3.4 Overview of generated datasets . . . . .	19
3.5 Quantifying the performance of inference method using synthetic data . . .	19
3.5.1 Indicator 1: Precision of the result . . . . .	20
3.5.2 Indicator 2: Precision of error estimate . . . . .	21
<b>4 Results: Point inference</b>	<b>22</b>
4.1 Model description . . . . .	22
4.1.1 Prior distribution . . . . .	22
4.1.2 Likelihood . . . . .	23
4.1.3 Inference algorithms . . . . .	23
4.2 Inference on simulated data . . . . .	23
4.3 Inference on data from database . . . . .	24

<b>5</b>	<b>Results: Profile inference with Gaussian processes</b>	<b>26</b>
5.1	Model description . . . . .	26
5.1.1	Prior distribution . . . . .	26
5.1.2	Likelihood . . . . .	27
5.1.3	Inference algorithms . . . . .	27
5.2	Inference on simulated data . . . . .	27
5.3	Inference on data from database . . . . .	28
<b>6</b>	<b>Discussion</b>	<b>32</b>
6.1	The development of the models . . . . .	32
6.2	Performance of the inference models based on the synthetic datasets . . . .	32
6.3	The unstable behaviour in edge areas with low signal-to-noise ratio . . . .	34
6.4	Processing archived data from COMPASS database . . . . .	34
6.5	Further development . . . . .	34
<b>A</b>	<b>List of probability distributions</b>	<b>38</b>
A.1	Normal and half-normal distribution . . . . .	38
A.2	Cauchy and half-Cauchy distribution . . . . .	39
A.3	Log-normal distribution . . . . .	39

# Introduction

Thermonuclear fusion is an ultimate source of energy, which powers the stars. While the Sun fuses large amounts of hydrogen isotopes every day, replicating this physical phenomenon in smaller scale on the surface of Earth is not a trivial task.

Current knowledge of controlled thermonuclear fusion indicates that the best approach is to generate and confine high temperature plasma. Usually plasma composed of hydrogen isotopes is used. Given enough time, the atoms from the high-energetic branch of the Maxwell spectrum initiate a powerful enough collision that results in fusion of the colliding particles and production of excess energy. There are several approaches to achieving the confinement of high temperature plasma. A very promising seems to be the use of strong magnetic fields which restrict the movement of the charged plasma particles and essentially capture them. This branch of fusion research is referred to as the magnetic confinement fusion. Most modern devices which utilize magnetic confinement are either tokamaks or stellarators. Both consist of a vacuum vessel, which contains the plasma, and a large number of strong electromagnets. Tokamaks induce current in the plasma in order to achieve an optimal configuration of the magnetic field. In contrast, the stellarators do not induce current and use unconventionally shaped electromagnets instead, which need to be manufactured with very high precision.

The advancement of fusion research relies heavily on effective methods of plasma diagnostics. A disadvantage of numerous diagnostics is that they measure only local properties (magnetic and electric probes) or volume/line integrated values (interferometry, spectroscopy, neutrons). The exception to this is the Thomson scattering (TS) diagnostics, which directly measures full spatial profiles of plasma temperature and density. Plasma profiles are very important in respect to the research of confinement modes.

This research task focuses on the data processing software of TS diagnostic system on the COMPASS tokamak ( $R = 0.56$  m,  $a = 0.18$  m,  $B_T = 0.8$ – $2.1$  T,  $\kappa = 1.6$ , plasma current up to 400 kA, NBI (neutral beam injection) heating  $2 \times 400$  kW). Current software is based on curve fitting algorithms and calculates the temperature and density spatial profiles one point at a time. However, in reality the points are not independent. According to the general theoretical understanding of plasma in tokamak it is assumed that the profiles are formed by continuous functions. Therefore all points in the plasma profile are spatially correlated.

An alternative approach to analysing the TS data has been proposed. It employs the statistical methods of Bayesian inference and Gaussian processes. The main objective is to develop a more robust data processing routine that is able to incorporate the spatial correlation of points. The result of profile inference would then be a continuous function instead of a set of independent points.



# Chapter 1

## Thomson scattering on the COMPASS tokamak

Thomson scattering (TS) is an important method of plasma diagnostics commonly used on many present-day fusion experiments that utilize the magnetic confinement approach to controlled fusion (tokamaks, stellarators)[1, 2]. It is installed on the COMPASS tokamak[3] and it is included in design plans for the new device COMPASS-U[4]. Typical application of the TS diagnostics is to measure spatial profiles of electron temperature  $T_e$  and density  $n_e$  with resolution in the order of 10 or even 1 mm. Temporal resolution is usually low (hundreds of Hz).

### 1.1 Fundamentals of TS

Theory of Thomson scattering diagnostics is based on a scattering phenomenon of the same name. Monochromatic light with well-defined spectral properties, typically a laser beam, is scattered in plasma by free electrons. Resulting scattered spectrum is detected and analysed. Because the electrons are moving, the scattered spectrum is deformed by Doppler shifts as seen in Fig. 1.1. Assuming that the electrons are in thermal equilibrium, the spectral intensity of scattered beam can be expressed as a function of  $T_e$  and  $n_e$ . Thus, original values of  $T_e$  and  $n_e$  can be inferred from scattered spectrum. There are several implementations of the TS diagnostics that use different type of lasers and/or spectroscopy systems.

Main pitfall of TS is the low signal-to-noise ratio. The cross section of the TS phenomenon  $\sigma_{TS}$  is a fundamental limiting factor. Combined with typical plasma density values in tokamak it results in fairly weak signal. Subsequently, high-power lasers and sensitive spectrum analysers need to be utilized.

The theory of TS is addressed in-depth in [6, 7]. More practical details aimed at the implementation of a TS diagnostic system can be found in [5, 8].

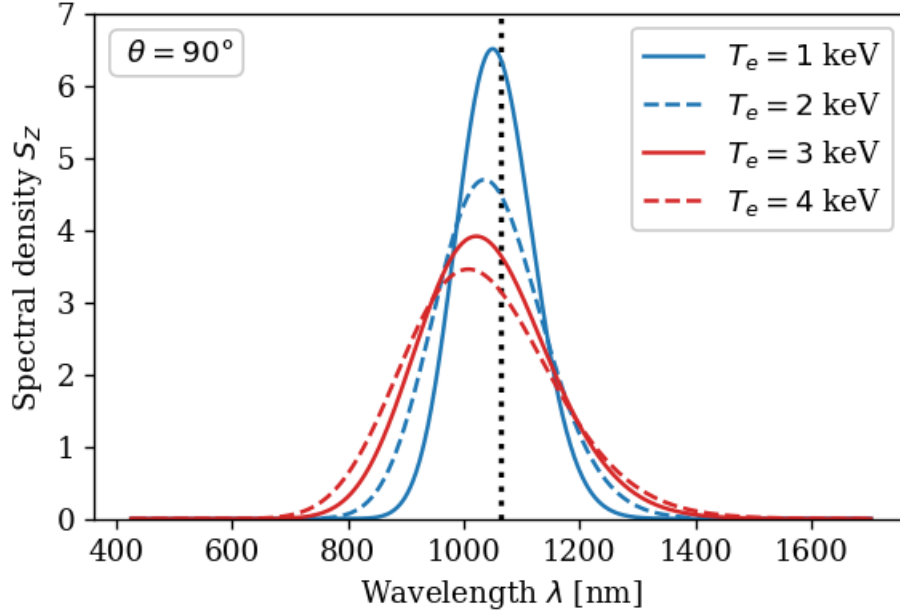


Fig. 1.1: An example of TS spectrum for a range of electron temperatures  $T_e$ .  $\theta$  is the angle between the incident beam and the observer. Vertical dotted line at  $\lambda = 1064$  nm marks the laser line of Nd:YAG laser.[5]

## 1.2 Layout of TS

The layout of TS diagnostics on the COMPASS tokamak is visualized in Fig. 1.2. Incidental beams are produced by 4 Nd:YAG lasers and the spectrum analysis is conducted by a set of 29 polychromators (5-channel). Scattered light is collected by two separate lenses from two adjacent areas spanning along the z-axis above the equatorial plane. There is the *core TS* and the *edge TS* area with 24 and 30 spatial points respectively. Note that a duplexing technique is employed in order to analyse signal from 54 points while using only 29 polychromators.

The edge TS has higher spatial resolution, 3.7 mm on average, compared to core TS, 9.9 mm on average. The actual distance between neighbouring points slightly varies depending on the viewing angle of the collecting optics. The total observed spatial distance is 336 mm.

The time resolution is given by the repetition rate of the lasers. Combined together the 4 lasers offer a continual repetition rate of 120 Hz. Alternatively, the lasers can be used in *burst mode* to perform fewer measurements with higher time resolution.

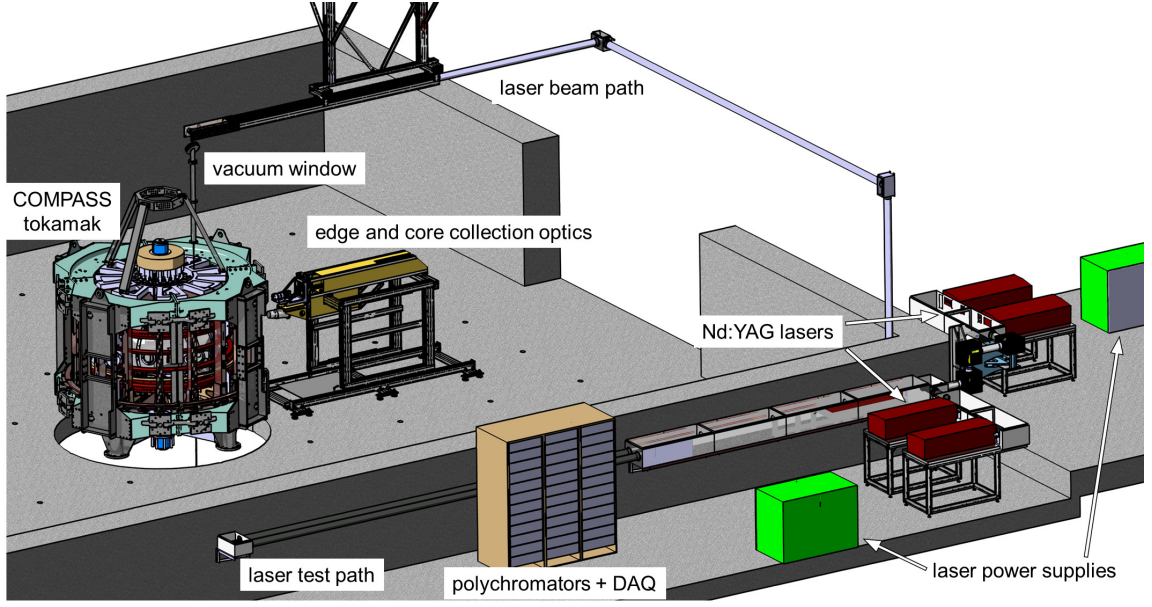


Fig. 1.2: Layout of TS diagnostic system on the COMPASS tokamak.[9]

### 1.3 TS data processing

On the COMPASS tokamak, polychromators are used for spectral analysis of scattered light. Each polychromator has 5 channels with APDs<sup>1</sup> that detect light in 5 adjacent spectral bands. Signal from each channel is integrated in time to retrieve the total detected intensity during a single laser pulse. With additional knowledge about the diagnostic system, which is acquired through a set of calibration processes, and theory of TS spectrum it is possible to infer the  $T_e$  and  $n_e$  using a following relation

$$V_i = \underbrace{(G\Delta\Omega T(\lambda_L)LQE)}_{\text{system constants}} \underbrace{\left(\frac{1}{hc/\lambda_L} \frac{d\sigma_{TS}}{d\Omega}\right)}_{\text{known constants}} n_e \underbrace{E_{laser}}_{\text{const.}} \int \frac{\phi_i(\lambda)}{\phi_i(\lambda_L)} \frac{S(\lambda; T_e, \theta, \lambda_L)}{\lambda_L} d\lambda. \quad (1.1)$$

Here  $V_i$  denotes the integrated intensity in one of the 5 channels that correspond to one observed spatial point. Function  $S$  is the spectral density of TS. Although under specific conditions an analytical formula of  $S$  exists, often it is reasonable to use an approximation instead. The approximation that was used in this research task is presented in [10]. Integration parameter  $\lambda$  is the wavelength and constant  $\lambda_L$  is the known wavelength of the laser source.

Contrary to the first impression, the relation (1.1) is rather simple because most of the symbols that are present are either physical constants or system constants acquired from calibrations. A more intuitive version of (1.1) can be formulated

$$V_i = \alpha n_e \int \beta_i(\lambda) S(\lambda; T_e, \theta, \lambda_L) d\lambda, \quad (1.2)$$

where  $\alpha$  represents all general constants and spatial-point-specific calibrations and  $\beta_i$  represents channel-specific calibrations. It is now obvious that  $V_i$  is function of  $T_e$  and  $n_e$  and as such it can be used to infer the electron temperature and density.

<sup>1</sup>APD = avalanche photo diode, optical detector with high sensitivity.

The calibrations of TS diagnostic system on COMPASS tokamak referenced in the previous text include the absolute calibration, the spatial calibration and the spectral calibration. During the absolute calibration, all constants of the system that are not channel-specific are measured. The spectral calibration is tied to the spectral properties of all polychromator channels. The spatial calibration is tied to the spatial properties, e.g. the exact position of spatial points in the chamber and the angles of observation. In addition to that, the laser pulse energy is measured.

More details about calibrations, the relation (1.1) and approximation of spectral density  $S$  of TS spectrum that is used can be found in [11].

## Chapter 2

# Bayesian inference

### 2.1 Motivation for Bayesian inference in experimental physics

Physical measurement is an inherently stochastic process and it requires the use of probability theory and statistics. The measured quantity can be modelled as a random variable with unknown probability distribution. Data obtained through the measurement represents random samples of this probability distribution. The processes of estimation of the results, that is the real value of the measured quantity and its error, is referred to as the statistical inference.

