Dirk Joachim Nille, B.Sc. M.Sc.

Bayesian Inference of Heat Load Patterns in ASDEX Upgrade

DOCTORAL THESIS

to achieve the university degree of

Doktor der technischen Wissenschaften

submitted to

Graz University of Technology

Supervisor

Priv.-Doz. Dr.rer.nat. Udo von Toussaint

Institute of Theoretical and Computational Physics

In Cooperation with IPP Max Planck Institute for Plasma Physics

Graz, Dezember 2019

AFFIDAVIT

I declare that I have authored this thesis independently, that I have not used other than the declared sources/resources, and that I have explicitly indicated all material which has been quoted either literally or by content from the sources used. The text document uploaded to TUGRAZonline is identical to the present doctoral thesis.

Date

Signature

Version from 19th December 2019

Zusammenfassung

Das derzeit vielversprechendste Design für ein Kernfusionskraftwerk ist der Tokamak – im Wesentlichen ein ringförmiges Magnetfeld, das als Käfig für ein Plasma mit Temperaturen von bis zu 10 keV fungiert. In den letzten Jahrzehnten wurden bereits viele Probleme gelöst - angefangen von theoretischen Aspekten des kinetischen Verhaltens und der Wechselwirkungen über numerische Schemata, um Antworten zu erhalten, bei denen keine analytischen Ergebnisse erzielt werden können, bis hin zur Konstruktion der beteiligten Strukturen, elektrischen Leiter und Materialien. Wenn das Plasma jedoch die Quelle von Gigawatt an Leistung ist, muss es auch eine Leistungssenke geben - die umgebenden Wände. Ein großer Teil der Wand wird durch Strahlung homogen mit Wärmeleistung beaufschlagt, ein kleiner Teil kommt jedoch direkt mit dem Plasma in Kontakt. Die Wärmestromdichte liegt typischerweise bei mehreren $\mathrm{MW\,m^{-2}}$, und die Form und das Verhalten des Profils sind für die Kernfusionsforschung wesentlich. Zum einen droht die Wärmebelastung das Material zu beschädigen oder gar zu schmelzen. Zum anderen ist die Form der Verteilung ein Fußabdruck der Transportmechanismen im Plasma, die noch zu verstehen sind.

Die Ermittlung der Wärmelastungsprofile mit Unsicherheiten ist wichtig für physi-kalische Fragen sowie für Sicherheitsaspekte bei Materialien und Bauteilen, die hohen Wärmebelastungen ausgesetzt sind. Da für diesen Wärmestrom keine direkte Messmethode zur Verfügung steht, wird der indirekte Effekt der Wärmelasten auf die Temperatur der Teile gemessen werden. Am bekanntesten sind Infrarotkameras, die Informationen über die Oberflächentemperatur liefern, die Wärmelasten sehen. Da der Transport von Wärmeenergie in den festen Wandmaterialien diffus ist, handelt es sich um ein schlecht gestelltes Problem.

In dieser Arbeit werden die Methoden vorgestellt, die entwickelt wurden um dieses und ähnliche Probleme mit einem auf adaptiven Kerneln basierenden Multi-Resolution-Modell in einem Bayes'schen Rahmen zu lösen. Das Modell erlaubt es, die Komplexität des Modells zu quantifizieren, um eine Überanpassung zu verhindern - nach der Idee von Ockhams Rasiermesser. Die Daten stammen von ASDEX Upgrade, einem Tokamak in Garching bei München. Aufgrund von Symmetrieüberlegungen an dieser Maschine liegt der Fokus auf eindimensionalen Profilen im Raum. Die Analyse von Experimenten mit resonanten magnetischen Störung wird vorgestellt, da Interesse an der Position von schwachen Beiträgen besteht, und somit eine gute Anwendung für die adaptiven Kernel darstellt.

Für andere Maschinen und Probleme wird eine Erweiterung des AK-Modells auf 2D vorgestellt. Obwohl die Erweiterung prinzipiell unkompliziert ist, treten bei großen Systemen Probleme bezüglich der Laufzeit und des Speicherbedarfs des Algorithmus auf. Ein alternatives Modell wird vorgestellt, bei der die räumlich aufgelöste Regularisierung beibehalten wird. Der Name adaptives Diffusions Modell wird vorgeschlagen, da der Diffusionsoperator für die Regularisierung eines skalaren Feld verwendet wird.

Abstract

The currently most promising design for a nuclear fusion power plant is the tokamak – in essence a toroidal magnetic field, acting as a cage for a plasma at temperatures of up to 10 keV. During the last decades many problems have already been solved – ranging from the theory aspects of the kinetic behaviour and interactions, over numerical schemes to obtain answers where analytic results are unobtainable, to the engineering of the structures, electrical conductors, and materials involved. However, when the plasma is the source of gigawatts of power, there also has to be a sink – the surrounding walls. A large fraction of the wall will be homogeneously exposed to power radiated from the plasma volume, but a small part will be in direct contact with the plasma. The heat flux density is typically on the order of several $\mathrm{MW\,m^{-2}}$ and knowledge about the shape and behaviour of the profile is essential for nuclear fusion research. Firstly, the heat load poses the threat to damage or even to melt the surfaces. Secondly, this distribution is a footprint of the transport mechanisms in the plasma, which are still to be understood.

Obtaining reliable heat load profiles with uncertainties is important for physical questions, as well as for safety considerations for the materials and components exposed to high heat loads. As no direct measurement for this heat load distribution is available, the effect of the heat loads on the temperature distribution in the parts can be measured. Most prominent are infrared (IR) cameras, which yield information about the surface temperature. As the transport of thermal energy in the solid wall materials is diffusive, the problem at hand is ill-posed.

This thesis presents methods to tackle this and similar ill-posed problems with a multi-resolution model – based on adaptive kernel (AK) – in a Bayesian framework. The model allows to quantify the model complexity in order to prevent overfitting – following the idea of Ockham's razor. The data presented in this thesis were obtained at ASDEX Upgrade, a tokamak in Garching near Munich, Germany. Due to symmetry considerations at this machine, the focus is on one-dimensional profiles in space. Analysis done on experiments with resonant magnetic perturbation (RMP's) is presented, as the position of weak features in the heat load pattern is of interest and appears as ideal test for the adaptive kernel (AK) model.

For other machines and problems, an extension of the AK model to 2D is presented. While the extension is straight forward in principle, significant issues about the run time and memory demand of the algorithm are encountered for large systems. An alternative model is presented, keeping the spatially resolved regularisation. The name adaptive diffusion model is suggested, as the diffusion operator is used on a scalar field for regularisation.

Contents

Acronyms

AD adaptive diffusion – an alternative to the adaptive kernel approach for density estimation

AK adaptive kernel – a multi-resolution model for density estimation

BayTh Bayesian THEODOR – the Bayesian extension of the THEODOR code

HDE heat diffusion equation

IR infrared

log-det logarithm of the determinant

MAP maximum a posterior – position in parameter space with the highest probability density

MCMC Markov chain Monte Carlo

PDE partial differential equation

SNR signal-to-noise ratio

SOL scrape-off-layer – a domain in magnetic confinement devices

STE stochastic trace estimation

SVD singular value decomposition

1. Motivation and background

An essential ingredient for our modern society is energy, and the demand is expected to grow [Age17]. A potential method to obtain electric energy with little harm to our environment is nuclear power. However, current nuclear fission reactors are facing safety issues, like nuclear waste [JJ96; Hed97]. An alternative is nuclear fusion. Copying this process from the core of our sun [Kik11] has the potential to deliver electrical power with little CO_2 emissions [PA15; WK00]. A major issue for the development of a future fusion power plant is power exhaust [You+16]. Understanding and predicting the behaviour of the hot fluid called plasma [aut+05, p. 20] and optimising the materials in contact with it are essential for fusion based on magnetically confined plasma. This understanding is achieved by analysing data from existing experiments, which is typically an ill-posed inverse problem.

In this work the use of inference methods is presented, aiming to improve the analysis of infrared data, in order to determine the heat load pattern. This chapter presents an overview of the basics of fusion and known confinement methods. Parts of this are based on the authors master thesis [Nil16]. For magnetic confinement of plasma the limiter and divertor configuration are introduced. For the more relevant divertor configuration a model describing the heat flux profile onto the wall is outlined. Quantifying the underlying processes is important for the design of a fusion power plant.
The essentials about inference and Bayes' Theorem are explained in the next chapter.

1.1. Magnetic confinement fusion

This section is meant as a brief introduction to magnetically confined fusion, which is a confinement concept to gain reliable electrical power from nuclear, with global research activity [Fed+14]. A comparison to nuclear fission is followed by the introduction of the stellarator and tokamak concepts. The limiter and divertor configuration for the power exhaust are then followed by a simple model for the heat load pattern on a divertor target.

1.1.1. Fusion as a source of energy

Nuclear fusion is closely linked to nuclear fission, which is an established method to gain electric energy based on the mass defect between nucleons and composite atoms. In the case of fission, a heavy nucleus like uranium is split, yielding products of lower total mass than the initial elements, as quantified by E = mc^2. There is a set of possible products, which can be stable, or radioactive themselves. This is where nuclear fusion is inherently beneficial. For further reading see [Bit04, p. 18]. If two light nuclei fuse to a heavier one, the net mass is reduced as well and energy is released. Iron is the boundary between fusion and fission in terms of energy gain, as it is the nucleus with the highest binding energy per nucleon. The energy released per reaction is on the order of mega electron-volt (MeV), and therefore about 6 orders of magnitude larger than for chemical reactions, like burning fossil hydrocarbon.

Fusion of two isotopes of hydrogen, deuterium (D) and tritium (T), is currently favoured for the use in a fusion power plant [Bol+02; Che16]. This reaction has the highest cross-section at the lowest relative kinetic energy of 64 keV of all fusion reactions, as shown in figure 1.1. This is because ^{5}He – the resulting nucleon – has an excited state 64 keV above the combined mass equivalent of D and T. This isotope of helium is not stable and decays to ^{4}He, a neutron and energy:

$$\mathrm{D} + \mathrm{T} \rightarrow {}^5\,\mathrm{He} \rightarrow {}^4\,\mathrm{He} + \mathrm{n} + 17.6\,\mathrm{MeV} \tag{1.1}$$

The final product is stable ^{4}He (α-particle) and a neutron. They together inherit the released 17.6 MeV according to their masses. The helium is considered an alpha particle, as it is ionised, with 3.5 MeV of kinetic energy and the neutron carries about 14.1 MeV. The neutron is absorbed in the first wall, facing the fusion plasma. Accordingly the wall material can be chosen to minimise radioactive transmutation. Ideally the neutron is caught by lithium-6 or boron-10 contained in the wall to breed tritium [Rub19] and achieve tritium self-sufficiency..

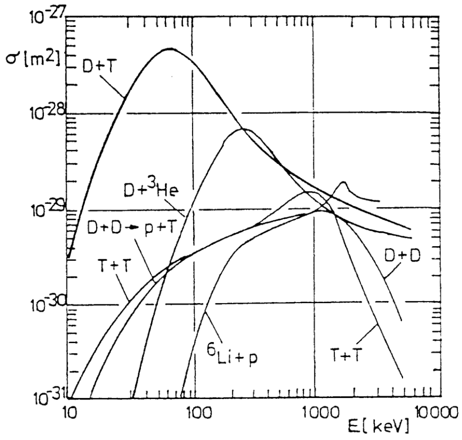


Figure 1.1.: Cross-section of fusion reaction for different nuclei combinations as log-log graph [Jen14]. Note the highest overall cross-section of the DT-reaction a comparably low energy of 64 keV.

Kinetic energies of several tens of keV are needed to overcome the repulsive electrostatic potential of the nucleons for the DT reaction. However, temperatures of about 10 keV – corresponding to about 100 million Kelvin – are sufficient to reach the maximum fusion rate.

The reason is the tunnel effect [Bit04, p. 410], allowing to overcome the barrier at energies below the potential barrier height, as well as the Maxwellian energy distribution [Bit04, p. 165] in the plasma. These temperatures are achievable using a magnetically confined plasma, where the fluid is not in contact with the solid wall material. Note that the necessary temperature corresponds to energies, which are several orders of magnitude above the $13.6\,\mathrm{eV}$ ionisation energy of hydrogen [aut+05]. Consequently the electrons are split from the nuclei which allows the manipulation of the fluid by electric and magnetic fields. This so called fourth state of matter – beside solid, liquid and gaseous – is called plasma. The kinetic behaviour is more complex than that of a neutral gas, as the charged particles are interacting with the external electric and magnetic field, due to their movement and spatial distribution.

1.1.2. Confining a fusion plasma

As was discussed before, high temperatures are mandatory to obtain fusion, prohibiting direct contact to solid matter used as containment. There are three ways to confine particles such hot: by gravity, as it is done in stars; by temporary use of inertia, where the fusion fuel is compressed quickly, and by magnetic confinement. Confinement by gravity requires masses only available at stars and comparable celestial bodies. Inertial fusion [Zoh17] is mainly researched for military purposes and suffers from micro-instabilities. An example is the National Ignition Facility (NIF), in Livermore California, operated by University of California [1]. Magnetic confinement is a promising way to achieve fusion in a future power plant.

Charged particles follow magnetic field lines, due to the Lorentz force

$$\vec{F}_{\mathrm{L}} = q\vec{v} \times \vec{B} \tag{1.2}$$

acting on a charge q with velocity $\vec{v}$ in the magnetic field $\vec{B}$. A linear magnetic field – like in an infinite coil – is able to confine charged particles, but practical considerations demand it to be of limited length. A local increase of the magnetic field magnitude along the main axis creates an area where particles are reflected due to the mirror force, see [Bit04, p. 77] for more details. However, the degree of reflection scales only with the logarithm of the ratio of the magnetic field at the bottleneck to the unperturbed magnetic field. Therefore, magnetic mirror machines are not suited to achieve in an efficient manner the high densities and temperatures needed for fusion.

An alternative is to bend the magnetic field lines to a circular shape. A purely toroidal magnetic field does not suffer from end-losses, but the curvature of the field affects the motion of the charged particles. This leads to positive and negative charges being separated along

---

[1] NIF Homepage: `https://lasers.llnl.gov/science/journal-articles`

the symmetry axis of the field due to the curvature drift

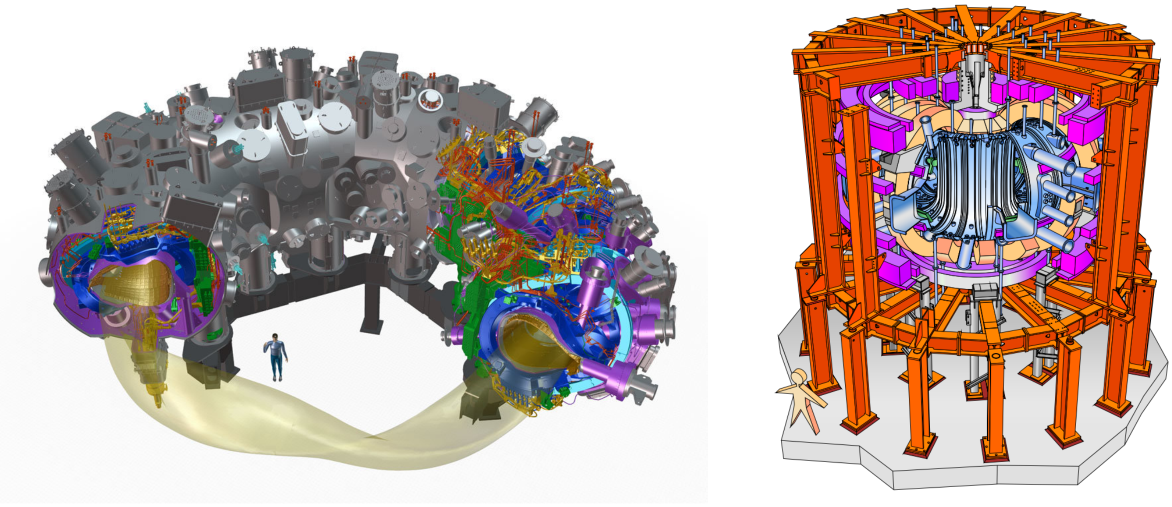
$$\vec{v}_{\text{curv}} = m v_{\parallel}^2 \frac{\vec{B} \times \nabla \vec{B}}{q B^3}$$

which depends on the sign of the charge. This charge separation leads to electric fields, which, together with the present magnetic field, drives the charges towards lower B-field density areas by the $E \times B$ drift

$$\vec{v}_{\vec{E}\times\vec{B}} = \frac{\vec{E} \times \vec{B}}{B^2} \tag{1.3}$$

which is oriented radially outward. Twisting the field lines helically around the torus is one way to minimize the losses.

Two different approaches to achieve this twist are the tokamak, relying on a toroidal current – typically on the order of MA – and the stellarator – relying on a clever configuration of the external magnetic coils. Figure 1.2 shows sketches of both concepts. The stellarator Wendelstein 7-X (W7-X) located at the IPP in Greifswald, Germany is illustrated in figure 1.2(a). A drawing of the tokamak ASDEX Upgrade (AUG) is shown in figure 1.2(b).



(a) Illustration of W7X from `https://www.ipp.mpg.de/2815279/technologie` – accessed 02.09.2019.

(b) Cut away drawing of ASDEX Upgrade from `www.aug.ipp.mpg.de` accessed at 02.09.19.

Figure 1.2.: Sketches of the two fusion experiments operated by IPP in Germany.

For a tokamak there are various external coils needed, as explained with reference to figure 1.3. Basic heating and a drive for the toroidal current are available from the central solenoid, indicated in green. This transformer also limits the pulse duration of classic tokamaks, as a toroidal electric field is the result of a changing current in the solenoid. A voltage on the order of $(1 - 10)\,\mathrm{V}$ is sufficient to drive currents on the order of $1\,\mathrm{MA}$, due to the low resistance of the plasma of about $1 \cdot 10^{-5}\,\Omega$ per toroidal turn. The toroidal magnetic field is provided

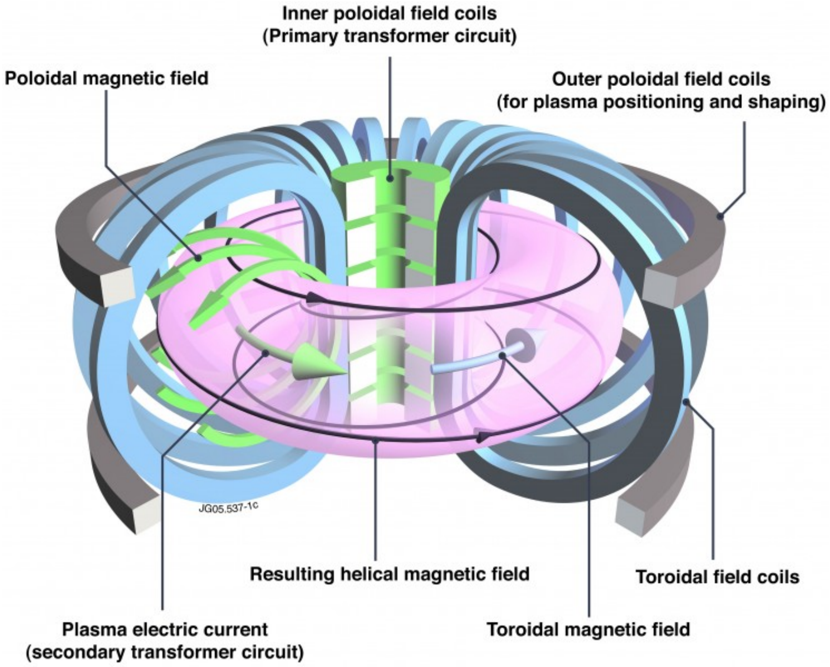


Figure 1.3.: Basic structure of a tokamak `www.euro-fusion.org` accessed at 03.06.19.

by the toroidal field coils, enclosing the plasma in a D-shape. To form the X-point for the divertor – explained in the next section – poloidal field coils are used. In a tokamak the toroidal current of the plasma itself induces a poloidal magnetic field, twisting the otherwise purely toroidal magnetic field. This twist connects the top with the bottom of the plasma and allows thereby to compensate accumulating charge from the $E \times B$-drift. In a stellarator the magnetic transform is provided by non-toroidal symmetric coil systems, like in W7-X. This can be a combination of planar toroidal coils plus helical coils, but modern machines aim at a modular design.

For a fusion power plant, besides good confinement, it is important to extract the ash – helium – and impurities from the confined region. Both lead to radiative losses, reducing the core temperature, and diluting the fusion fuel. This is the subject of the next section and applies for tokamaks as well as stellarators.