Conventionally<sup>1</sup> it is assumed that the measured quantity has normal distribution of probability and repeated measurements increase the precision. The argument why such assumption is reasonable is given by the central limit theorem. Then, arithmetic mean and standard deviation can be used to estimate the measured value and error. The downside of this approach is that the probability distribution, which is unknown, could be for example deformed (asymmetric, bimodal) under the influence of a non-ideal measurement method. In that case, systematically incorrect results are produced. Often it is a complicated task to both indicate the deviation from normal distribution and to implement systematic correction of results.

Bayesian inference offers an alternative approach. It is the process of using Bayesian statistics to analyse data and deduce the properties (shape) of an underlying probability distribution directly. In other words, the result of the inference is the probability distribution. Probability distribution contains the maximum information about a given stochastic process and thus, in theory, it is possible to estimate the measured value and error more rigorously. Apart from that, Bayesian inference also offers a straightforward way to increase precision of results by incorporating additional information about the measurement method and underlying physics of the measured object. Examples of application on plasma diagnostics are presented in [1] and [2].

---

<sup>1</sup>In relation to the courses of statistics and experimental physics taught at FNSPE CTU in Prague. Keywords: frequentist approach to statistics, the central limit theorem, arithmetic average and standard deviation.

## 2.2 Bayes' theorem and terminology

The fundamental structure and philosophy of Bayesian inference is derived from the Bayes' theorem, which can be stated as

$$P(A|D) = \frac{P(D|A) \cdot P(A)}{P(D)}, \quad (2.1)$$

where  $A$  represents the observed quantity and  $D$  represents data from experiment. Notation  $P(\dots)$  and  $P(\dots|\dots)$  refers to probability and conditional probability respectively.

The aforementioned structure of Bayesian inference is more obvious when the equation (2.1) is dissected into separate terms. In this context only the first three terms, discussed below, are important. The fourth term  $P(D)$ , which resides in denominator, is the summary (marginal) probability distribution of all possible data values. It is a normalization term and in practice it is automatically determined during the computation.[12, 13]

- Term  $P(A)$  is the *prior distribution of probability* or simply *prior*. It reflects any information regarding the observed quantity  $A$  that is known *prior* to the experiment. It should not be connected in any way to the data  $D$  or the measurement process (*likelihood*). A simple example of a prior would be an order estimate for  $A$ . A different model may incorporate results of a complex physical simulation instead, while risking the fact that the simulation may not reflect the reality in whole. The amount and validity of information that is provided affects both the final result and its uncertainty. Choosing the correct prior for a given situation is a vital task.
- Term  $P(D|A)$  is the *likelihood*. Note that  $P(D|A)$  is a conditional probability: the probability of measuring data  $D$  given a specific value of  $A$ . In other words, it refers to a relation between  $D$  and  $A$  in the form of

$$D = f(A), \quad \text{or rather} \quad D \sim \text{distribution}(f(A), \dots), \quad (2.2)$$

where *distribution* is an unspecified probability distribution that represents the stochastic character of the measurement process. The function  $f$  can be referred to as the forward model of the measurement process. Forward model is a function that accepts values of observed quantity  $A$  and yields corresponding data values  $D$ . For example, in case of linear regression  $D = y = A \cdot x$  the forward model is  $f(A) = A \cdot x$ , where the horizontal coordinates  $x$  are fixed known values. Contrary to prior distribution, forward model is rarely subject to choice based on scientist's preferences. Often for a given problem there is only a limited number of options to choose from and the main difference is usually their precision.

- Term  $P(A|D)$  is the *posterior distribution of probability* or simply *posterior*. Posterior is a conditional probability, similar to likelihood but reversed: the probability distribution of observed quantity  $A$  given that data  $D$  has been measured. It is the result of the inference. The word *posterior* refers to the fact, that at this stage of the inference process the experiment has been conducted and measured data  $D$  has been considered.

Considering the new terminology, an intuitive interpretation of the Bayesian approach can be formulated: *An experiment is conducted with the motivation to update a prior belief with new knowledge and form a more precise posterior belief.* Here the new knowledge includes both the measured data and the forward model. The emphasis is on the fact that

the inference is a process of updating beliefs or knowledge. An argument is often made that this is equivalent to the natural decision making process of human mind<sup>2</sup> and thus more intuitive than other approaches to statistics.

In practice, the Bayes' theorem (2.1) is hidden away as a part of the numerical algorithms that sample or approximate the posterior distribution. When building an inference model, the user instead comes into contact with the concepts of prior, likelihood and posterior as it is described in the following section.

## 2.3 Calculating the posterior distribution

Aside from special cases, it is not possible to calculate full posterior distribution in finite time. Instead, a class of non-deterministic numerical algorithms called Markov chain Monte Carlo (MCMC) is used. These algorithms are able to draw samples from an unknown posterior distribution based on the known prior distribution, likelihood and data. A draw of  $N$  samples is considered the approximation of a final result that converges for  $N \rightarrow \infty$ . Samples can be used to calculate posterior properties such as median, mean and credibility interval or to visualize and plot the distribution. Popular MCMC sampling algorithms include:

- Metropolis-Hastings, published 1970 [14]
- Hamiltonian or Hybrid Mote Carlo (HMC), published 1987 [15]
- No U-Turn Sampler (NUTS), published 2011 [16], extension of HMC

The NUTS sampler was used for posterior sampling in this research task.

An alternative to posterior sampling algorithms is offered by variational inference (VI) algorithms. The purpose of VI algorithms is to infer an approximation of the posterior distribution and then to draw samples of posterior from this approximation with high efficiency. VI algorithms usually sacrifice the precision of results exchange for smaller computational demand. Some algorithms may also sacrifice the guarantee of valid results when applied to a completely general case. Therefore it is recommended to probe the posterior using the more robust MCMC algorithms in order to verify applicability of a specific VI algorithm.

## 2.4 Introduction to the PyMC3 library for Python

The MCMC sampling algorithms and related advances in this field both popularized the Bayesian approach to statistics and inspired a number of publicly available computational platforms for Bayesian inference. Notable instances of such platforms include the Stan[17] (specialized programming language), Pyro[18] (library for Python language) and PyMC3[19] (library for Python language). The PyMC3 library was decided to be the best fit for the purposes of this research task. One of the main reasons being its both detailed and accessible documentation that also includes a variety of introductory materials and examples.

---

<sup>2</sup>Consider a Constitution class starship on its way to the Federation-Romulan border to check-up on a monitoring outpost that fell silent. Captain Kirk keeps an open mind and thinks that a malfunction is more probable than a Romulan attack. Next, a contact is lost with two more outposts and a third one broadcasts emergency signal. Captain Kirk updates his believes and raises red alert.

### 2.4.1 Constructing the inference model

The general concept of PyMC3 is to offer a friendly Python interface that allows the user to specify the inference model. The computation itself is implemented using optimized background libraries and on-the-fly compiled C code and therefore avoids any computational inefficiency related to Python. Each model is specified by a set of parameters (the unknowns, e.g.  $T_e$  and  $n_e$ ) and an observed variable (the measured quantity, e.g. channel intensity  $V_i$ ), while the forward model connects them together. A probability distribution needs to be assigned both to parameters and to the observed variable because they are all random variables. To recall the Bayesian terminology, the distribution of parameters and observed variable corresponds to the *prior* distribution and the *likelihood* respectively. Finally, the observed variable requires the actual measured data. An example of a linear regression model defined in PyMC3:

```
import pymc3 as pm
import numpy as np

# Simulating data from an experiment:
x = np.array([1, 2, 3, 4, 5])
data = np.array([1.11, 1.23, 1.30, 1.39, 1.55])
data_err = np.array([0.01, 0.02, 0.01, 0.02, 0.03])

# The forward model for linear regression
def f(A, B):
    return A*x + B

with pm.Model() as model:
    # Model parameters (prior), slope and intersection
    var_A = pm.Uniform('A', lower=-5, upper=5)
    var_B = pm.Uniform('B', lower=-10, upper=10)

    # Observed variable (likelihood)
    y = pm.Normal('y', mu=f(var_A, var_B), sigma=data_err, observed=data)
```

### 2.4.2 Sampling the posterior

When a model has been specified, the posterior distribution can be sampled and the results acquired. In PyMC3 framework it means the following:

```
with model:
    trace = pm.sample(draws=2000, tune=1000, chains=8)
# As soon as the sampling process is finished, the variable 'trace'
# will contain samples from posterior distribution of 'A' and 'B',
# where 'A' is the slope and 'B' is the intersection
```

The function `sample()` accepts various arguments that control the sampling process. Listing the three most important arguments:

- **draws** — Number of samples to draw from the posterior distribution per chain. A reasonable amount is 2000 for a simple model and 10 000 for a complex model.
- **tune** — Number of tune samples that are used to tune the algorithm. A reasonable amount is for example a half of **draws**.
- **chains** — Number of computational chains. Chains can be run in parallel, utilizing the efficiency of a multi-core CPU. It is recommended to compute at least 2 chains so that the results can be compared.



It is good practice to always check and diagnose results of posterior sampling. One common issue that can occur during posterior sampling is a divergent sample. If a divergent sample is present, the results should be deemed unreliable. Another issue is the autocorrelation of samples. When high autocorrelation between samples is detected, it signals low information density and therefore low effective sample size of the result. Both of these checks are calculated and logged by PyMC3 automatically, there are, however, numerous other indicators and tests for model and results diagnostics.[20]

Sampling issues can be countered either by increasing the `tune` parameter or by revising the model: its structure, the choice of parameters and their prior distributions.

### 2.4.3 Using variational methods to approximate the posterior

PyMC3 also implements the variational methods mentioned in section 2.3. One of the methods available in the library is the Automatic differentiation variational inference (ADVI) that uses an approximation based on single-mode normal distribution. Another is the Stein variational gradient descent (SVGD). While the methods (mainly ADVI) were successfully utilized in order to acquire fast results, in general they go beyond the scope of this research task.

## 2.5 Gaussian processes

Typically, the domain of Bayesian inference is the number-space. The use of Gaussian processes (GP) extends this domain to the function-space,  $f : \mathbb{R} \rightarrow \mathbb{R}$ . To clarify, instead of inferring electron temperature  $T_e$  at discrete spatial points, it is possible to infer continuous spatial profile  $T_e = T_e(z)$  from the same data. The advantage of this approach is that additional information, more specifically the correlation between points in the same spatial profile, is incorporated and that it is possible to predict the function values at new points. The PyMC3 library offers a convenient implementation of GP.[19]

Gaussian processes are used similarly to any other *prior distribution*. In theory they are represented by a Multivariate Gaussian distribution of infinite size that is collapsed (marginalized) into a finite size equal to the number of data points. A Multivariate Gaussian, and therefore GP as well, is defined by a mean vector and a covariance matrix. The combination of mean and covariance determines the properties of functions that are sampled from the GP prior. While the mean can be used to specify a general trend or to provide correct scale, the main properties are influenced by the covariance.[13]

### 2.5.1 Kernels

In the context of GP, covariance matrix is usually constructed using a kernel, sometimes referred to as the covariance function. Kernel can be naively described as a symmetric weighing function that accepts two inputs and returns a positive value

$$k : \mathbb{X} \times \mathbb{X} \rightarrow \mathbb{R}^+, \quad \text{where } \mathbb{X} \text{ is an unspecified input space} \quad (2.3)$$

The output value can be interpreted as a measure of similarity between the two inputs. Notice that if a kernel measures the similarity of inputs inversely proportional to their distance, it actually describes a property analogous to the continuity of a function. Throughout the years, many different kernels have been invented describing different classes of functions. One of the main features that a kernel can determine is the order of differentiability — the smoothness. More detailed theoretical introduction to the concept of kernels

as covariance functions of GP can be found in [21, Chapter 4]. A list of kernels selected for the purpose of this research task is presented:

- Exponentiated quadratic (Gaussian), Fig. 2.1,

$$k(x, x') = \exp \left[ -\frac{(x - x')^2}{2\ell^2} \right] \quad (2.4)$$

where the parameter  $\ell$  defines the length-scale.

The exponentiated quadratic kernel (EQ) describes a class of infinitely differentiable functions. According to [21] it has been argued that such strong smoothness assumptions are unrealistic for modelling many physical processes and the Matérn class is recommended instead.

- Matérn  $\nu = \frac{5}{2}$ , Fig. 2.2,

$$k(x, x') = \left( 1 + \frac{\sqrt{5(x - x')^2}}{\ell} + \frac{5(x - x')^2}{3\ell^2} \right) \exp \left[ -\frac{\sqrt{5(x - x')^2}}{\ell} \right] \quad (2.5)$$

and  $\nu = \frac{3}{2}$ , Fig. 2.3,

$$k(x, x') = \left( 1 + \frac{\sqrt{3(x - x')^2}}{\ell} \right) \exp \left[ -\frac{\sqrt{3(x - x')^2}}{\ell} \right], \quad (2.6)$$

where in both equations parameter  $\ell$  defines the length-scale.

The Matérn kernel uses an additional parameter  $\nu$  to describe a class of functions which are  $k$ -times differentiable if and only if  $\nu > k$ . For  $\nu \rightarrow \infty$  it becomes the EQ kernel. The general defining formula of this kernel, which can be found in [21], is complex, however, for special cases of  $\nu$  it can be significantly simplified. Two such cases,  $\nu = \frac{5}{2}$  and  $\nu = \frac{3}{2}$ , are presented above. The examples in Fig. 2.2 and Fig. 2.3 visibly reflect more diverse features than the infinitely smooth EQ kernel in Fig. 2.1.

- Exponential, Fig. 2.4,

$$k(x, x') = \exp \left[ -\frac{||x - x'||}{\ell} \right] \quad (2.7)$$

where the parameter  $\ell$  defines the length-scale.

The exponential kernel is in fact another special case of the Matérn kernel, where  $\nu = \frac{1}{2}$ . According to the statement about the relation of  $\nu$  and the differentiability, it describes a class of continuous but not differentiable functions. Such assumptions about smoothness appear to be very conservative and possibly unrealistic in the context of many physical processes. It can be seen from the example in Fig. 2.4 that the functions are comprised of very sharp peaks.