1.1.3. Limiter and divertor configuration

The most widespread configuration up to the 1980's was the limiter configuration. The left sketch in figure 1.4 shows the cross-section of ASDEX Upgrade with the magnetic field lines as used in the limiter configuration. In this poloidal cross-section the x-axis corresponds to the major radius and the y-axis to the symmetry axis of the machine, the toroidal position is fixed. The circular structure represents field lines, which are projected into the poloidal plane and actually revolve the magnetic axis like a helix. The confined area of the plasma with closed field lines – meaning field lines not intersecting wall elements – is indicated in red. The field line separating the closed field lines from the open ones – intersecting a wall – is called the last closed flux surface. In the limiter configuration plasma is transported radially into the scrape-off-layer (SOL) – marked in orange – with the material wall interaction being in direct vicinity to the core plasma. To separate these areas the divertor configuration is used in most modern tokamaks. This is illustrated in the right side of figure 1.4. A magnetic X-point is formed, where the poloidal field is zero and the magnetic field is purely toroidal like on the magnetic axis. The field line where this separations happens is called the separatrix. The SOL is *diverted* away from the confined area into a volume dedicated for exhaust. This includes a pumping system to prevent the neutralised gas to penetrate into the confined area. The green marked area is called private flux region and is characterised by magnetic field lines which are separated from the SOL as well as the confined area.

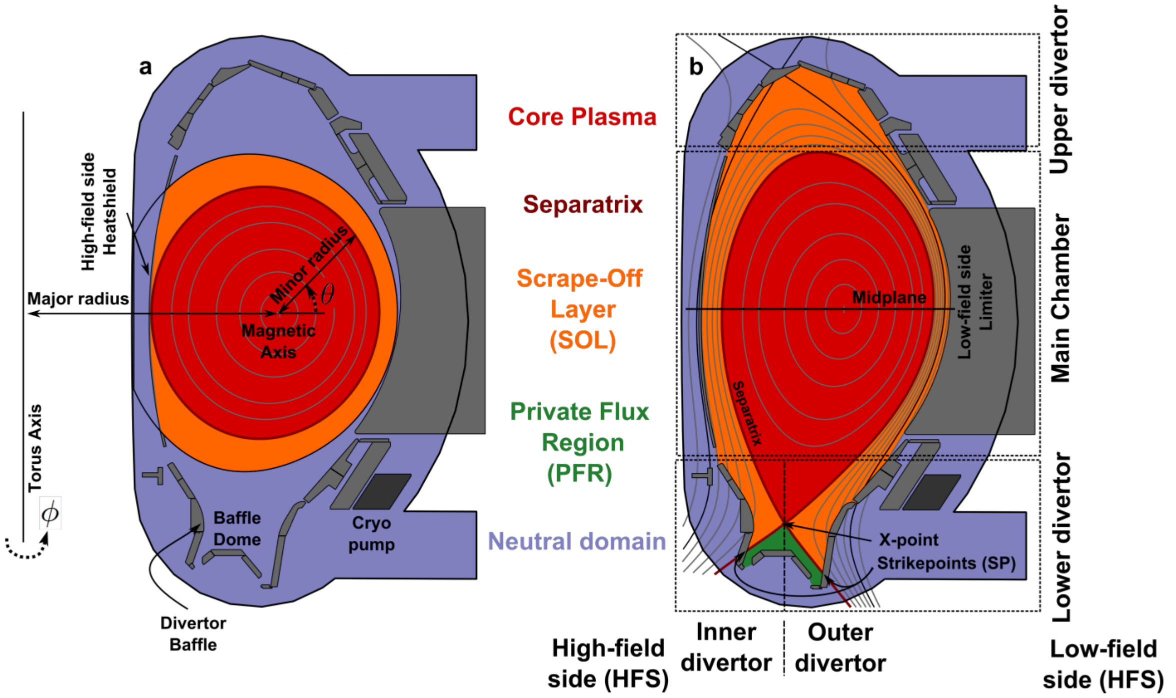


Figure 1.4.: Poloidal cross-section for a) limiter geometry and b) divertor geometry [Rei14]. The midplane (mp) is defined by the magnetic axis in the poloidal cross-section for both configurations.

Figure 1.4 b) shows a lower single null configuration in ASDEX Upgrade, where only the lower X-point is *active*. The upper X-point is located behind the target plates, so the strike points – where the highest heat flux arrives by plasma transport – form only at the bottom of the machine.

The spatial distance between the confined plasma and the target surface increases the travel time of the particles, which have passed outside of the last closed field line. During this time energy is lost by dissipation like volume radiation and charge-exchange [Sta00, p. 219]. At a sufficiently low plasma temperature in front of the target, a cushion of neutral particles can form, slowing down the incoming charged particles [Sta00, ch. 16.4]. In addition to the lowered temperature in front of the target, the impurities produced by sputtering of the target are less likely to reach the core plasma. This will be vital for large fusion devices such as ITER, which is envisaged to produce ten times the fusion power by DT-reactions compared to the auxiliary heating power. The heat flux density parallel to the magnetic field for ITER at the outer midplane can be approximated by the heat crossing the separatrix $P_{\mathrm{sep}} \simeq 100\,\mathrm{MW}$ divided by the divertor area given by the major radius $R = 6.2\,\mathrm{m}$ and the extent of the SOL at the outer midplane of $\lambda_q \simeq 1\,\mathrm{mm}$ [Eic+13]:

$$q_{\mathrm{ITER,div}} = \frac{P_{\mathrm{sep}}}{2\pi R \lambda_q} \cdot \frac{\sin(\alpha_{\mathrm{pitch,div}})}{\sin(\alpha_{\mathrm{pitch,mp}})} \simeq 770\,\mathrm{MW\,m^{-2}}\,. \tag{1.4}$$

The angles $\alpha_{\mathrm{pitch,mp}} \simeq 10°$ and $\alpha_{\mathrm{pitch,div}} \simeq 3°$ represent the inclination of the magnetic field lines at the midplane (mp) and at the divertor (div).

Radiative losses and spreading of the profile have to reduce the peak heat flux density to around $10\,\mathrm{MW\,m^{-2}}$ perpendicular to the divertor target, which can be handled by actively cooled components [Mao+18].

Figure 1.5 shows two different configurations for the divertor geometry in ASDEX Upgrade, called *open divertor* (Div I) and *closed divertor* (Div IIa). While the open divertor delivers target surfaces distant to the confined region, the closed divertor in addition supports a higher pressure of neutral particles. This is important for large machines, as the pumping flow is proportional to the particle density [Rei14].

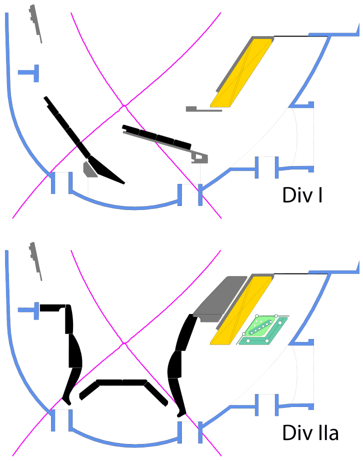


Figure 1.5.: Comparison between two divertor configurations in ASDEX Upgrade, from `www.aug.ipp.mpg.de` accessed at 02.09.19. Upper part: *open divertor configuration* of divertor *Div I*. Lower part: *closed divertor configuration* of divertor *Div IIa*.

1.1.4. Heat flux density profile

Heat reaching the target plates in the divertor is examined in poloidal profiles, perpendicular to the toroidal orientation, on their surface. They are readily referenced as heat flux profiles, though the quantity shown is a heat flux density – power per surface area. For convenience these labels are kept interchangeable. The profile peak value and gradient influence the lifetime of the material. Fluxes of heat and particles degrade the material, spatially and temporally varying heat loads lead to cracking [LY15]. Predicting and controlling these profiles is necessary in order to design and operate a fusion power plant. Steady state heat loads of several tens and intermittent events with hundreds of $\mathrm{MW\,m^{-2}}$ have been measured in present day devices, usually operating for some seconds [Sie+13]. However, the experimental reactor

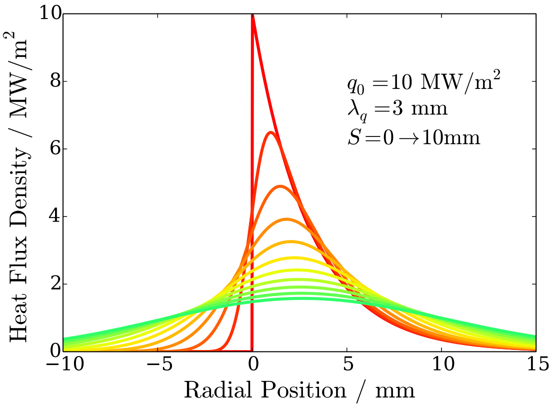


Figure 1.6.: Spreading of an exponentially decaying heat flux profile at the X-point to the target by diffusion for different spreading values S. From red to green S increases from 0 mm to 10 mm in steps of 1 mm.

ITER is only specified for values of $10\,\mathrm{MW\,m^{-2}}$ [Mao+18] in steady state. A good description of the heat flux profile at the X-point is an exponential with fall-off length λ_q and peak q_0, represented in figure 1.6 as a blue line. Spreading the heat lowers the peak heat flux and the gradients. Better understanding of this spreading process is important for the design of future divertors and yields information about the heat transport in the SOL.
More details on the formation of this profile are given in section 2.2.

1.1.5. External magnetic perturbation

As the toroidal field is the result of a finite set of toroidal field coils, the resulting field is not perfectly toroidal. Instead, a ripple [aut+05, p. 208] of the field influences the trajectories of the particles. Imperfections like this can be counteracted by external field coils, which apply a correction field. However, these coils can be used to introduce an external magnetic perturbation as well. For a given set of coils the choice of their individual polarity and current allows to impose various modes. A perturbation is called resonant when the perturbation fields of different coils are coherent along a single magnetic field line in the edge. The effect on the plasma is especially strong when there is locking to internal modes [Wil+16] or density pump-out [Gar+13].

Typically only the outermost part of the plasma is affected the most, because the conducting plasma shields the external perturbation before the core of the plasma is affected [Wil+16; RC09]. Due to the perturbation in the SOL, the heat load profiles change in presence of external magnetic perturbation [Sie+17].

One of the occurring features of the external perturbation is strike line splitting, which will be presented in the result chapter 8.

1.2. Scope and structure of this thesis

This thesis addresses a set of questions. Foremost: how to deduce heat load profiles from IR data? This is a time-dependent inverse problem. How reliable are the profiles and how much do they differ from a deterministic approach, circumventing to solve the inverse problem? To deal with the dynamic signal to noise ratio and varying profile complexity, a multi-resolution model is used.
The heat content of events, which are shorter than the interval between two samples of the diagnostic, is also of interest, as these are potentially dangerous for the material.
Important topics are plasma physics, data recording, numeric integration, and inference. The following structure overview is meant as a guide for the reader:
Chapter 2 introduces the basics of the theory of Bayesian inference and the probabilistic analysis used. Basics like the concept of prior knowledge and Ockham's Razor are introduced in section 2.7. Section 2.6 defines the general idea behind the solution of inverse problems.

Chapter 3 introduces the infrared diagnostic used to obtain the data presented and analysed in this thesis. This includes an example of the signal-to-noise ratio during two typical experiments.

Chapter 4 introduces the numerical method to represent the physical system and explains how the heat diffusion equation is solved. Considerations about the uncertainty of the solving schemes are given.

Chapter 5 introduces the methods and tools to navigate probability space and how the final solution is determined. In the end, the knowledge about the distributions has to be expressed in scalar quantities, which are used to make decisions.

Chapter 6 introduces the models used to describe the quantity of interest. Section 6.1.2 introduces the definition of and inference methods for the adaptive kernel model. This is followed by section 6.1.5 with an alternative formulation for density estimation: adaptive diffusion.

Chapter 7 presents the technical approach used to connect the single pieces to obtain reliable results. This might be a good starting point for readers which are familiar with the basics of inference, as the other relevant sections are references.

After introducing the physical problem, and the methods and approaches used to solve them, the application is presented in the three chapters 8, 9 and 10. The results are separated into 1D heat loads presented in chapter 8. The fast events – as well in 1D but with a more detailed model about the heat load in time – are presented in chapter 9. Finally an overview about 2D results is given in chapter 10.

The results are summarised and concluded in chapter 11, followed by the outlook.

2. Theory

This chapter begins with an introduction to heat transport, in order to answer the question *why* this work was carried out. Of central importance is the transport of heat via diffusion in a solid (2.1). This is proceeded by transport via diffusion and convection in a plasma. The heat load pattern on the divertor can be quantified via plasma parameters – section 2.2. Essential for the diagnostic are Planck's law (2.3) and bremsstrahlung (2.4). This first part is in parts based on [Nil16].

The second part is discussing the *how*: obtaining information from measurements is best done using the language of statistics (2.5). Inferring a quantity of interest is a typical inverse problem (2.6), which is approached via Bayes' theorem (2.7). This way prior information (2.7.2) is combined with the measurement via the likelihood (2.7.3). To prevent over-interpretation of the data – also known as over-fitting – the principle of Ockham's Razor should be applied (2.7.4), which is quantified in Bayes' theorem.

2.1. Heat transport

This section outlines the three basic mechanisms of transferring heat. While radiation is carried by electromagnetic processes, diffusive and convective transport result from particle interaction and movement. Diffusion is able to transfer heat in solids with quasi-static atoms, whereas convection describes moving particles. All transport mechanisms are important in plasma, but diffusion is essential for the forward model of the inference – see chapter 4.

Convection describes heat flux carried by particle motion [Jij09]. This movement can build up temperature gradients and can be in equilibrium with diffusive processes. It is dominant in the convection limited regime in the SOL, when a net particle flow carries heat without the presence of a parallel temperature gradient [Sta00, p. 95].

Radiation is emitted by several mechanisms, with thermal radiation being the most important in this thesis, as it is the basis for the measurements [Rie01]. Objects with finite temperature emit *thermal radiation* according to Planck's law, further explained in section 3. Thermal radiation is used to determine the temperature of surfaces, for example the divertor plates.

Other kinds of radiation in a magnetised plasma are line radiation, bremsstrahlung and cyclotron radiation [Hut02, ch. 5]. Bremsstrahlung is dominant in the confined plasma, acting as heat sink. Line radiation is emitted by excited atoms and is dominant in the divertor region. Cyclotron radiation is emitted and absorbed by gyrating, charged particles. These mechanisms are out of the scope of this thesis and therefore are not further discussed.

Diffusion is found in all phases of matter, where heat is transported by particle collisions, reducing temperature gradients [BS98]. This process is explained in more detail in the rest of this section. It is an important transport mechanism in the plasma in the scrape-off-layer (SOL) and as well in the solid divertor target.

For the scope of this work diffusion is most important – for the transport in the plasma as well as in the forward model – and hence an introduction is given.

Diffusion in general is based on particle interaction on the scale of the average particle distance, leading to a transport of particles or energy in a quasi-static medium. Gradients of quantities like particle density or concentration, or temperature are spatially equalised by diffusive processes. Particle transport is used to introduce basic properties, which are then applied to heat diffusion.

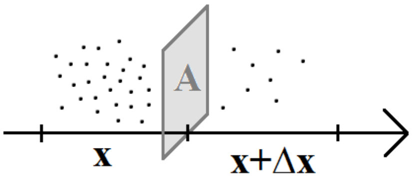


Figure 2.1.: Sketch for particle diffusion used for Fick's law. In this example the result is a net flux to the right, equalising the particle density.

The flux of particles across an area A, see figure 2.1, can be expressed by the flux of particles from the left to the right minus the flux from the right to the left. Assumption is an unbiased random walk, where particles have the same probability of moving to the left or to the right, with step sizes Δt and step length Δx. The length with extent Δx along the x axis to the left of A is denoted x and the length to the right $x + \Delta x$. In the associated volumes $A \cdot \Delta x$ the particle numbers are denoted $N(x)$ and $N(x + \Delta x)$. The flux through A is found to be

$$J(x,t) = -\frac{1}{2}\left(\frac{N(x+\Delta x,t)}{A\Delta t} - \frac{N(x,t)}{A\Delta t}\right) \tag{2.1}$$

Expanding by the square of the step width $(\Delta x)^2$ and introducing the particle density

$$n(x,t) = \frac{N(x,t)}{V} = \frac{N(x,t)}{\Delta x \cdot A} \tag{2.2}$$

leads to:

$$J(x,t) = -\frac{(\Delta x)^2}{2\Delta t}\left(\frac{n(x+\Delta x,t) - n(x,t)}{\Delta x}\right) \tag{2.3}$$

For infinitesimal step width we find the derivative of n with respect to x on the right hand side and define the diffusivity

$$D = \frac{(\Delta x)^2}{2\Delta t} \tag{2.4}$$

yielding

$$J(x,t) = -D\frac{\partial n(x,t)}{\partial x} \tag{2.5}$$

known as Fick's first law [BS98, p. 3]. It can be generalised to more than one dimension by using the corresponding differential operator:

$$\vec{J}(x,t) = -D\,\nabla n(x,t) \tag{2.6}$$

The diffusivity D in general is a tensor, but for homogeneous materials like tungsten it is expressed as scalar. This derivation is also valid for other quantities like heat instead of particle density. The corresponding solution for the heat flux density is called Fourier's law

$$\vec{q} = -\kappa \nabla T \tag{2.7}$$

with heat conductivity κ being related to the diffusivity by the mass density ρ and the specific heat capacity c_p:

$$D = \frac{\kappa}{\rho c_p} \, . \tag{2.8}$$

Applying the continuity equation for the thermal energy to Fourier's law

$$\nabla \cdot \vec{q} \, + \, \rho c_p \frac{\partial T}{\partial t} = 0 \tag{2.9}$$

leads to the heat diffusion equation [BS98, p. 4]:

$$\frac{\partial T}{\partial t} = \frac{1}{\rho c_p} \nabla \cdot (\kappa(T) \nabla T) \tag{2.10}$$

This second order partial differential equation (PDE) exhibits a non-linearity if κ is a function of T.

This equation is simplified by introducing the heat potential

$$u(T) \equiv \int_0^T \kappa(T')dT' \tag{2.11}$$

showing the properties

$$\frac{\partial u}{\partial t} = \frac{\partial u}{\partial T}\frac{\partial T}{\partial t} = \kappa \frac{\partial T}{\partial t} \tag{2.12}$$

$$\frac{\partial u}{\partial x} = \frac{\partial u}{\partial T}\frac{\partial T}{\partial x} = \kappa \frac{\partial T}{\partial x} \tag{2.13}$$

$$\frac{\partial^2 u}{\partial^2 x} = \frac{\partial}{\partial x}\left(\kappa \frac{\partial T}{\partial x}\right) \tag{2.14}$$

leading to

$$\vec{q} = -\nabla u \tag{2.15}$$

and

$$\frac{\partial u}{\partial t} = D(u)\Delta u \ . \tag{2.16}$$

Substituting the temperature with the heat potential, we end up with a semi-linear PDE. This is beneficial for numerical solving schemes, allowing to use simpler and hence faster algorithms and reduces numerical errors. Equation (2.16) is the basis for the code called THEODOR (*THermal Energy Onto DivertOR*) [Her+95]. Its purpose is to calculate the heat flux onto a surface given a measured surface temperature profile and models for the other – typically unobserved – boundaries.

2.2. Divertor target heat load profiles

This section is based on the publication [Nil+19] of the thesis author.

To describe the heat flux density profile on the divertor target, a model for the heat transport in the SOL assuming only diffusive parallel and perpendicular electron conduction is commonly used [Eic+13]. All plasma temperatures and densities in this section refer to the electrons, being the dominant species for parallel diffusive transport for comparable ion and electron temperatures, as seen in the Braginskii equations, see [Sta00, ch. 9]. For diffusive transport parallel to the field lines in the divertor volume the transport time

$$\tau_{\|} = \frac{L^2}{\chi_{\|}} \qquad [\mathrm{s}] \tag{2.17}$$

is an expression of connection length L – from the divertor entrance to the target – and parallel diffusivity $\chi_{\|}$. The parallel diffusion time is equal to the perpendicular diffusion time for heat entering the divertor region. The perpendicular diffusion length is thus given by

$$S = \sqrt{\tau_\| \cdot \chi_\perp} = L\sqrt{\frac{\chi_\perp}{\chi_\|}} \qquad [\mathrm{m}] \quad (2.18)$$

and is also called divertor broadening. Measurements of the heat flux profiles are readily done by infrared thermography in target coordinates called s with separatrix position s_0. Quantities following the magnetic field lines can be mapped to the outer midplane, to the radial coordinate called x, for comparison between different magnetic geometries and machines. The coordinates are related by the effective flux expansion $f_x = f_{x,\mathrm{magn}} \cdot f_{x,\mathrm{geom}}$, which is the product of the magnetic flux expansion $f_{x,\mathrm{mag}}$ and poloidal inclination of the tile with respect to the field lines. The position mapped to the outer midplane is then

$$x = \frac{s - s_0}{f_x} \qquad (2.19)$$

with $x = 0$ representing the separatrix position. For perpendicular transport being described as 1D diffusion, a power density profile given by a delta peak entering the divertor area is spread to a Gaussian of width S when reaching the target without flux expansion. The measure on the target is $S_{\mathrm{tar}} = S \cdot f_x$. In this work S refers to the divertor broadening mapped to the outer midplane.