Note that kernels often have one or more parameters that need to be specified, e.g. the lengthscale  $\ell$ . Instead of assigning a fixed value, these parameters are usually regarded as additional unknowns of the model and their own prior distribution is specified. Because they parametrize the prior distribution of the original parameters, they are sometimes referred to as the *hyperparameters*.

In order to describe functions with more complex properties, it is possible to combine several kernels (covariance matrices) with the use of following operations:

- $k(x, x') = k_1(x, x') + k_2(x, x')$  ... element-wise addition
- $k(x, x') = k_1(x, x') \cdot k_2(x, x')$  ... element-wise multiplication
- $k(x, x') = \eta^2 \cdot k_1(x, x')$  ... multiplication by a positive scalar

The scalar multiplication should be always utilized to introduce a scaling hyperparameter  $\eta$  as an additional degree of freedom. The kernels are normalized by default.

### 2.5.2 Gaussian processes with positive domain

As was mentioned before, GP are represented by a Multivariate Gaussian distribution which is defined over (a vector of) all real numbers. In some cases, strictly positive values could be required. To obtain a GP prior defined over positive numbers it is possible to apply exponential transformation and use the GP prior in logarithmic space:

$$T_e \sim \exp[\mathcal{GP}(\vec{\mu}, \Sigma)] \quad \text{instead of} \quad T_e \sim \mathcal{GP}(\dots), \quad (2.8)$$

where  $\vec{\mu}$  is the mean vector and  $\Sigma$  is the covariance matrix.

The exponential transformation causes several notable side effects. Functions characterized by the kernels listed above resemble fluctuations around mean that are symmetrically distributed. The transformation deforms them so that their amplitude above mean is greater than below mean (see Fig. 2.1). It is caused by the fact that the absolute fluctuations in logarithmic space are transformed into relative fluctuations in respect to the mean value. Therefore, if a non-constant mean were to be used, additional deformation would appear. The transformation also affects the hyperparameters, namely the scaling factor  $\eta$ . It is convenient to redefine it so that it scales the kernel in following form:

$$[\log(1 + \eta)]^2 \cdot k(x, x'). \quad (2.9)$$

In this form the  $\eta \in (0, +\infty)$  correctly parametrizes the relative scale.

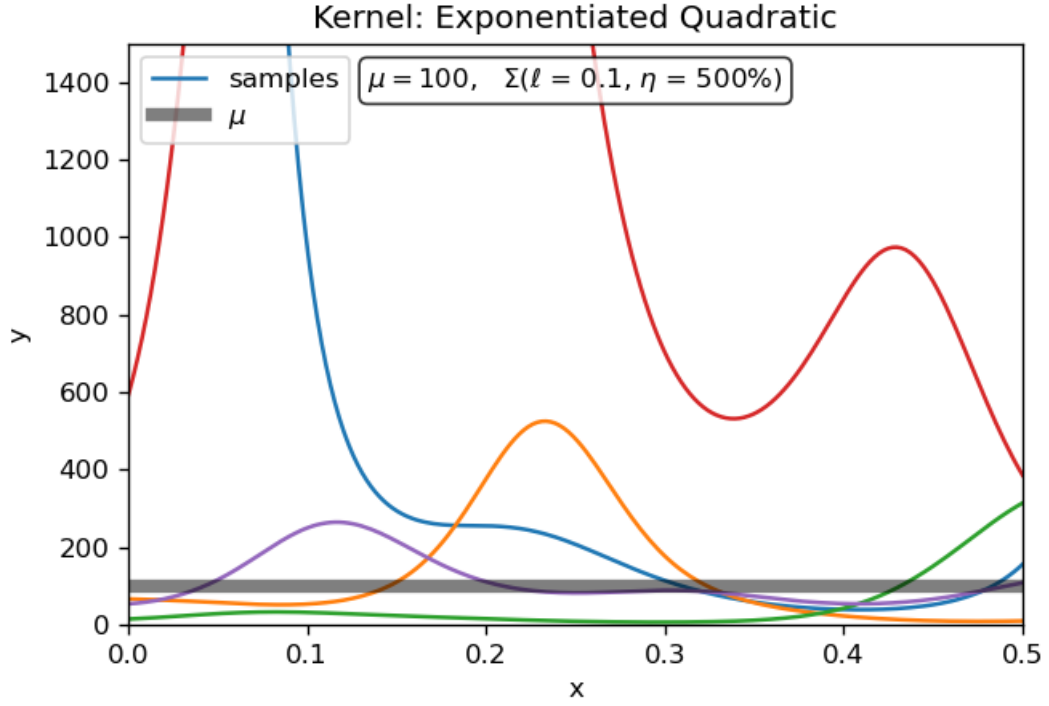


Fig. 2.1: Functions sampled from GP prior based on the exponentiated quadratic kernel (2.4). Transformed from log-space, see section 2.5.2.

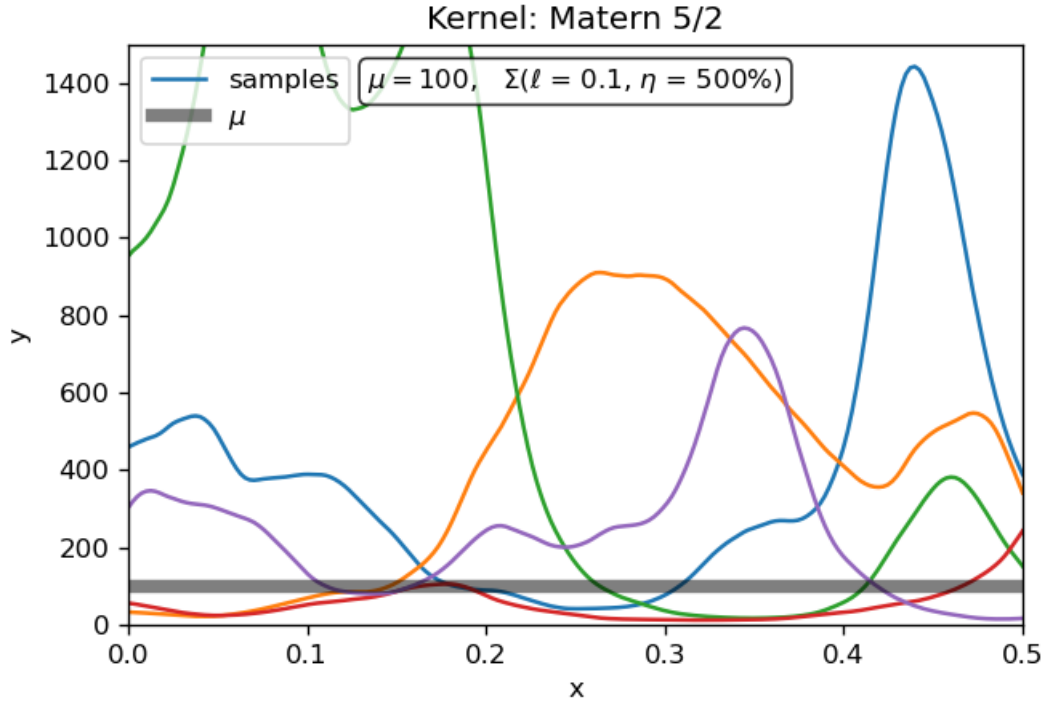


Fig. 2.2: Functions sampled from GP prior based on the Matérn kernel (2.5) with  $\nu = \frac{5}{2}$ . Transformed from log-space, see section 2.5.2.

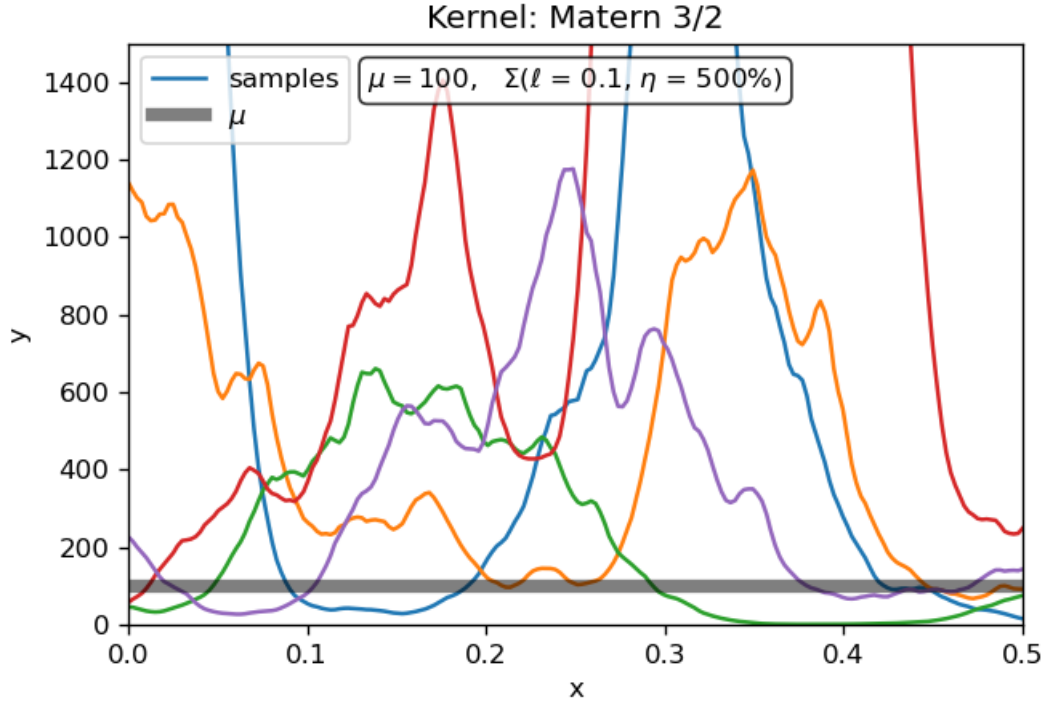


Fig. 2.3: Functions sampled from GP prior based on the Matérn kernel (2.6) with  $\nu = \frac{3}{2}$ . Transformed from log-space, see section 2.5.2.

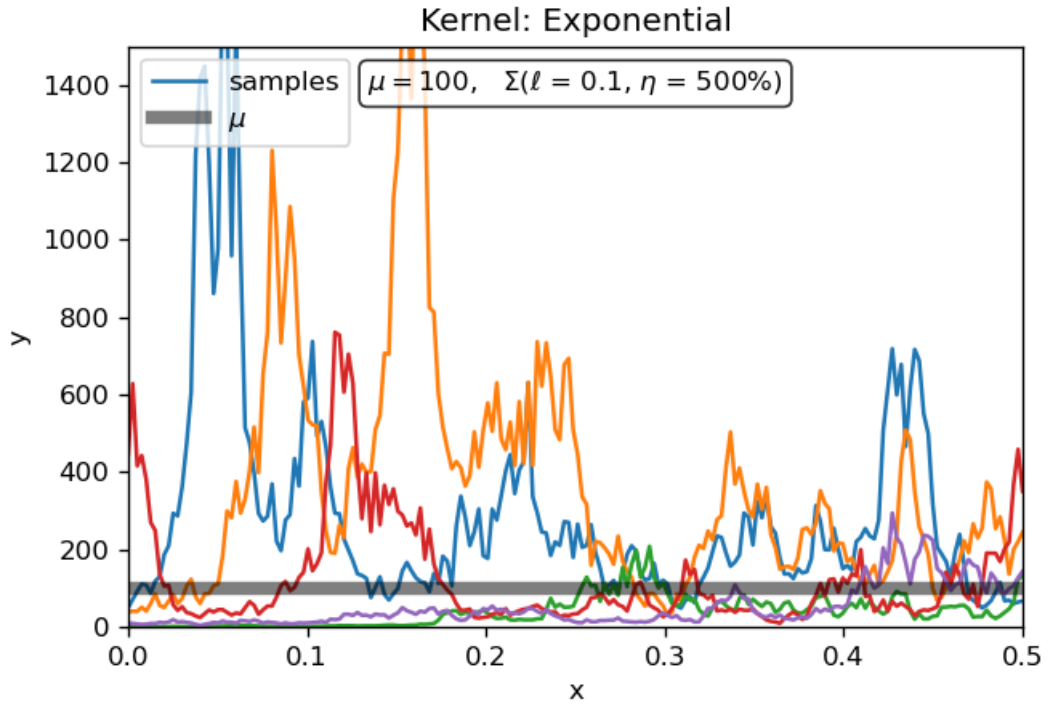


Fig. 2.4: Functions sampled from GP prior based on the exponential kernel (2.7). Transformed from log-space, see section 2.5.2.

## Chapter 3

# Synthetic TS data

### 3.1 Source of synthetic profiles

In order to generate synthetic TS data, a pair of electron temperature  $T_e$  and electron density  $n_e$  profiles is required. The first method that can be used to generate the profiles is to simulate plasma discharge in tokamak using available numerical models. Despite the fact that this approach would probably produce more precise and authentic results, it is unnecessarily complex and time demanding. Instead, an alternative approach was chosen, which utilizes the models that are used for fitting of TS profiles and determining important parameters.

#### 3.1.1 Model for TS profile fitting

A model for TS profile fitting that is currently adopted on the COMPASS tokamak is presented in [22]. It is tailored for H-mode temperature and density profiles. The H-mode is a mode of high plasma confinement that can be achieved on a tokamak with divertor configuration and supersedes the less effective L-mode. Currently, it is considered one of the main building blocks of successful tokamak fusion. A characteristic phenomenon of H-mode is the edge transport barrier, which decreases the diffusion of particles from plasma and counters various instabilities. This results in a modified shape of the plasma pressure profile, which appears to be *raised up on a pedestal* as seen in Fig. 3.1. The term *pedestal* is often used in reference to this concept. The shape of pressure profile also influences the temperature and density profiles in a similar manner.