Note that the poloidal inclination between field lines and target is not included in this discussion, as it is only a mapping from parallel heat flux along the plasma onto what is seen as perpendicular heat flux of the target material. For the sheath condition this angle, which is machine and configuration dependent, has to be included [Sta00, ch. 1.5].

The X-point heat flux density profile is described [Mak+12] by an exponential with peak value q_0 at the separatrix and decay length λ_q at the midplane with the radial coordinate x [Eic+13, eq. 1]:

$$q(x) = q_0 \cdot \exp\left(-\frac{x}{\lambda_q}\right) : x > 0 \ . \qquad [\mathrm{MW\,m^{-2}}] \quad (2.20)$$

Following the simplified model for perpendicular diffusion, the target heat flux profile is described by the X-point profile convolved with a Gaussian of width $\sigma = \sqrt{2} \cdot S$., representing the broadening in the divertor region:

$$q_\|(s) \ = \frac{q_0}{2} \exp\left(\left(\frac{S}{2\lambda_q}\right)^2 - \left(\frac{s - s_0}{f_x \lambda_q}\right)\right) \cdot \mathrm{erfc}\left(\frac{S}{2\lambda_q} - \frac{s - s_0}{f_x S}\right) \ . \qquad (2.21)$$

Figure 2.2 shows the flattening of the heat flux density profile from the raw exponential in deep red – starting at the strike point at $s_0 = 0$ – up to a value of $S = 10\,\mathrm{mm}$ in green in steps of 1 mm for S, keeping $\lambda_q = 3\,\mathrm{mm}$ and $q_0 = 10\,\mathrm{MW\,m^{-2}}$ fixed.

The peak heat flux $\hat{q}$ onto the target is used as design parameter and related to the integrated

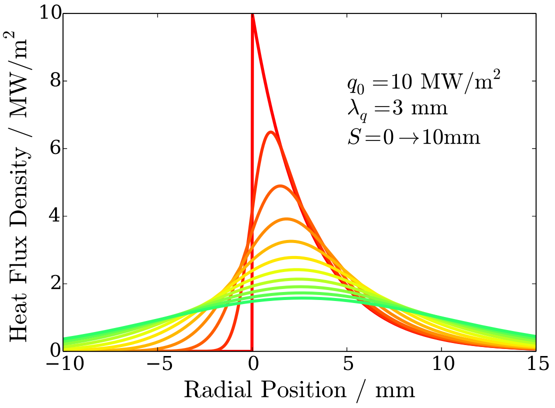


Figure 2.2.: Influence of divertor broadening S: Target heat flux profiles with steps of 1 mm in the the divertor broadening S, starting from the unperturbed X-point distribution in red up to $S = 10\,\text{mm}$ in green.

power profile of arbitrary shape $q(s)$ onto the target element by the integral power decay length [Mak+12, eq. 3]

$$\lambda_{\text{int}} \equiv \int \frac{q(s)}{\hat{q}} \text{d}s \,. \qquad [\text{m}] \qquad (2.22)$$

The benefit of an additional divertor broadening S on λ_{int} compared to an exponential with decay length λ_q – as described in equation (2.21) – is approximated [Mak+12] by

$$\lambda_{\text{int}} \simeq \lambda_q + 1.64 \cdot S \,. \qquad (2.23)$$

Studies predict a value for the heat decay length of about $\lambda_q \approx 1\,\text{mm}$ for future fusion relevant machines like ITER, which is smaller than for current machines like AUG and JET [Eic+13]. Therefore the divertor broadening gains importance to meet the material limits of the divertor target with respect to the incoming heat flux density. Scaling laws for S are available for AUG [Sie+13; Sie+16], investigating S for target electron temperatures above 20 eV. Below this value, the increasing volume radiation prevents the inference of S from the IR signal in AUG with the current system and method. An analysis of the effect of the plasma transport on the divertor spreading is found in [Nil+19].

2.3. Planck's law

Planck's law of radiation [Pla00a; Pla00b] describes the emission of electromagnetic radiation from an ideal black body with finite temperature. The spectral radiance M_e emitted by a black body of temperature T at wavelength λ and surface area $\mathrm{d}A$ is described by

$$M_e(T,\lambda)\,\mathrm{d}A\,\mathrm{d}\lambda = \frac{2\pi hc^2}{\lambda^5}\frac{1}{\exp\left(\dfrac{hc}{\lambda kT}\right)-1}\,\mathrm{d}A\,\mathrm{d}\lambda \qquad \left[\frac{\mathrm{W}}{\mathrm{m}^2\,\mathrm{\mu m}\,\mathrm{sr}}\right] \tag{2.24}$$

where k denotes the Boltzmann constant, h the Planck constant and c the speed of light.

The cameras used in the experiments measure the photon flux M_γ instead of the spectral radiance M_e. The rate of photons is expressed by taking the energy per photon at a certain wavelength

$$E_\gamma = \frac{hc}{\lambda} \tag{2.25}$$

into account:

$$M_\gamma(T,\lambda)\,\mathrm{d}A\,\mathrm{d}\lambda = \frac{2\pi c}{\lambda^4}\frac{1}{\exp\left(\frac{hc}{\lambda kT}\right)-1}\,\mathrm{d}A\,\mathrm{d}\lambda \qquad \left[\frac{1}{\mathrm{m}^2\,\mathrm{\mu m}\,\mathrm{sr}\,\mathrm{s}}\right] \tag{2.26}$$

Real surfaces are not ideal black bodies, but a part of the incident radiation is reflected. The grey body is a first approximation to real surfaces. From the incoming radiation intensity I_0 the reflected intensity I_R is given by the wavelength independent reflectivity

$$R = \frac{I_R}{I_0}\,. \tag{2.27}$$

A grey body emits Planck radiation only with the emissivity defined as:

$$\epsilon = 1 - R\,. \tag{2.28}$$

The radiated fluxes for a grey body are $\epsilon 1$ times the black body radiation:

$$M_{\gamma,gb}(T,\lambda)\,\mathrm{d}A\,\mathrm{d}\lambda = \epsilon\cdot\frac{2\pi c}{\lambda^4}\frac{1}{\exp\left(\frac{hc}{\lambda kT}\right)-1}\,\mathrm{d}A\,\mathrm{d}\lambda \qquad \left[\frac{1}{\mathrm{m}^2\,\mathrm{\mu m}\,\mathrm{sr}\,\mathrm{s}}\right] \tag{2.29}$$

The photon flux emitted by a surface into the half space is calculated by computing the integral

$$\Gamma(T) = \int\!\!\int \epsilon\frac{2\pi c}{\lambda^4}\frac{1}{\exp\left(\frac{hc}{\lambda kT}\right)-1}\,\mathrm{d}A\,\mathrm{d}\lambda \tag{2.30}$$

which can be approximated for a narrow wavelength window by the effective wavelength λ_{eff}. The calibration coefficient c_0 depends on the properties of the optical transmission line, like aperture and losses, and the properties of the camera like photon efficiency and pixel area. This yields

$$\Gamma(T) \simeq c_0\frac{2\pi c\epsilon}{\lambda_{\text{eff}}^4}\frac{1}{\exp\left(\frac{hc}{\lambda_{\text{eff}}kT}\right)-1} \tag{2.31}$$

Thermography uses Planck's law to determine surface temperatures. Implicitly also the width of the spectrum is included. Solving this equation for the temperature yields:

$$T(\Gamma) = \frac{hc}{\lambda_{\text{eff}} k} \frac{1}{\ln\left(\frac{2\pi c\epsilon}{\lambda_{\text{eff}}^4} \cdot \frac{c_0}{\Gamma} + 1\right)} \tag{2.32}$$

2.4. Bremsstrahlung

Electric charges subject to a change in their velocity emit radiation known as bremsstrahlung. This radiation is found in ionised plasma and is mostly caused by electrons, which are deflected by ions. For a thermalised plasma – with Maxwellian velocity distribution of the electrons – the power spectrum is found to be [Gri97]

$$P_\lambda \mathrm{d}\lambda = \frac{2\pi c 8\sqrt{2}}{\lambda^2 3\sqrt{\pi}} \left[1 - \left(\frac{n_e e^2 \lambda^2}{4\pi^2 c^2 \epsilon_0 m_e}\right)\right]^{1/2} \cdot n_e^2 r_e^3 \left[\frac{(m_e c^2)^{3/2}}{(k_b T_e)^{1/2}}\right] \cdot E_1\left(\frac{1}{2}\left(\frac{hc}{k_b \lambda T_e}\right)^2\right) \cdot d\lambda \tag{2.33}$$

in units of $\mathrm{W\,m^{-3}\,m}$. Relevant quantities are n_e the volume density of electrons, r_e the classical radius of an electron, e the elementary charge, m_e the electron mass, k_B the Boltzmann constant and c the speed of light. E_1 is a function known as first exponential integral and defined as [Hut02, eq. 6.3.4]

$$E_1(y) = \int_y^\infty \mathrm{d}s \frac{\exp(-s)}{s} \; . \tag{2.34}$$

2.5. Probability

Note: While the sections before were treating the physical aspects like heat diffusion and Planck's law, this and the following sections will focus on how to formulate and treat the emerging inverse problem mathematically.

Bayes' theorem is a rule about updating probability distributions, which has a quite long history in science. Blaise Pascal and Pierre de Fermat gave first quantitative definitions in the 17th century. The classic interpretation of probability is derived from frequencies or relative occurrences of events. When a coin is tossed N times and the outcome head is found k times the probability of this event is k/N for $N \to \infty$. This has the obvious drawback, that the *actual* probability value is only available after an infinite number of trials, while it is not clear if more than one realisations is possible.

A more Bayesian understanding is that *true* and *false* are labels applying to propositions like 'It will rain tomorrow'. The probability is a measure for the truth of a proposition, which is a real number between 1 (guarantee or 100% true) and 0 (impossible or 0% true $\equiv$ false).