The aforementioned TS profile fitting model consists of two main components. The first component is the function  $F_{\text{ped}}$ , which is based on modified hyperbolic tangent and describes the pedestal and the edge transport barrier using 5 independent parameters. The second component is an exponential function, which describes the plasma centre and has 3 parameters. The complete formula can be written as follows

$$F(r, a, b) = F_{\text{ped}}(r, b) + \{a_{\text{height}} - F_{\text{ped}}(r, b)\} \cdot e^{-\left(\frac{r}{a_{\text{width}}}\right)^{a_{\text{exp}}}}$$
$$F_{\text{ped}}(r, b) = \frac{b_{\text{height}} - b_{\text{SOL}}}{2} \left\{ \text{mtanh} \left( \frac{b_{\text{pos}} - r}{2b_{\text{width}}}, b_{\text{slope}} \right) + 1 \right\} + b_{\text{SOL}} \quad (3.1)$$
$$\text{mtanh}(x, b_{\text{slope}}) = \frac{(1 + b_{\text{slope}}x)e^x - e^{-x}}{e^x + e^{-x}},$$

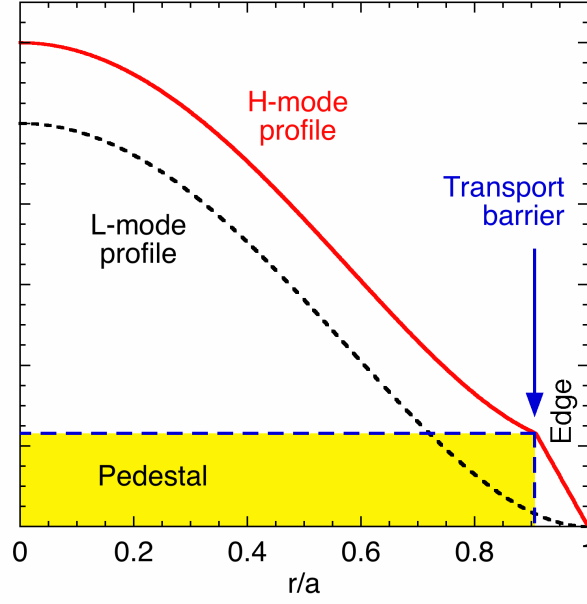


Fig. 3.1: Sketch of H-mode plasma profile and pedestal with comparison to L-mode.[23]

where parameters  $a_{\text{height}}$ ,  $a_{\text{width}}$  and  $a_{\text{exp}}$  describe the plasma centre and parameters  $b_{\text{height}}$ ,  $b_{\text{SOL}}$ ,  $b_{\text{width}}$ ,  $b_{\text{pos}}$ ,  $b_{\text{slope}}$  describe the pedestal. The precise interpretation of these parameters, which is discussed in [22], is not important in this context. An example plot of the function (3.1) is in Fig. 3.2.

The idea is that by generating a set of 16 parameters, 8 for temperature profile and 8 for density profile, it is possible to produce synthetic profiles which have properties similar to those of the real plasma profiles. However, note that the values of the parameters need to be restricted based on empirical experience and archived data in order to produce the expected results. Example of generated profiles is in Fig. 3.3.

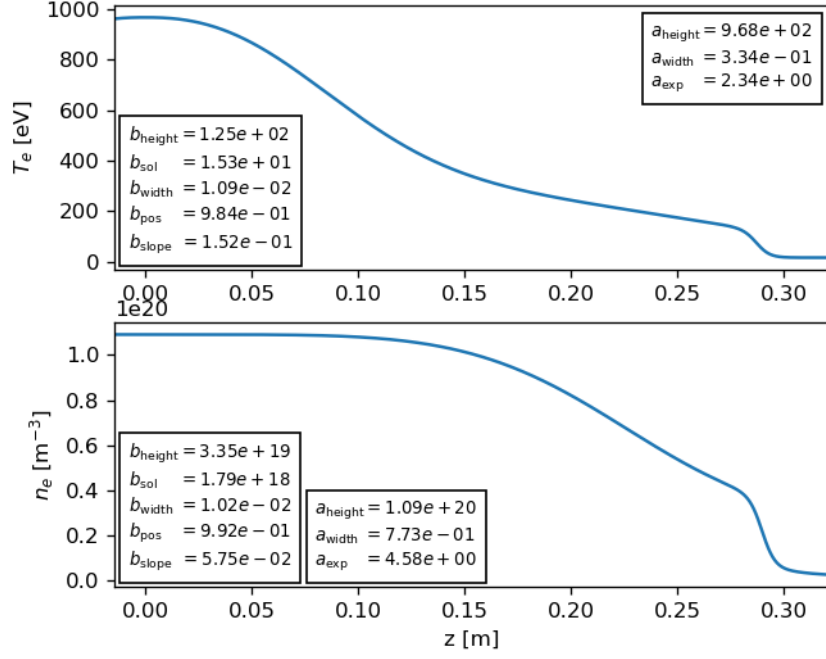


Fig. 3.2: An example of temperature  $T_e$  and density  $n_e$  profile function (3.1).

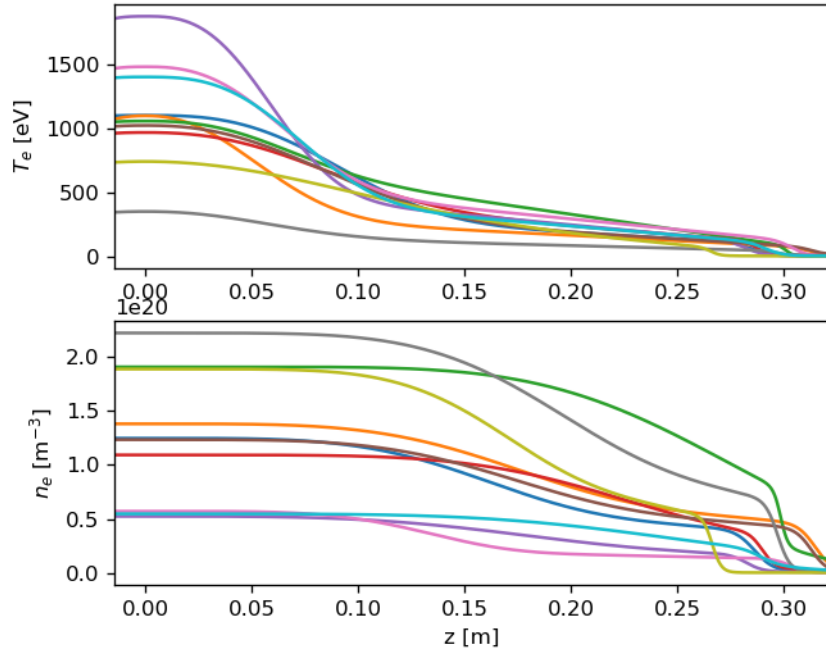


Fig. 3.3: A number of synthetic temperature  $T_e$  and density  $n_e$  profiles generated using the fitting function (3.1).



## 3.2 Simulating profile fluctuations

The fitting model (3.1) is an empirical approximation and captures only a limited subset of components that form a real plasma profile. In other words, better authenticity can be possibly achieved by introducing small continuous variations or fluctuations to the profile, which disturb its unrealistic smoothness and provide a crude imitation of stochastic components.

Gaussian processes (GP) and the multivariate normal, described in section 2.5, are well-suited for this purpose.<sup>1</sup> As proposed in section 2.5.2, an exponential transformation can be utilized to sample from classes of strictly positive functions. Then the fluctuation of the profile can be introduced as a relative multiplicative factor  $f(z)$

$$\tilde{A}(z) = f(z) \cdot A(z), \quad (3.2)$$

where  $A(z)$  is the original profile and  $\tilde{A}(z)$  is the modified profile with fluctuations.

GP requires a kernel and a mean function. Because the fluctuation modifier  $f(z)$  is a relative factor, the mean is set to 0 in log-space, which results in 1 or 100 % in the original space. The kernel recommended for this purpose is Matérn  $\frac{3}{2}$  or Matérn  $\frac{5}{2}$  because of their convenient degree of smoothness. Both kernels are parametrized by the length-scale parameter  $\ell$  and scaling factor  $\eta$ , which further specify the properties of the resulting fluctuations. Example of profiles with Matérn  $\frac{3}{2}$  simulated fluctuation is in Fig. 3.4.

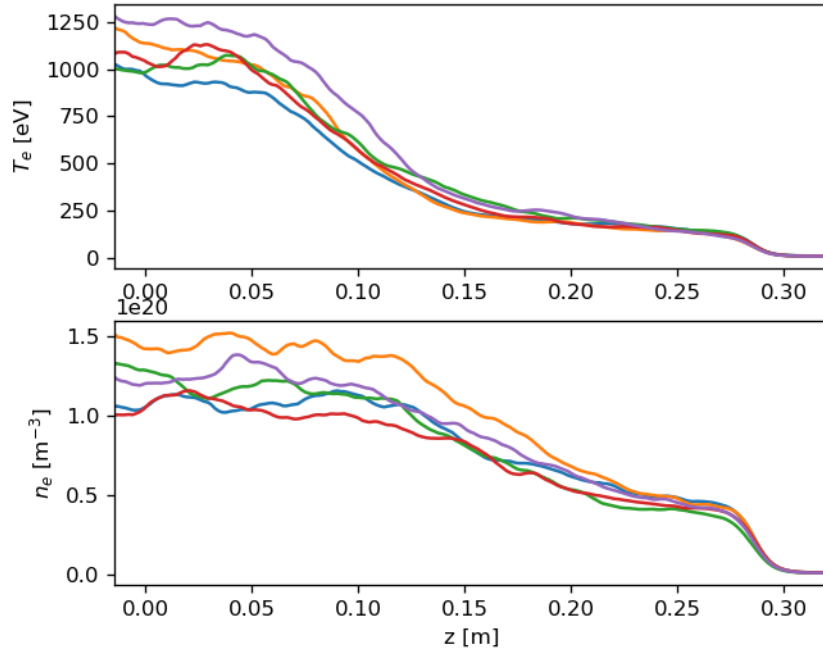


Fig. 3.4: An example of a synthetic temperature  $T_e$  and density  $n_e$  profile with simulated fluctuations (5 random samples).

<sup>1</sup>It could be argued that using a part of the inference model, the GP, to generate data is self-serving and logically must result in better apparent performance of the model. However, the alternative is to use somewhat more artificial methods, e.g. white noise, which are not particularly grounded in general physical understanding of the problem.

### 3.3 Simulating measurement noise

A measurement is always affected by noise and therefore a noise component should be incorporated into the synthetic data as well. Note that this refers to the intensity data of TS diagnostics that is calculated using the forward model (1.1), rather than the profile data discussed in the previous text. The simplest method to incorporate noise is to select a suitable probability distribution, specify its parameters and sample the noised data from it.

Usually the universal candidate is the normal distribution (A.1), where the mean parameter  $\mu$  represents the original data value and the standard deviation  $\sigma$  specifies the magnitude of the noise. However, the data of TS diagnostics is strictly positive, while the normal distribution produces real values. The requirement is fulfilled by the log-normal distribution (A.3), which can be parametrized by mean  $\mu$  and standard deviation  $\sigma$ , see (A.4) and (A.5), and applied in the same manner.

### 3.4 Overview of generated datasets

Based on empirical knowledge about the fitting model (3.1) and archived data, 10 sets of the fitting parameters  $a[...]$ ,  $b[...]$  that describe H-mode plasma profiles of  $T_e$  and  $n_e$  were generated. The profiles were scaled according to the  $z$  axis of the COMPASS tokamak and sampled at spatial points of the TS diagnostics. Then the synthetic TS data, i.e. the integrated intensity per channel, was derived using the TS forward model (1.1) and calibration data of an arbitrarily chosen discharge (#20555). Each profile was processed with 3 different configurations of the data generator resulting in 3 datasets:

- **Dataset 1:** no noise and no fluctuations
- **Dataset 2:** measurement noise  
Measurement noise is 5 % relative error + 0.005 absolute error.
- **Dataset 3:** measurement noise + profile fluctuations  
Measurement noise is 5 % relative error + 0.005 absolute error.  
The profile fluctuations are based on Matérn  $\frac{3}{2}$  kernel in log-space. Two components are used, the long-scale component  $\{\ell = 0.1 \text{ m}; \eta = 10 \%\}$  and the short-scale component  $\{\ell = 0.02 \text{ m}; \eta = 5 \%\}$ .

Notice that the magnitude of simulated measurement noise has a small absolute component, in addition to the relative error. It sets a lower bound for the error, which is inherent for any measurement device. The selected value 0.005 is roughly equal to the resolution digitalized signal from polychromator channels.

### 3.5 Quantifying the performance of inference method using synthetic data

The sole purpose of generating synthetic data is to assess the performance of different inference methods. That is accomplished by comparing the inferred  $T_e$  and  $n_e$  profiles with the original profiles that were used as a source for the data. To compare two 1-dimensional spatial profiles, it is necessary to select a norm.

In the case of continuous functions  $f_1 = f_1(z)$  and  $f_2 = f_2(z)$ , the  $L^2$  norm of their difference  $\Delta f = f_2 - f_1$

$$\|\Delta f\| = \left( \int_{\mathbb{R}} |\Delta f(z)|^2 dz \right)^{1/2} \quad (3.3)$$

is often employed in mathematics to quantify their similarity. Because the profile data is assumed to be a finite discretization of continuous functions, i.e.  $\tilde{f}_1 = \tilde{f}_1(z_i)$  and  $\tilde{f}_2 = \tilde{f}_2(z_i)$  where  $i = 1, 2, \dots, n$ ;  $n \in \mathbb{N}$ , an analogous norm was derived based on a discrete approximation of integral

$$\|\Delta \tilde{f}\| = \left( \sum_{i=1}^n |\Delta \tilde{f}(z_i)|^2 \cdot \frac{(\Delta z)_i}{z_n - z_0} \right)^{1/2}, \quad (3.4)$$

where  $z_1 < z_2 < \dots < z_n$  is the set of  $n \in \mathbb{N}$  of spatial points (on the  $z$  axis) which define the discretization. Notice the expression

$$\frac{(\Delta z)_i}{z_n - z_0},$$

which assumes the role of a weighing factor based on the distance between points. It is important to include this factor in order to compensate for the different spatial resolution of the core TS and the edge TS area, which is discussed in section 1.2. If the variations caused by the viewing angle are neglected, then the weighing factor can be approximated as

$$\begin{aligned} (\Delta z)_i &= \begin{cases} 3.7 \text{ mm}, & \text{if } z_i \text{ belongs to the edge area} \\ 9.9 \text{ mm}, & \text{if } z_i \text{ belongs to the core area} \end{cases} \\ z_n - z_0 &= 336.0 \text{ mm} \end{aligned} \quad (3.5)$$

A simple difference of the original synthetic profile and the inferred profile  $\|\Delta \tilde{f}\|$  is unsuitable for an objective statistical summary, because it logically depends on the synthetic profile itself. To resolve this, the difference needs to be normalized by an arbitrary factor  $\xi$  derived from the profile data (original or inferred). The norm of normalized difference

$$\left\| \frac{\Delta \tilde{f}}{\xi} \right\| \quad (3.6)$$

can then be applied on profiles inferred from the synthetic datasets (see section 3.4) to summarize and assess the inference method. For this purpose, two summarizing indicators were devised that focus on the precision of results and error estimation:

### 3.5.1 Indicator 1: Precision of the result

The normalizing factor  $\xi$  is the maximal value in the synthetic profile  $\tilde{f}_{\text{synth}}$ .

$$\xi^{(1)} = \max \left\{ \tilde{f}_{\text{synth.}}(z_i) \mid i = 1, 2, \dots, n \right\}. \quad (3.7)$$

This indicator describes how similar is the inferred profile and the original synthetic profile. The value of the norm (3.6) itself does not have any specific absolute scale in this case, but it can be used to compare different inference methods or their performance on different datasets.

### 3.5.2 Indicator 2: Precision of error estimate

Another option is to select the estimated error at given point as the normalizing factor  $\xi$ . An error is usually estimated as the standard deviation  $\sigma$ . However, in Bayesian statistics the highest density interval (hdi), which is more universal, is used instead. An equivalent of the standard deviation is the half of a 68 % hdi, labelled as  $\sigma_{68}$

$$\xi^{(2)} = \sigma_{68}(z_i) \tag{3.8}$$

This indicator shows whether the error estimate  $\sigma_{68}$  is reasonable. If a value of the norm (3.6) with this factor  $\xi^{(2)}$  is close to 1, then the difference between the original and the inferred profile is on average approximately  $\sigma_{68}$ . If the value is significantly lower or higher than 1, the errors are over- or underestimated respectively.

## Chapter 4

# Results: Point inference

The topic of this research task is the process of inferring electron temperature  $T_e$  and density  $n_e$  spatial profiles from data produced by Thomson scattering diagnostics on the COMPASS tokamak. Two inference models of different complexity were proposed and implemented. The first model, described in this chapter, is less complex and implements inference of the spatial profile on a per-point basis without assuming any correlation between the points. Its purpose is to emulate the process of the conventional regression method, which is currently used on the COMPASS tokamak, with the use of the Bayesian inference.

### 4.1 Model description

The unknown of this model is a pair of  $T_e$  and  $n_e$  values at several spatial points where data was measured. Because no spatial correlation between the points is assumed, the complexity of the model is effectively lowered from the 1-dimensional inference of a profile to the 0-dimensional inference of a single point.

#### 4.1.1 Prior distribution

The prior distribution of  $T_e$  and  $n_e$  is the half-normal distribution with  $\sigma$  parameters equal to arbitrarily chosen order-of-magnitude estimates

$$T_e \sim \text{half-normal}(\sigma = 1000 \text{ eV}) \quad (4.1)$$

$$n_e \sim \text{half-normal}(\sigma = 10^{20} \text{ m}^{-3}) \quad (4.2)$$

Half-normal distribution, see appendix A.1, is considered to be a weakly-informative prior implying that it maintains the objectivity of the inference. Another property of this distribution is its positive domain, which is required by the physical context of temperature and density. Apart from half-normal distribution, other variants were considered including half-Cauchy and log-normal distributions, see Fig. 4.1 for comparison. While the half-normal distribution was selected as the most appropriate, preliminary testing had showed that all of the considered options produce very similar results.

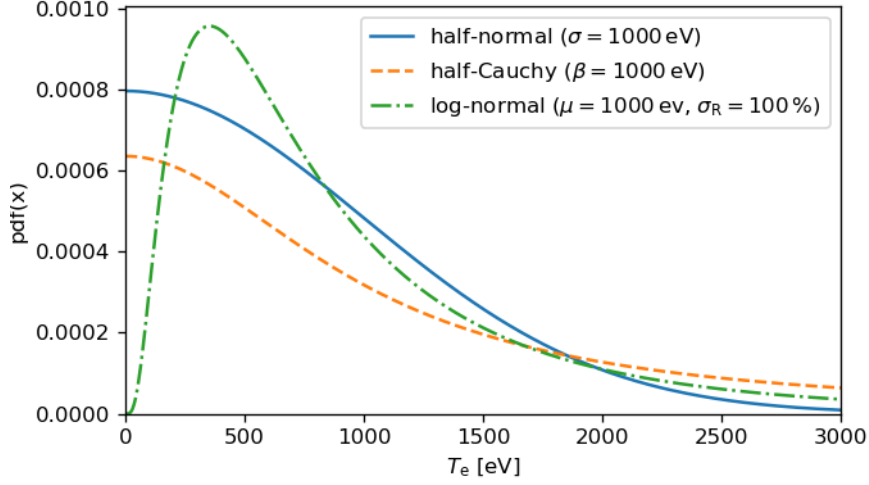


Fig. 4.1: Comparison of different options for prior distribution of  $T_e$  in the *point* model. The plot shows that the half-Cauchy distribution constrains large  $T_e$  values less than the half-normal and that the log-normal distribution significantly disadvantages near-zero values.

#### 4.1.2 Likelihood

The second building block of a model, the likelihood, is defined by the log-normal distribution with transformed parameters  $\mu$ ,  $\sigma_R$  as specified in appendix (A.3)

$$V_i \sim \text{log-normal} \left( \mu = f_{\text{TS}}(T_e, n_e), \sigma_R = \frac{D_i^{(\text{err})}}{D_i} \right). \quad (4.3)$$

Here  $f_{\text{TS}} = f_{\text{TS}}(T_e, n_e)$  refers to the forward model of TS diagnostics specified by the equation (1.1) and employed by the regression method currently used on the COMPASS tokamak.  $V_i$  is the intensity in the  $i$ -th polychromator channel (see section 1.3) as a random variable, while  $D_i$  and  $D_i^{(\text{err})}$  refer to the actual data: the measured value of  $V_i$  and its error estimate respectively. Although the normal distribution is a popular universal choice, the log-normal distribution was chosen instead to better reflect the fact that the intensity values  $V_i$  are always positive. This brings more stability to the algorithm. Apart from log-normal, several other distributions with positive domain were considered, see Fig. A.4 for comparison.

#### 4.1.3 Inference algorithms

As a result of low complexity of the model, it was possible to utilize the posterior sampling method NUTS to infer the results without the need of an unreasonable amount of computational time. It was observed that the variational method ADVI brings only small increase in computational speed in this case.

## 4.2 Inference on simulated data

The inference model was tested on 3 synthetic datasets, which are described in chapter 3.4. The two indicator quantities from section 3.5 were used to compare inferred profiles

with the original synthetic profiles. The results from each dataset were summarized using mean and standard deviation and plotted to graphs, see Fig. 4.2.

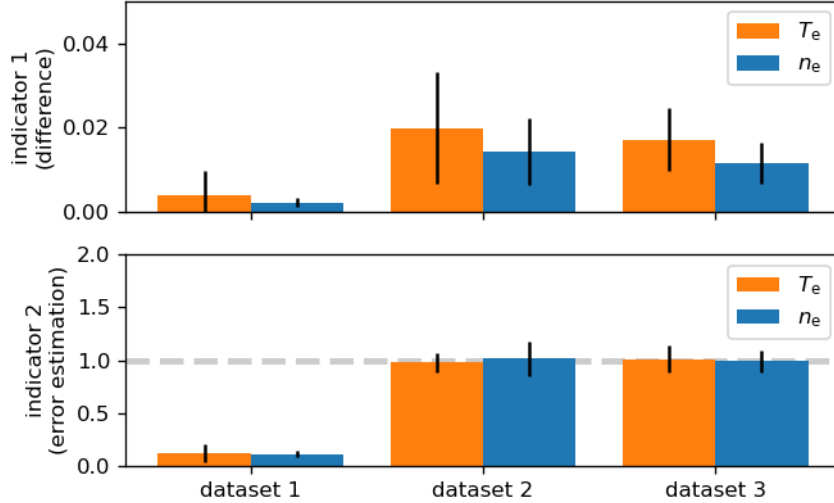


Fig. 4.2: Summarized performance of inference model *point* when applied on synthetic datasets 1, 2 and 3. The top plot describes the difference from the original synthetic profile. The bottom plot shows, whether the errors are over- or underestimated. See section 4.2 for more details.

Note that only a subset of 44 spatial points (out of 54) was used for calculating these statistics in order to avoid the far-edge area, where the signal is very low. Because of the low signal the inference method loses reliability. This is discussed in detail in section 6.3. Apart from that, the graphs in Fig. 4.2 seem to indicate that the inference method estimates its errors well (values of indicator 2 are close to 1.0) and is able to approximate the synthetic profile. Both indicators are very small in the case of dataset 1 (no measurement error) as expected. The effect of the profile fluctuations added in dataset 3 is inconclusive with respect to the estimated standard deviation of the indicators.

A general comparison with the other inference method and with the data processing routine used on the COMPASS tokamak is in Fig. 6.1.

### 4.3 Inference on data from database

In addition to synthetic data, the inference model was also applied on archived data from three standard shots on the COMPASS tokamak (shots # 20447, # 20555 and # 20664). An example of inferred temperature and density profile is in Fig. 4.3. Results of the COMPASS data processing routine are added for comparison.

The example in Fig. 4.3 illustrates that the results of the Bayesian method and the COMPASS processing routine are very similar in respect to the density profile. Minor differences usually appear in the core area of the temperature profile. In addition to that, very low density in the far edge area often causes significant deviations. This is further discussed in section 6.3.

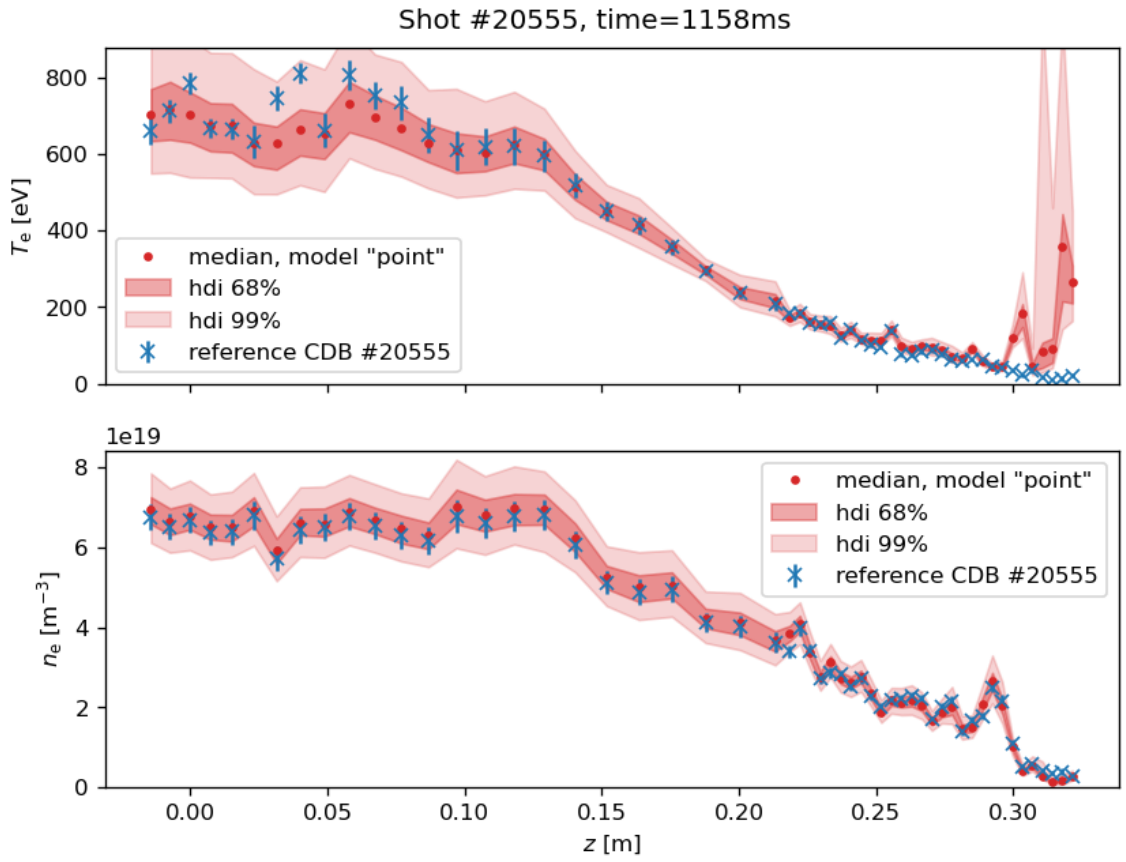


Fig. 4.3: Plasma temperature and density profile inferred from archived data from the COMPASS database (CDB) using the *point* inference model, section 4. Hdi = highest density interval.



## Chapter 5

# Results: Profile inference with Gaussian processes

This chapter is focused on the description and implementation of the second model. In comparison with the first model (Point inference), it adds the complexity of spatial correlation between points of the profile. The motivation is to include more information in the process of inference. The correlation is specified using an advanced type of prior distribution that is based on Gaussian processes.

### 5.1 Model description

The unknown of this model is a continuous 1-dimensional spatial profile of  $T_e$  and  $n_e$ .