From symmetry considerations a set of rules is found, which probabilities must obey [SK19]. For a proposition of hypothesis A the associated probability is denoted $P(A)$ and its complement is denoted $P(\overline{A})$. For A and a second proposition B the rules are:

1. normalisation: $P(A) + P(\overline{A}) = 1$
2. sum rule: $P(A \vee B) = P(A) + P(B) - P(A \wedge B)$
3. product rule: $P(A, B) = P(A \mid B)\, P(B)$
4. marginalisation rule: $P(A) = P(A, B) + P(A, \overline{B})$

The probability of a hypothesis H is typically denoted

$$P(H \mid I) \tag{2.35}$$

and depends on background information I. While this background information can be obvious like 'My coin has two sides.' every aspect of relevance should be declared explicitly. The detail 'My coin never lands on the edge.' is often included implicitly, but does not always hold true in practice. A good rule of thumb is

There is no such thing as an unconditional probability.

When two scientists find different answers to the same question given the same data, their background information is probably different. This is easily revealed when they present them.

The concept of probabilities measuring the truth of a proposition applies to discrete as well as continuous problems. It applies to cases in court as well as to the mass of celestial bodies. A planet has a *true* value for its mass at one point in time, that is simply not known with absolute precision. This uncertainty is then expressed by a probability distribution. The sentence 'The earth has a mass of $5.9723(6) \cdot 10^{24}\,\mathrm{kg}$' does not mean that there are several – or infinite – universes where the mass of the earth is different in each and they average to $5.9723 \cdot 10^{24}\,\mathrm{kg}$ with a standard deviation of $6 \cdot 10^{20}\,\mathrm{kg}$. The sentence makes a reference to one instance of earth, with an uncertain value for the associated mass.

It is convenient to perform calculations on the logarithmic scale, as probability densities can become very small and large. In addition the sign is flipped, so that the search for the highest probability density corresponds to a minimisation problem. These negative log-probabilities are import for numerical analysis.

Some notes on the quantities used. For the continuous distributions of a random variable x the probability density function is denoted $\mathrm{pdf}(x)$ or $p(x)$. How this density changes under parameter transform is presented in section 5.5. The cumulative distribution function (cdf) is its integral

$$\mathrm{cdf}(x) = \int_{-\infty}^{x} \mathrm{d}x'\, \mathrm{pdf}(x') \tag{2.36}$$

and is of interest when determining the median position of a distribution.

2.6. Inverse problems

In physics terms an inverse problem can be thought of as the task to infer a quantity u which cannot be measured directly, but only via its effect on another quantity – y. The connection is given by a physical system, reacting to our quantity of interest. Ideally the behaviour allows us to infer information about u. A simple example is a linear operator $\mathbf{A}$ acting on $\vec{u}$:

$$\vec{y} = \mathbf{A} \cdot \vec{u} \,. \tag{2.37}$$

A more detailed introduction is given in [Gla11, ch. 1] and in context of regression in [LDT14, sec. 21.1.5]. As usually only a finite set of measurements is available, the quantities are denoted as vectors, while they can be scalars, vectors or higher dimensional fields. If $\mathbf{A}$ is perfectly known and can be inverted, and a perfect measurement of $\vec{y}$ would allow us to invert $\mathbf{A}$ and find $\vec{u}$:

$$\vec{u} = \mathbf{A}^{-1}\vec{y} \,. \tag{2.38}$$

However, for real systems there is an uncertainty in our measurement of $\vec{y}$. For a set of data $\vec{d}$, realisations of $\vec{u}$ the uncertainty – often seen as noise ϵ – leads to

$$\vec{d} = \mathbf{A} \cdot \vec{u} + \vec{\epsilon} \,. \tag{2.39}$$

For a badly conditioned system $\mathbf{A}$ the direct inversion will lead to an amplification of noise

$$\vec{u} = \mathbf{A}^{-1}\left(\vec{d} - \vec{\epsilon}\right) \,. \tag{2.40}$$

and hence renders the result useless. This can be seen, when looking at the eigendecomposition of $\mathbf{A}$

$$\mathbf{A} = \mathbf{Q}\mathrm{diag}(\vec{\lambda})\mathbf{Q}^{-1} \tag{2.41}$$

with eigenvalues λ in the diagonal matrix $\mathbf{D}$. Using the linearity of this model system gives further insight

$$\vec{u} = \mathbf{A}^{-1}(\vec{y} - \vec{\epsilon}) = \mathbf{A}^{-1}\vec{y} - \mathbf{Q}\mathbf{D}^{-1}\mathbf{Q}^{-1}\vec{\epsilon} \,. \tag{2.42}$$

The first operation on the noise component

$$\vec{\delta} = \mathbf{Q}^{-1}\vec{\epsilon} \tag{2.43}$$

is the noise term transformed into the eigenspace of the operator $\mathbf{A}^{-1}$. The following application of $\mathbf{D}^{-1}\vec{\delta}$ amplifies these noise components with the inverse of the eigenvalues of $\mathbf{A}$. The last application of $\mathbf{Q}$ transforms the amplified noise back into the initial coordinate system. The wider the range of eigenvalues of $\mathbf{A}$ – in other words the worse its condition number $\kappa = \lambda_{\mathrm{max}}/\lambda_{\mathrm{min}}$ – the stronger the noise is amplified in this process, relative to the *good* components with large eigenvalues.

Note: The operator **A** can also describe a complicated connection – say non-linear with phase-transitions. The presented derivation holds for linear and finite operators.

A powerful alternative to the presented direct inversion, is the probabilistic approach. This is presented in the next section. An important benefit is that the uncertainty of the measurements are considered, allowing to quantify a credibility range for the solution.

2.7. Bayesian inference

Bayes' theorem is foremost a rule about updating probability distributions. One application is to combine knowledge prior to an experiment with information obtained from new data. This relationship allows to quantify the probability distribution of parameters, which model a quantity of interest related to the experimental data – or any uncertain measure. The involved components for this analysis are explained in this section. Good sources are for example [Jay03] and [LDT14].

The central physics problem in this thesis is to infer the heat load to a surface, given temperature information from that surface, which represents an inverse problem. The *classic* approach is explained in section 4.3, where the temperature information is used as a boundary condition to solve a diffusion problem in time. This classic approach avoids the inverse problem entirely, but this comes with limitations, explained in chapter 4.

2.7.1. Bayes' theorem

To derive Bayes' theorem, only basic properties of probabilities are necessary, here for propositions M and D. Using the product rule $P(M|D) = P(M, D)/P(D)$, and the commutativity $P(M, D) = P(D, M)$ we find Bayes' theorem

$$P(M|D) = \frac{P(D|M)\,P(M)}{P(D)} \tag{2.44}$$

This can be seen also as the sum rule together with independence of M and D. A typical scenario is to update a model M given data D. M and D can be continuous or discrete. There are four parts in this equation, which are typically denoted as:

- $P(M \mid D)$: posterior – probability for the model (parameters) given the data
- $P(D \mid M)$: likelihood – how probable is the realisation D given a model M?
- $P(M)$: prior – how probable is model M, without considering data D?
- $P(D)$: evidence – how probable is the data?

The posterior is typically the quantity of interest. However, it is often a high-dimensional quantity and needs to be summarised in a suitable way. Often only the position with the highest probability density is reported as the maximum a-posteriori (MAP) solution. By approximating the pdf locally with a Gaussian, its standard deviation is used as a measure for the uncertainty or stability of the solution. For multimodal distributions these approximations can be misleading.

The likelihood quantifies the mismatch between model prediction and available measurement. The choice depends on the problem at hand. Often a Gaussian distribution is used for continuous or the Poisson distribution for discrete distributions. Using– and neglecting – a constant prior density, the maximum likelihood solution corresponds to the classic frequentist view. This often yields an incomplete picture, but is typically somewhat easier to determine. See [LDT14, sec. 2.3] for a more detailed explanation with examples.

The prior $P(M)$ describes the probability, for the model to be true before including the data D. This can include knowledge from previous measurements, so that sequential use of new data allows to sharpen the distribution for every new bit of data. The frequentist analysis often uses a flat prior, constant for each parameter, which can however only be normalised when limits are set. Introducing hard boundaries corresponds to setting the prior to zero beyond these limits, so that a frequentist analysis often includes this prior without knowing about it.

The evidence

$$P(D) = \int \mathrm{d}M \; P(D \mid M) \cdot P(M) \tag{2.45}$$

is also known as marginal likelihood or integrated likelihood or model evidence. For a single model this value is constant and often dropped from the equation. However, when comparing various models, this factor is essential.

Another view point is to see the evidence as the flexibility of the model with respect to the data. A very versatile model can describe any data, but for a simple model the given data is compatible with a larger fraction of the parameter space. In the words of J. Skilling [Ski98, p. 3]: "Different priors will induce different evidence values $Pr(D)$ (strictly, $Pr(D \mid$prior and other assumptions)). An inappropriate prior shows up through a relatively low evidence value. The art of setting a prior is to allow sufficient flexibility to cover all the inputs X that one might plausibly need in order to fit the data reasonably, without widening the field so much that the particular data D will become intrinsically implausible. In other words, ask a question appropriate to your data."

The evidence is the Bayesian equivalent to Ockham's Razor, see [LDT14, sec. 3.3] and [Jay03, ch. 20]. When a simple model explains the data good enough, don't employ more unknowns or wild speculations to your explanation. To quantify this, the ratio of two posteriors for models M_1 and M_2 is employed [Jay03, eq. 20.3]:

$$\frac{P(M_1 \mid D)}{P(M_2 \mid D)} = \frac{P(D \mid M_1)}{P(D \mid M_2)} \cdot \frac{P(M_1)}{P(M_2)} \tag{2.46}$$

The first fraction – comparing the likelihoods – is referred to as Bayes factor, the second fraction as prior odds. For equal prior probability, the likelihood is enough to look at. For more complex models the prior volume is typically larger, which reduces the average prior probability density. A more complex model is then justified, only when the likelihood is improved at a similar level.

2.7.2. Priors

Priors encode the knowledge available before the experiment was conducted. When a first data analysis was performed and a posterior for the first data set was obtained, this is the prior information for subsequent data. There is plenty of literature on how to choose a prior depending on the problem [KW96]. Properties to consider are e.g. scale independence and stability. The choice of prior and likelihood defines the distribution class of the posterior. When the posterior is described by the same distribution class as the prior, the prior is called conjugate for the used likelihood [RS61]. This can be valuable for sequential data evaluation.

Flat prior

The flat or uniform prior is a family of densities, which are constant on their support. The boundaries of this support are defined by two real values x_L and x_U with

$$\infty < x_L < x_U < \infty \ . \tag{2.47}$$

For a finite support, the probability density is finite and the distribution is well defined with norm

$$Z_f = x_U - x_L \ . \tag{2.48}$$

The density is accordingly

$$p(x) = \begin{cases} \dfrac{1}{Z_f} & : x_L <= x <= x_U \\ 0 & : else \end{cases} \tag{2.49}$$

and depends solely on the upper and lower limits. Without those limits, the probability distribution cannot be normalised. The hard cuts can be a reasonable choice for some problems, but represent actually a very precise knowledge about this transition.

The flat prior is implicitly used in every frequentist analysis, though typically without limits. The resulting densities are therefore not normalised and treated as $p(x) = 1 \in \mathbb{R}$.

While the flat prior is sometimes called non-informative, it actually expresses information about the scale of the problem.

Gaussian prior

As the name suggests it is the Gaussian distribution, used as prior. See [Tho18, ch. 8] for more details. In its multivariate form it is written as

$$p(\vec{x} \mid \vec{\mu}, \mathbf{\Sigma}) = \frac{1}{(2\pi)^{N/2}} \frac{1}{\mid \mathbf{\Sigma} \mid^{1/2}} \cdot \exp\left[-\frac{1}{2} \cdot (\vec{x} - \vec{\mu})^T \, \mathbf{\Sigma}^{-1} \, (\vec{x} - \vec{\mu})\right] \tag{2.50}$$

with covariance matrix $\mathbf{\Sigma}$ and mean vector $\vec{\mu}$ for N parameter. The inverse of the covariance matrix is the Hessian matrix

$$\mathbf{H} = \mathbf{\Sigma}^{-1} \, . \tag{2.51}$$

When the covariance matrix is diagonal, it can represented by the vector $\vec{\sigma}$ containing the corresponding standard deviations $\sigma_i^2 = \Sigma_{i,i}$. This can be used to reduce (2.50) to

$$p(\vec{x} \mid \vec{\mu}, \vec{\sigma}) = \frac{1}{(2\pi)^{N/2}} \cdot \frac{1}{\prod_{i=1}^{N} \sigma_i} \cdot \exp\left[-\frac{1}{2} \cdot \sum_{i=1}^{N} \left(\frac{x_i - \mu_i}{\sigma_i}\right)^2\right] \, . \tag{2.52}$$

The negative log-probability is

$$-\ln(p(\vec{x} \mid \vec{\mu}, \vec{\sigma})) = \frac{N}{2} \cdot \ln(2\pi) + \sum_{i=1}^{N} log(\sigma_i) + \frac{1}{2} \cdot \sum_{i=1}^{N} \left(\frac{x_i - \mu_i}{\sigma_i}\right)^2 \tag{2.53}$$

and has a quadratic dependency on each element of $\vec{x}$.
A specific property is the trivial marginalisation [RW06]. When being interested in $p(\vec{x}')$ for x' being a subset of x, e.g. without the j'th parameter:

$$x' = (x_1, \, \ldots \, , \, x_{j-1}, \, x_{j+1}, \, \ldots \, , \, x_N)$$

the j'th entry in $\vec{\mu}$ as well as the j'th line and row in the covariance matrix $\mathbf{\Sigma}$ are removed, giving $\vec{\mu'}$ and $\mathbf{\Sigma}'$

$$p(\vec{x'}|\vec{\mu'}, \Sigma') = \int \mathrm{d}x_j p(\vec{x}|\vec{\mu}, \Sigma) \tag{2.54}$$

$$= \mathcal{N}(\vec{\mu'}, \Sigma') \tag{2.55}$$

The Gaussian distribution is a reasonable choice for parameters, where the range of plausible values is known. Choosing the standard deviation generously, the distribution is close to uniform around its centre. In contrast to the uniform distributions there are no hard limits, which is beneficial in numerical applications.

Cauchy prior

The Cauchy distribution is defined as [LDT14, sec. 7.5.8]

$$\operatorname{pdf}(x \mid t, s) = \frac{1}{\pi} \cdot \frac{s}{s^2 + (x-t)^2} \tag{2.56}$$

with the centre t and width s. Its cumulative distribution function is

$$\operatorname{cdf}(x \mid t, s) = \frac{1}{2} + \frac{1}{\pi} \cdot \arctan\left(\frac{x-t}{s}\right) . \tag{2.57}$$

Around the centre value t the pdf behaves similarly to the Gaussian, but it differs in various aspects. The heavy tails give less weight to outliers than the Gaussian distribution, when used as likelihood function. As a result of the slow decay no higher moments than the zeroth exist. This implies, that the mean does not exist strictly. However, the principal value is equivalent to the parameter t. Also the second moment is not finite, the standard deviation is infinite.

The similarity to the normal distribution becomes more visible in negative log space

$$-\ln(p(x \mid t, s)) = \ln(\pi/s) + \ln\left(1 + \left(\frac{x-t}{s}\right)^2\right) \tag{2.58}$$

Around $x = t$ the logarithm is proportional to $\left(\frac{x-t}{s}\right)^2$ – which is a quadratic dependence. For larger values the increase is however not quadratic, but only two times the logarithm of x/s. However, its use can affect the exploration, as the distribution is not concave and the first derivative approaches zero for $\mid x - t \mid \gg s$.

Entropic prior

Entropy is a useful tool in thermodynamics, as well as in statistics. Plenty of literature is available [CP04; Ski89a; Jay03]. Employing maximum entropy corresponds to choosing a solution with maximum uncertainty or ignorance towards what we do not know. A short introduction is found in [Kes09].

The starting point is the information entropy

$$S(h) = \sum_i^N \left(h_i - m_i - h_i \ln\left(\frac{h_i}{m_i}\right)\right) \tag{2.59}$$

for the hidden image values h_i and the standard model m_i for each amplitude. The distribution of the standard model is chosen to be uniform $m_i = m \forall i$ for the rest of this section. The probability distribution of the hidden image values given the regularisation parameter α is then

$$\mathrm{pdf}(h|\alpha) = \left(\Pi_{i=1}^{N}\left(\frac{\alpha}{2\pi}\right)^{1/2}\right)\exp\left(\alpha S\left(h\right)\right)\ . \tag{2.60}$$

α has to be marginalised in this case, but its most likely value can be determined if set as free parameter.

In any case α is treated as a scale parameter with the prior

$$p(\alpha) = \frac{1}{\alpha}\cdot\frac{1}{\ln(\alpha_{\mathrm{max}}/\alpha_{\mathrm{min}})} \tag{2.61}$$

with the second term arising from the normalisation.

For implementation the negative logarithm is used, yielding

$$\begin{aligned}
-\ln(p(h|\alpha)) &= -\ln\left(\Pi_i^N(\frac{\alpha}{2\pi})^{1/2}\right) - \alpha S(h) && (2.62)\\
&= \frac{1}{2}\sum_i^N\left(-\ln(\frac{\alpha}{2\pi})\right) - \alpha\left(\sum_i^N\left(h_i - m - h_i\ln(\frac{h_i}{m})\right)\right) && (2.63)\\
&= \frac{1}{2}\sum_i^N\left(-N/2\cdot\ln(\frac{\alpha}{2\pi})\right) - \alpha\left(\sum_i^N\left(h_i - h_i\ln(\frac{h_i}{m})\right) - N\cdot m\right) && (2.64)\\
&= \sum_i^N\left(-\alpha h_i\left(1-\ln(\frac{h_i}{m})\right)\right) - N/2\cdot\ln(\frac{\alpha}{2\pi}) + \alpha\left(N\cdot m\right) && (2.65)
\end{aligned}$$

The factor α can be seen as *weight* or the amount of regularisation. It intrinsically depends on the data and cannot be set a-priori. A comparison of approximations versus integration is found in [FVD96].

The entropic prior introduces no correlation to the parameters. When applied to an image, there is no correlation between the amplitudes. The prior is not invariant to the resolution of the problem and scales with the number of the amplitudes.

The marginalisation of α is not trivial. An analytic integration is typically not possible and numerical integration can be cumbersome. Depending on the prior for α the integral may diverge towards 0 and ∞. An alternative is to compute the MAP $p(\hat{\alpha} \mid h, d)$. In this case the prior is limited to a finite range – e.g. from 10^{-4} to 10^{+4}. In this range the regularisation is expected to work and a diverging factor α is then seen as an indicator for problems.

Marginalisation via tabulation

As presented in [SWW93] the integral to marginalise α from equation (2.60)

$$p(h) = \int p(h \mid \alpha)\cdot p(\alpha)\mathrm{d}\alpha \tag{2.66}$$

can be solved efficiently for independent amplitudes. The normalisation element $Z_S(\alpha)$ in the conditional entropy

$$p(h \mid \alpha) = \frac{\exp(\alpha S(h))}{Z_S(\alpha) \cdot \prod_{i=1}^{N} \sqrt{h_i}} \tag{2.67}$$

can be factorised into the product of N one-dimensional integrals:

$$Z_S(\alpha) = \prod_{i=1}^{N} \sqrt{em}\, Z_1(em\alpha) \tag{2.68}$$

The integral hidden in

$$Z_1(\beta = \alpha \cdot em) = \int_0^\infty \mathrm{d}g\, \frac{\exp\left[-\beta \cdot \left(g \ln(g) + \frac{1}{e}\right)\right]}{\sqrt{g}} \tag{2.69}$$

is made available using tabulation. The integrand diverges for $g \to 0$, but is integrable. Some results are $Z_1(10^{-3}) \approx 24.61$, $Z_1(1) \approx 2.14$, $Z_1(10^{+3}) \approx 0.079$.

Both integrals – (2.66) and (2.69) – are solved on a log-scale for the respective argument. In addition, the logarithm of the integrand values I_i are stored and their peak value $M = \max(I)$ is determined. When summing the contributions to approximate the integral an offset of -M is added in log-space. This allows to evaluation the exponential function without overflow and negligible underflow:

$$\sum_{i=1}^{N} I_i = \sum_{i=1}^{N} \exp\left(\ln(I_i)\right) = \exp(M) \sum_{i=1}^{N} \exp\left(\ln(I_i) - M\right) \ . \tag{2.70}$$

The largest argument appearing in the exponential is 0, preventing overflow. At the same time, cutoff is experienced for an argument of about -700, giving sufficient headroom for small but non-negligible values. The summation is carried out using a Neumaier sum [Neu74] to extent the relative precision to effectively 10^{30} instead of the typical value of 10^{15} for 64-bit bit precision.