#### 5.1.1 Prior distribution

The prior distribution is implemented as Gaussian processes (GP), see section 2.5. Because both temperature and density can be expected to be always positive, the exponential transformation (2.8) is utilized. The mean vector  $\vec{\mu}$  is specified by a constant generating function  $\mu = \mu(z)$  in order to provide a simple order-of-magnitude estimate for the algorithm similarly to (4.1) and (4.2)

$$\mu_{T_e}(z) = 100 \text{ eV} \quad (5.1)$$

$$\mu_{n_e}(z) = 10^{19} \text{ m}^{-3} \quad (5.2)$$

The Matérn  $\frac{3}{2}$  kernel (2.6) is used for the covariance matrix  $\Sigma$ . Illustrative examples of functions sampled from a GP distribution with exponential transformation described by this kernel and a constant mean function can be seen in Fig. 2.3. The covariance is parametrized by the length-scale parameter  $\ell$  and the transformed scale parameter  $\eta$ , see (2.9).

Because of the exponential transformation of GP, the covariance reflects the *relative* behaviour of the functions, while their absolute scale is determined by the mean. It is expected that the general *relative* behaviour of  $T_e$  and  $n_e$  profiles is comparable and thus, for convenience, identical weakly-informative prior distributions can be used for hyperparameters of both covariances  $\Sigma_{T_e}$  and  $\Sigma_{n_e}$

$$\ell \sim \text{gamma}(\mu = 0.1 \text{ m}, \sigma = 0.05 \text{ m}) \quad (5.3)$$

$$\eta \sim \text{half-Cauchy}(\beta = 1000 \%) \quad (5.4)$$

The priors are plotted in Fig. 5.1. Their parameters were chosen based on empirical knowledge about the kernel (see Fig. 2.3) and plasma profiles on the COMPASS tokamak.

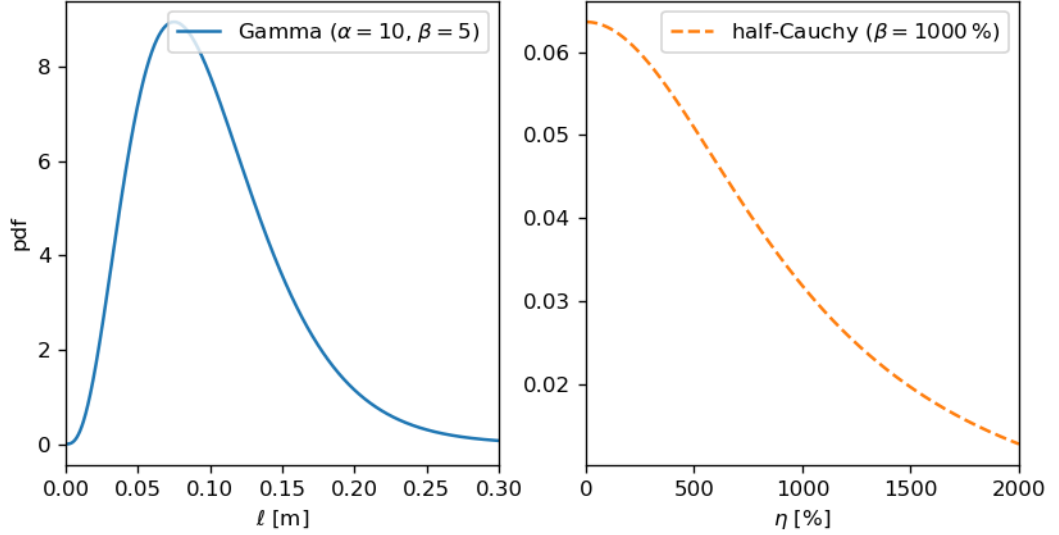


Fig. 5.1: Prior distribution of hyperparameters  $\ell$  and  $\eta$  of the GP prior in the *profile* model, see chapter 5.

### 5.1.2 Likelihood

The likelihood is identical to that of Point inference model, which was the combination of log-normal distribution (4.3) and the forward model specified by (1.1). According to the semantics of Bayesian models, the likelihood describes the experimental method which is the same for both models (the same data is used). The sole difference between the models is in the prior knowledge and expectations about  $T_e$  and  $n_e$  profiles, which is encoded in the prior distributions.

### 5.1.3 Inference algorithms

Models based on GP bring additional computational complexity. In this case, the computational time required by the posterior sampling method NUTS varied significantly depending on the presence of data points with low signal-to-noise ratio, i.e. the points where density  $n_e$  was low. Consequently, variational method ADVI was employed for the purpose of the larger-scale statistics on synthetic data.

## 5.2 Inference on simulated data

In analogy to section 4.2, the profile inference model was also tested on the 3 synthetic datasets from chapter 3.4 with the use of indicator quantities from section 3.5. The subset of 44 (out of 54) spatial points was also used in order to avoid the known negative effects low signal-to-noise ratio, which are discussed in 6.3. The summarized results are plotted in Fig. 5.2.

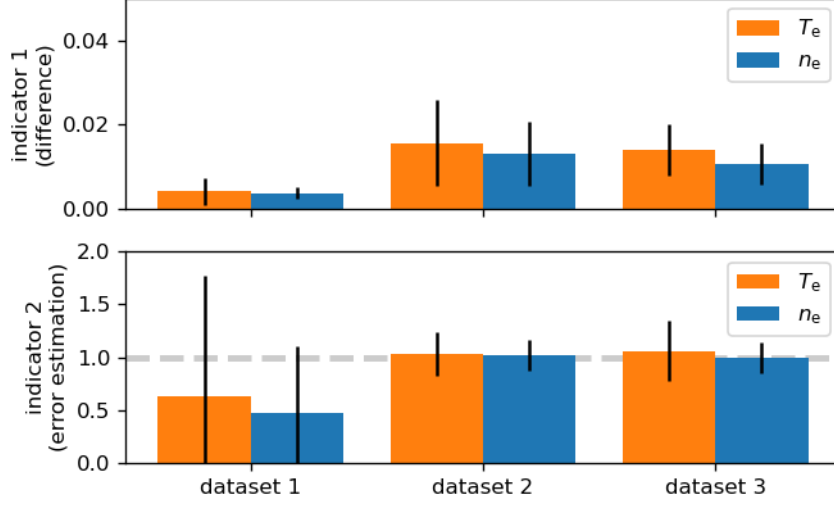


Fig. 5.2: Summarized performance of inference model *profile* when applied on synthetic datasets 1, 2 and 3. The top plot describes the difference from the original synthetic profile. The bottom plot shows, whether the errors are over- or underestimated. See section 4.2 for more details.

The results in Fig. 5.2 show that the profile model estimates the errors well, because the indicator 2 is close to value 1.0. The difference from the synthetic profile is slightly larger in case of the temperature  $T_e$ . However, compared to the point model results shown in Fig. 4.2, the profile model performs better.

A general comparison of both inference methods and the data processing routine used on the COMPASS tokamak is in Fig. 6.1.

### 5.3 Inference on data from database

The *profile* inference model was also applied on archived data from three standard shots on the COMPASS tokamak (shots # 20447, # 20555 and # 20664). An example of inferred temperature and density profile is in Fig. 5.3. Once again, the unstable edge behaviour is present. Results of the COMPASS data processing routine are added for comparison.

As was discussed in section 2.5, the GP prior distribution relates to the space of *functions* and therefore in theory the true result is a function. The results shown in Fig. 5.3 can be thought of as the estimation of several function values at points where the data was measured. The information about the function itself is contained in the posterior distribution of parameters of the covariance matrix  $\ell$  and  $\eta$ . It is possible to utilize this information and process the results further in order to robustly predict posterior distribution of the function values at different spatial points, for example in between the original points. This is referred to as the posterior predictive sampling, see [20]. Profiles from Fig. 5.3 that were predicted at 200 new equidistant points are shown in Fig. 5.4.

The highest density interval of profiles in Fig. 5.4 features a distinctive wave-like modulation, which reflects the fact that the uncertainty rises with the distance from points where the data was measured. The magnitude of this effect depends on the length-scale  $\ell$  parameter of the GP kernel with respect to the distance between the measurement points.

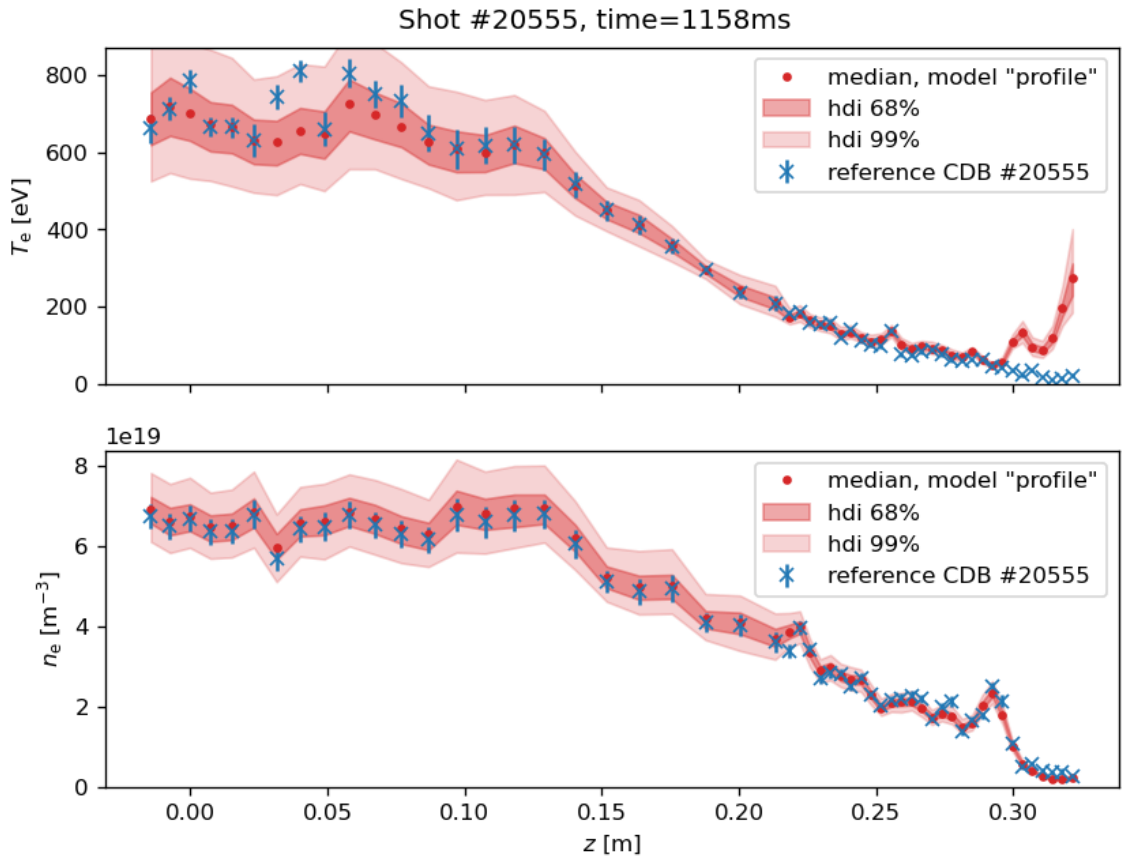


Fig. 5.3: Plasma temperature and density profile inferred from archived data from the COMPASS database (CDB) using the *profile* inference model, section 5. Hdi = highest density interval.

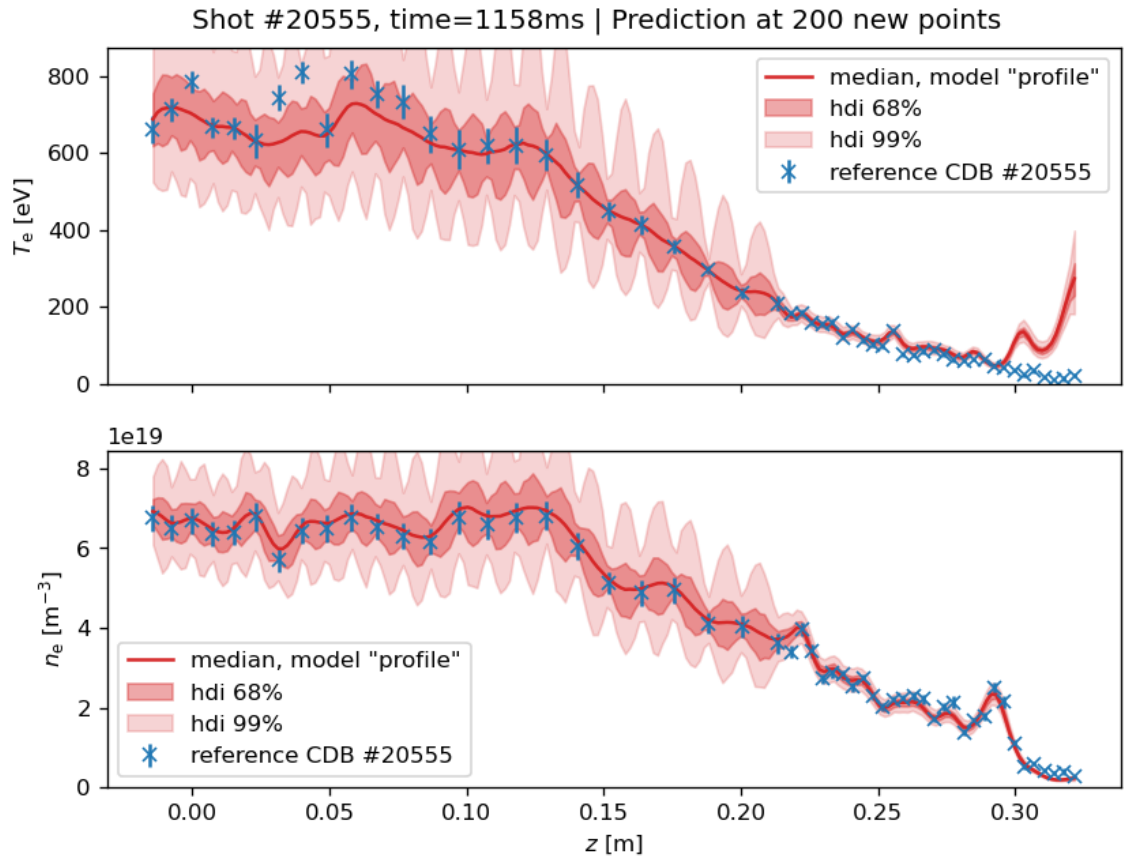


Fig. 5.4: Prediction of temperature and density profile at 200 new  $z$  coordinates based on inference from Fig. 5.3. Hdi = highest density interval.

Therefore the points in edge area are far less affected.

# Chapter 6

## Discussion

### 6.1 The development of the models

Sections 4 and 5 present the final form of the models. However, numerous alternatives were explored and tested.

The likelihood for both models was defined by the log-normal distribution. It was selected because of its positive domain and because its properties are similar to those of normal distribution when  $\sigma \ll \mu$ . The positive domain is in compliance with the fact that the TS data is composed of positive light intensity values. In comparison to the likelihood based on normal distribution, the posterior sampling algorithm was more stable, i.e. divergent samples were produced less often. Despite that, several preliminary test-results indicated that the normal distribution is not afflicted by the edge instability (see section 6.3) as much as the log-normal. Therefore the properties of log-normal distribution seem to be the cause of the instability.

The GP covariance of the profile model is defined using the Matérn  $\frac{3}{2}$  kernel. As an alternative, the exponentiated quadratic (EQ) was tested. However, it was very unstable and produced significant amount of divergent samples. To counter that, a combined covariance given by a sum of these two kernels was also introduced. However, it was discovered that the EQ kernel is usually overpowered by the Matérn kernel and does not contribute to the result.

### 6.2 Performance of the inference models based on the synthetic datasets

Comparison of performance of both inference models and the software used on the COMPASS tokamak is plotted in Fig. 6.1. The performance is quantified by the two indicators defined in section 3.5.

Based on the indicator 2 (bottom plot) it can be argued that the estimation of errors does not present an issue for any of the models. The difference from the synthetic profile described by the indicator 1 (top plot) is mostly constant in the case of density  $n_e$ . This is expected because in the forward model (1.1) the density holds the role of a simple linear factor and therefore it does not pose problems for the algorithms.

By comparing the temperature  $T_e$  profiles it can be claimed that the spatial correlation of points in the *profile* model does improve its performance over the *point* model. However, the COMPASS software appears to perform better. It is not conclusive whether the Bayesian approach offers less precision or whether the models simply require more tuning

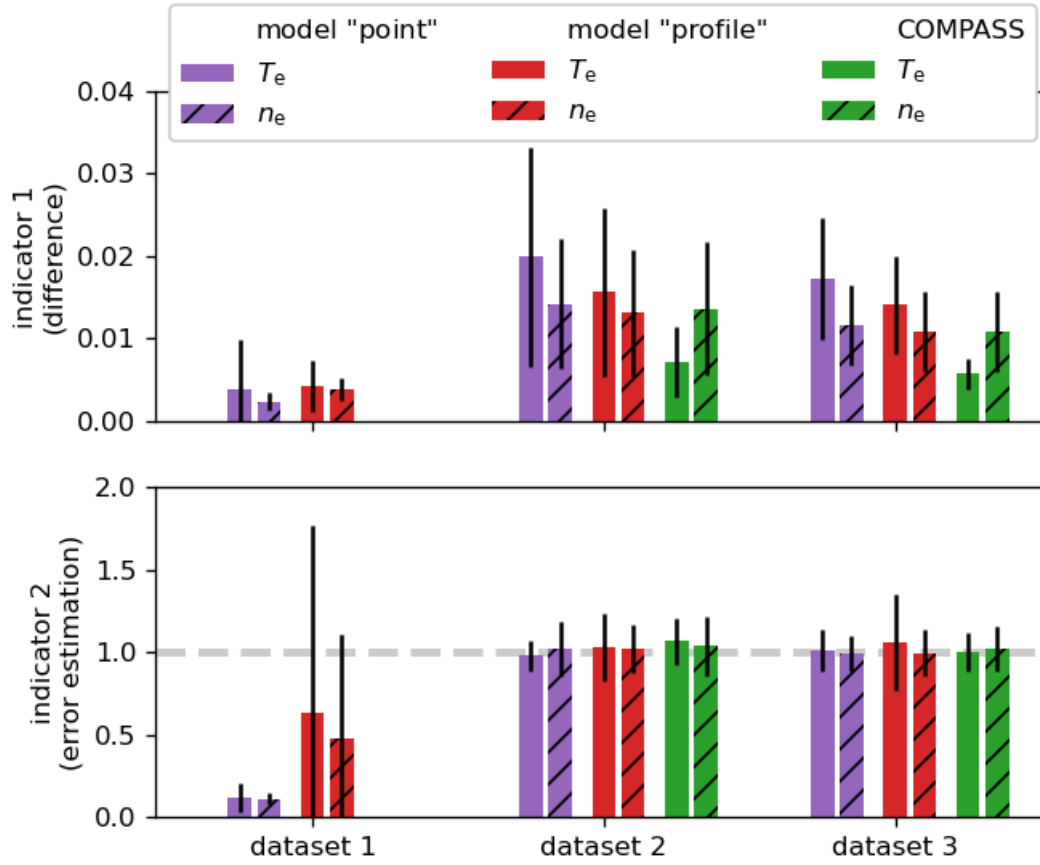


Fig. 6.1: Comparison of performance of the two inference methods, shown separately in Fig. 4.2 and Fig. 5.2, and the software used on the COMPASS tokamak. The top plot describes the difference from the original synthetic profile. The bottom plot shows, whether the errors are over- or underestimated.



and longer calculation (i.e. more posterior samples). Note that the results for *profile* model were calculated using the approximative algorithm ADVI because the full posterior sampling was deemed to be unnecessarily time intensive for this purpose.

### 6.3 The unstable behaviour in edge areas with low signal-to-noise ratio

Both Bayesian models are afflicted with unstable behaviour in areas, where the signal-to-noise ratio is very low (usually far edge area). This results in significant deviations in the temperature profile as can be seen in Fig. 4.3, Fig. 5.3 and Fig. 5.4. It also causes divergent samples. The deviations can be observed both in results inferred from COMPASS database data and from synthetic data with high simulated measurement error.

In section 6.1 it was discussed that the properties of log-normal likelihood could represent the main cause. If that is the case, new likelihood options should be researched and tested. Although it is possible that all conventional likelihoods with positive domain share a similar property that causes this undesirable phenomenon.

Finally, the situation can be approached more pragmatically by discarding all points with signal-to-noise ratio below certain threshold. It can be argued that data with low signal-to-noise does not have a potential to produce reliable results in general.

### 6.4 Processing archived data from COMPASS database

Both models were successfully applied on the data from COMPASS database, the results are shown in Fig. 4.3, Fig. 5.3 and Fig. 5.4. Apart from the edge behaviour discussed in (6.3), both models produce results in good agreement with the COMPASS software. Minor deviations appear mainly in the temperature profile. The difference might be caused because of a stray-light compensation algorithm, which is implemented in the COMPASS software but not in the Bayesian models.

In summary, the models are from a technical standpoint prepared for use on the data from COMPASS database. They are, however, in an experimental state and a further research is required to guarantee their robustness and precision.

### 6.5 Further development

The main purpose of the *point* model was to provide an additional reference point when comparing the *profile* model with COMPASS software. The only motivation for its further development is the potential research of alternative likelihood options discussed in section 6.3.

The main focus is the *profile* model. The current conclusion derived from results in Fig. 6.1 is that it lacks precision in respect to the temperature profiles. First step would be to confront the edge instabilities described in section 6.3. Improvement can also be achieved through an empirical research of optimal settings for the posterior sampling algorithm (number of tuning steps and sample draws).

In addition, the statistical survey of model performance (section 6.2) should be extended in volume in order to provide more conclusive results that can be used to steer the further research.

# Conclusion

A brief introduction to the theory of Thomson scattering (TS) and its application in high temperature plasma diagnostics was presented in chapter 1. The TS forward model (1.1) was introduced and the fundamental components TS diagnostic system on the COMPASS tokamak were described.

The concept of Bayesian inference and its application for data analysis was presented in chapter 2. The library PyMC3 for the Python language that implements a framework for constructing inference models and posterior sampling algorithms was introduced. A description of Gaussian processes and their function was offered.

The model for fitting of H-mode electron temperature and density profiles on the COMPASS tokamak was utilized to produce 3 synthetic datasets of TS data (chapter 3). Two Bayesian inference models, the *point* model and the *profile* model, were formulated. The *point* model was less complex and roughly emulated the native software on the COMPASS tokamak. The *profile* model incorporated spatial correlation between the points of the profile with the use of Gaussian processes. The models were described in detail in chapters 4 and 5. The synthetic datasets were utilized to compare performance of these models and the native COMPASS software, see Fig. 6.1. An example of inference of data from the COMPASS database was presented as well, see Fig. 4.3, Fig. 5.3 and Fig. 5.4.

In conclusion, two Bayesian models were successfully implemented and tested on both synthetic data and the COMPASS database data. A common instability of the models in edge area, which was probably caused by a low signal-to-noise ratio, was observed. In addition to that, it was indicated that the performance of the models in terms of precision was lower than the performance of the COMPASS software. A set of possible causes concerning both mentioned topics was discussed. A direction of further research with the objective of improving the models, primarily the *profile* model, was proposed.

# Bibliography

- [1] G. von Nessi and M. Hole, “Using Bayesian analysis and Gaussian processes to infer electron temperature and density profiles on the Mega-Ampere Spherical Tokamak experiment,” *Review of Scientific Instruments*, vol. 84, no. 6, p. 063505, 2013.
- [2] R. Fischer, A. Dinklage, and E. Pasch, “Bayesian modelling of fusion diagnostics,” *Plasma physics and controlled fusion*, vol. 45, no. 7, p. 1095, 2003.
- [3] M. Aftanas, P. Bohm, P. Bilkova, V. Weinzettl, J. Zajac, F. Zacek, J. Stockel, M. Hron, R. Panek, R. Scannell, and M. J. Walsh, “High-resolution Thomson scattering system on the COMPASS tokamak: Evaluation of plasma parameters and error analysis,” *Review of Scientific Instruments*, vol. 83, no. 10, p. 10E350, 2012.
- [4] V. Weinzettl, J. Adamek, P. Bilkova, J. Havlicek, R. Panek, M. Hron, O. Bogar, P. Bohm, A. Casolari, J. Cavalier, *et al.*, “Constraints on conceptual design of diagnostics for the high magnetic field COMPASS-U tokamak with hot walls,” *Fusion Engineering and Design*, vol. 146, pp. 1703–1707, 2019.
- [5] J. Hecko, *Error analysis of plasma diagnostic system based on Thomson scattering on the COMPASS tokamak*. CTU: Bachelor’s thesis, 2019.
- [6] S. L. Prunty, “A primer on the theory of Thomson scattering for high-temperature fusion plasmas,” *Physica Scripta*, vol. 89, no. 12, p. 128001, 2014.
- [7] J. Sheffield *et al.*, *Plasma Scattering of Electromagnetic Radiation*. Burlington, USA: Academic Press, 2011.
- [8] P. Bohm, *Temporally and spatially resolved evolution of plasma in the COMPASS tokamak*. PhD thesis, CTU, 2012.
- [9] P. Bilkova, P. Bohm, M. Aftanas, M. Sos, A. Havranek, D. Sestak, V. Weinzettl, M. Hron, R. Panek, and the COMPASS team, “High resolution thomson scattering on the COMPASS tokamak—extending edge plasma view and increasing repetition rate,” *Journal of Instrumentation*, vol. 13, no. 01, p. C01024, 2018.
- [10] A. C. Selden, “Simple analytic form of the relativistic Thomson scattering spectrum,” *Physics Letters A*, vol. 79, no. 5, p. 405 – 406, 1980.
- [11] M. Sos, *Optimisation of processing raw data from Thomson scattering diagnostic on the COMPASS tokamak*. CTU: Master’s thesis, 2018.
- [12] A. B. Downey, *Think Bayes*. Sebastopol, USA: O’Reilly Media, 2013.
- [13] O. Martin, *Bayesian Analysis with Python*. Birmingham, UK: Packt Publishing, 2013.

- [14] W. Hastings, “Monte Carlo sampling methods using Markov chains and their applications,” *Biometrika*, vol. 57, no. 1, pp. 97–109, 1970.
- [15] S. Duane, A. D. Kennedy, B. J. Pendleton, and D. Roweth, “Hybrid Monte Carlo,” *Physics Letters B*, vol. 195, no. 2, pp. 216–222, 1987.
- [16] M. D. Hoffman and A. Gelman, “The No-U-Turn sampler: Adaptively setting path lengths in Hamiltonian Monte Carlo,” *Journal of Machine Learning Research*, vol. 15, pp. 1593–1623, 2014.
- [17] B. Carpenter, A. Gelman, M. D. Hoffman, D. Lee, B. Goodrich, M. Betancourt, M. Brubaker, J. Guo, P. Li, and A. Riddell, “Stan: A probabilistic programming language,” *Journal of Statistical Software*, vol. 76, 1 2017.
- [18] E. Bingham, J. P. Chen, M. Jankowiak, F. Obermeyer, N. Pradhan, T. Karaletsos, R. Singh, P. Szerlip, P. Horsfall, and N. D. Goodman, “Pyro: Deep universal probabilistic programming,” *J. Mach. Learn. Res.*, vol. 20, p. 973–978, Jan. 2019.
- [19] C. F. J. Salvatier, T.V. Wiecki, “Probabilistic programming in Python using PyMC3,” *PeerJ Computer Science*, vol. 2, p. e55, 2016.
- [20] The PyMC Development Team, “PyMC3 Documentation.” <https://docs.pymc.io/>. [accessed September 15, 2020].
- [21] C. E. Rasmussen and C. K. I. Williams, *Gaussian Processes for Machine Learning*. Cambridge, USA: The MIT Press, 2006.
- [22] E. Stefanikova, M. Peterka, P. Bohm, P. Bilkova, M. Aftanas, M. Sos, J. Urban, M. Hron, and R. Panek, “Fitting of the Thomson scattering density and temperature profiles on the COMPASS tokamak,” *Review of Scientific Instruments*, vol. 87, no. 11, p. 11E536, 2016.
- [23] “Pedestal — FusionWiki.” <http://fusionwiki.ciemat.es/wiki/Pedestal/>. [accessed September 15, 2020].

## Appendix A

# List of probability distributions

This appendix presents a brief overview of continuous probability distributions that were used or mentioned in this research task.

### A.1 Normal and half-normal distribution

The probability distribution function (pdf) of normal distribution is defined as

$$f(x | \mu, \sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp \left\{ -\frac{(x - \mu)^2}{2\sigma^2} \right\} \quad (\text{A.1})$$

where  $\mu$  denotes the mean (and median) of random variable  $X$  and  $\sigma$  is its standard deviation. Half-normal distribution is created from normal distribution with  $\mu = 0$  by truncating it to positive domain. Probability distribution functions of normal and half-normal distribution is plotted for comparison in Fig. A.1.

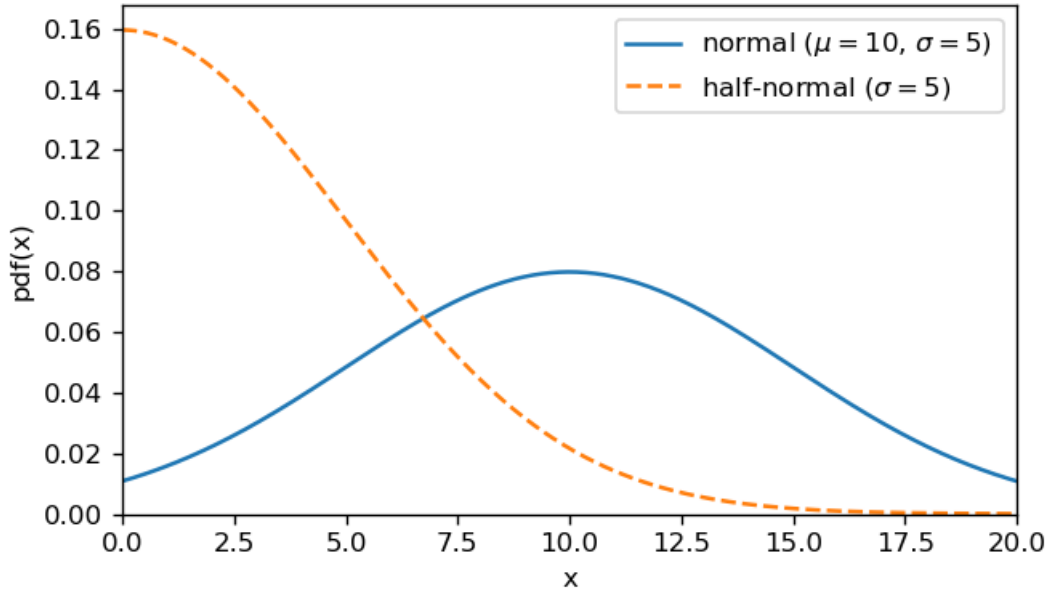


Fig. A.1: Probability distribution function of normal and half-normal distributions.

## A.2 Cauchy and half-Cauchy distribution

The probability distribution function (pdf) of Cauchy distribution is defined as

$$f(x | \alpha, \beta) = \frac{1}{\pi\beta \left[1 + \left(\frac{x-\alpha}{\beta}\right)^2\right]} \quad (\text{A.2})$$

where  $\alpha$  and  $\beta$  are parameters which are analogous to mean  $\mu$  and standard deviation  $\sigma$  respectively. However, it should be noted that neither mean nor standard deviation can be defined in the context of the Cauchy distribution. It is often described as being heavy-tailed and because of that it can be used as a less restrictive alternative to the normal distribution.

Half-Cauchy is created from Cauchy distribution with  $\alpha = 0$  by truncating it to positive domain. Both probability distribution functions are plotted in Fig. A.2.

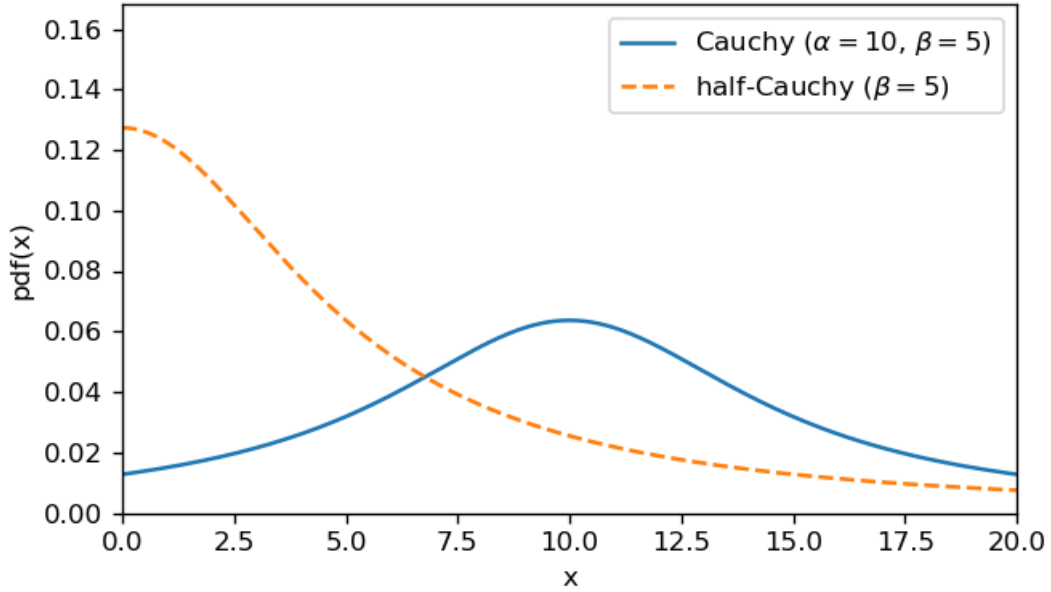


Fig. A.2: Probability distribution function of Cauchy and half-Cauchy distributions.

## A.3 Log-normal distribution

Distribution of any random variable  $X$  whose logarithm  $\ln(X)$  is normally distributed. A variable might be modeled as log-normal if it can be thought of as the multiplicative product of many small independent factors.[20]

Log-normal distribution has positive domain. Its probability distribution function is

$$f(x | \tilde{\mu}, \tilde{\sigma}) = \frac{1}{x} \frac{1}{\sqrt{2\pi\tilde{\sigma}^2}} \exp \left\{ -\frac{(\ln(x) - \tilde{\mu})^2}{2\tilde{\sigma}^2} \right\}. \quad (\text{A.3})$$

The native  $\tilde{\mu}$  and  $\tilde{\sigma}$  parameters refer to the mean and standard deviation of the underlying normal distribution of  $\ln(X)$ . These parameters are used in the implementation of log-normal distribution in the PyMC3[19] library.

In some cases it is convenient to introduce different parametrization based on mean  $\mu$  and standard deviation  $\sigma$  of  $X$ . Relation between these and the native parameters is given by the following set of transformation equations

$$\tilde{\mu} = \ln(\mu) - \frac{1}{2} \ln \left( 1 + \frac{\sigma^2}{\mu^2} \right) \quad (\text{A.4})$$

$$\tilde{\sigma} = \sqrt{\ln \left( 1 + \frac{\sigma^2}{\mu^2} \right)} \quad (\text{A.5})$$

Notice that parameter  $\sigma$  always appears in a form of expression  $\frac{\sigma^2}{\mu^2}$  in both equations. Therefore, it is straightforward to derive a third parametrization, which uses a relative standard deviation  $\sigma_R$

$$\sigma_R = \frac{\sigma}{\mu}. \quad (\text{A.6})$$

Example of probability distribution function of log-normal distribution parametrized as  $f(x | \mu, \sigma_R)$  is plotted in Fig. A.3.

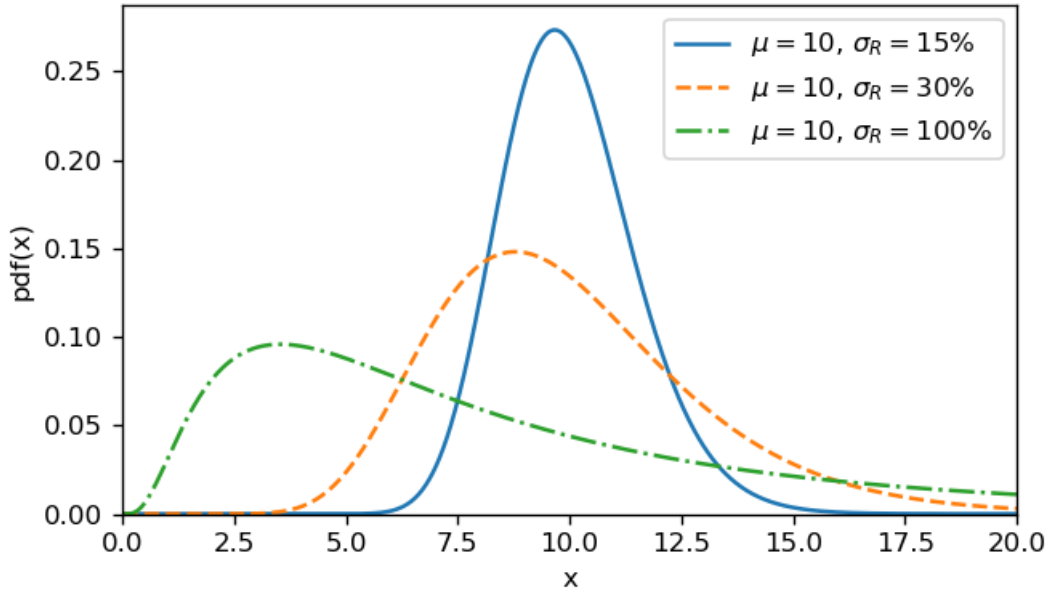


Fig. A.3: Probability distribution function of log-normal distribution with transformed parameters  $\mu$  and  $\sigma_R$ .

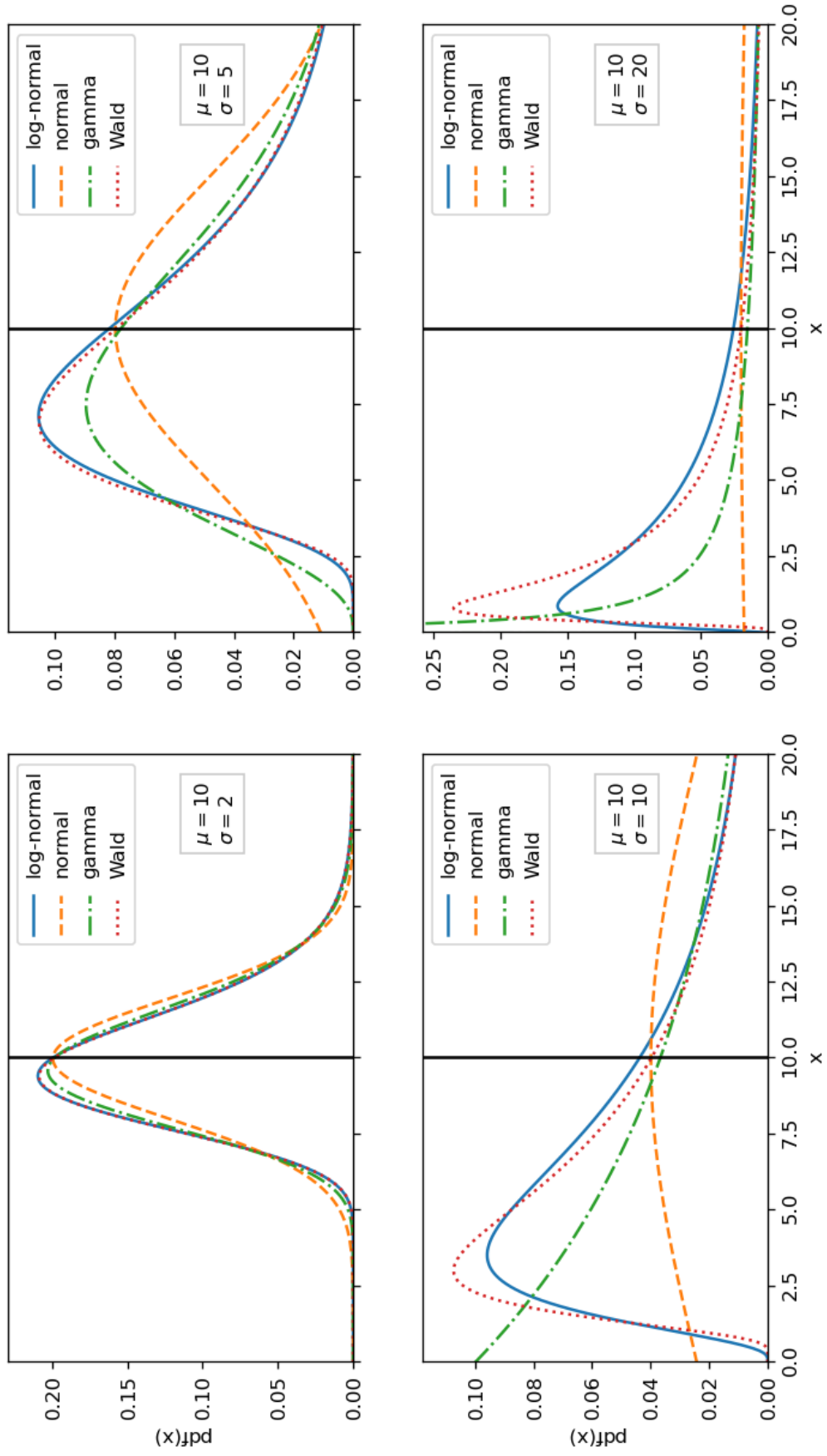


Fig. A.4: Comparison of probability distribution functions of selected distributions with positive domain and the normal distribution. These distributions were considered for the likelihood of the Bayesian models described in sections 4.1 and 5.1.