Remark: the offset could also be set to a higher value like $600 - M$, evaluating to a peak value of $3.77 \cdot 10^{260}$, still leaving ample headroom to the largest value of $1,7 \cdot 10^{308}$. This extends the headroom to the cutoff to 1300 in log-space. However, the numerical summation process will lead to truncation for small values anyway.

2.7.3. Likelihood

The probability for a measurement $\vec{d}$ given a model M and its parameters $\vec{\Theta}$ is given by the likelihood $\mathrm{pdf}(\vec{d} \mid \vec{\Theta}, M, I)$. Its form depends on the problem at hand, e.g. continuous or discrete quantities.

The Poisson distribution [Tho18, p. 39] the probability of observing k events in a counting process, for λ being the mean count

$$p(k \mid \lambda) = \frac{\lambda^k}{k!} \exp(-\lambda) \; . \tag{2.71}$$

For an increasing number of mean events λ the normal distribution serves as an approximation, which can be considered as an almost prefect surrogate for $\lambda > 100$. For this reason the Gaussian distribution is often used as pdf for measurements with sufficient counts.

However, a single outlier in the data can perturb the entire parameter estimation. When rare events perturb the measurement by more than a few standard deviations, a more benign – i.e. heavy tailed – distribution should be used. A common choice is the Cauchy distribution introduced in the previous section 2.7.2.

2.7.4. Ockham's razor

Ockham's razor is a principle, according to which simple explanations should be preferred to complicated ones, at least when both explanations make predictions of similarly good quality. This principle is rooted in Bayes' theorem. While the likelihood and the prior are usually simple to compute, the evidence – often a high-dimensional integral – is more difficult to compute and therefore often ignored. Nested sampling is one example of an efficient numerical integration scheme for the computation of the evidence [Ski06; FS13].

2.7.5. Summarising high-dimensional pdf's in a single number

While Bayes' theorem yields a full probability distribution for all parameters of interest, technical use often relies on a decision. How this choice is made depends on the problem at hand and the posterior distribution itself. The typical method for parameter estimation is to make an Gaussian expansion around the mode or mean of the distribution. The mean of this distribution is a simple point-estimate, which can be seen as finite-dimensional answer. The covariance matrix – or the derived vector with standard deviations – quantifies the uncertainty and interaction of the parameters around the point of the maximum a-posterior *MAP* density.

3. Infrared Thermography

This chapter introduces the infrared (IR) diagnostic used to measure the experimental data.

3.1. Basics

Thermography is a method to deduce the temperature of a surface by measuring its emitted thermal radiation. The underlying basis is Planck's law of radiation, as introduced in section 2.3, describing the emission of electromagnetic radiation from an ideal black body with finite temperature.
IR cameras are used to determine the spatial temperature distribution of surfaces. They are suited for remote measurements on the divertor targets, because they do not suffer from the harsh conditions in the divertor region. The temperature evolution is used to estimate heat flux densities onto the surface. This is done by solving the heat diffusion equation (2.15) with the surface information of the thermography as boundary condition.

Figure 3.1.: Lower Divertor in ASDEX Upgrade, showing the higher surface roughness – therefore lower reflectivity – of the sandblasted area on the tiles [IPP18].

Sources of error are the incomplete knowledge of the temperature dependent emissivity $\epsilon(T)$ of the tungsten targets and reflections due to their high reflectivity ($\epsilon \simeq 0.2$). Treating the target plates of interest with sandblasting results in surfaces with higher emissivity, see figure 3.1, lowering both uncertainties.

Optical filters are used, so that the camera system is only sensitive to a wavelength region around 4.7 µm. At this wavelength the measurement is reliable and also sufficiently sensitive in the most relevant temperature region from 300 K to 2500 K [Sie14]. The lower (higher) the wavelength, the lower the sensitivity – the increase of photon flux with temperature – at lower (higher) temperatures. In addition, the systematic error due to line radiation produced in the divertor region is smaller, compared to measurements at lower wavelengths.

3.2. Notation

All pixels of the detector are integrating – i.e. being sensitive to – the incoming photon flux at the same time and are read out as single frame. A single pixel value of a frame is called a sample. The time associated with a sample refers to the end of the finite integration time. This is important, as the integration time can change on a frame-to-frame basis, for the system used. The sample interval refers to the distance in time between two samples. For a constant sample rate this is the inverse of the sample frequency.

A part of the data analysis sections uses for simplicity the notation of a scalar samples in space, dropping the spatial index. As the samples of one frame are obtained at the same time, the spatial index can be added without ambiguity.

3.3. Sensitivity

Here the likelihood of the measurement and the resulting sensitivity of the diagnostic are introduced. The sensitivity of the used IR thermography system is derived together with the uncertainty of an obtained sample. The deviation when using a normal distribution in temperature space as likelihood is analysed.

The sensor effectively integrates an incoming photon rate $\Gamma(t)$ time interval t_{int} to a total photon count of

$$N = \int_{t_0}^{t_0+t_{\text{int}}} \mathrm{d}t\Gamma(t) \; . \tag{3.1}$$

That is the total number of photons, which interacted with electrons. The interaction with the electrons leads to a change of electric charge, which is in the end measured as discrete counts (cts) and is proportional to the detected events

$$\text{cts} = N/\,\text{ppc} \tag{3.2}$$

with ppc representing the discretisation in terms of *photons per count*. Typically the average count rate is used, to determine the surface temperature. Relation (3.2) also holds for the respective average rates – count rate (ctr) and photon rate Γ:

$$\mathrm{ctr} = \Gamma / \mathrm{ppc} \ . \qquad [1/\mathrm{s}] \quad (3.3)$$

For sake of simplicity equation (2.32) is simplified by introducing calibration parameters of the camera as constants a and b

$$T(\mathrm{ctr}) = \frac{a}{\ln\left(\frac{b}{\mathrm{ctr}\cdot\mathrm{ppc}} + 1\right)} \qquad (3.4)$$

in terms of the count rate. The calibration depends on aspects like the camera sensitivity, discretisation and amplification and also on the optical system. The first parameter is

$$a \propto \frac{hc}{\lambda_{\mathrm{eff}} k} \qquad (3.5)$$

where the proportionality sign indicates that the actual value depends on the optics in between the surface and the camera and the photon efficiency of the sensor. λ_{eff} is the effective wavelength used in Planck's law for monochromatic radiation, whereas the sensor is sensitive to a finite wavelength range. The second parameter is defined as

$$b = \frac{2\pi c\epsilon \cdot c_0}{\lambda_{\mathrm{eff}}^4} \ . \qquad (3.6)$$

The wavelength dependent sensitivity and transmission of the optical system are challenging to determine in practice as separate quantities. For this reason the system is calibrated as a whole, leaving only parameters a and b to be inferred. Figure 3.2 shows the response of the system at 4.7 µm wavelength on a linear scale (left) and on a logarithmic scale (right). The count rate scales not linear with the temperature, implying a curvature in the sensitivity. Accordingly, the shape of probability density functions changes under the transformation, see section 5.5.

The pdf of the detected photons is described by the Poisson distribution, which is however approximated by a normal distribution. This is justified, as for this system N is on the order of 10^3 to 10^6. The distribution is accordingly

$$\mathrm{pdf}(N) = \mathcal{N}(N_0, \sqrt{N_0}) \ . \qquad (3.7)$$

The distribution for the count rate is related via the integration time and the discretisation:

$$\sigma_{\mathrm{ctr}} = \sqrt{\mathrm{ctr}_0 \cdot t_{\mathrm{int}}} / \mathrm{ppc} \ , \qquad (3.8)$$

$$\mathrm{pdf}(\mathrm{ctr}) = \mathcal{N}(\mathrm{ctr}_0, \sqrt{\mathrm{ctr}_0 \cdot t_{\mathrm{int}}} / \mathrm{ppc}) \ . \qquad (3.9)$$

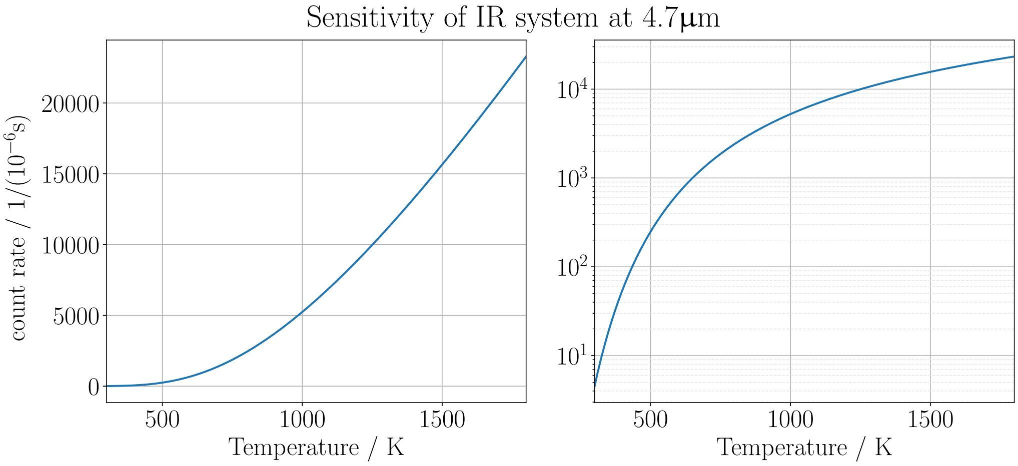


Figure 3.2.: Count rate of the IR-camera at highest sensitivity level in lin-space (left) and log-space (right). The temperature is in the range of (0 – 1500) °C.

For the use of THEODOR the surface temperature is calculated from the measured data. For the non-linear Planck law and sensitivity functions, the normal distributed count rate does not in general correspond to a normal distribution in temperature space. Not using the right pdf can lead to a bias in the mean temperature and obscure the uncertainty analysis, see section 5.5. To investigate the difference between the correct pdf and a Gaussian approximation in temperature space, the mapping is presented and applied to two distributions.
The transformation – based on equation (5.16) – is

$$\mathrm{pdf}_T(T) = \mathrm{pdf}_{\mathrm{ctr}}(\mathrm{ctr}(T)) \cdot \left| \frac{\partial T}{\partial \, \mathrm{ctr}}^{-1} \right| . \tag{3.10}$$

The derivative of (3.4) with respect to the count rate is

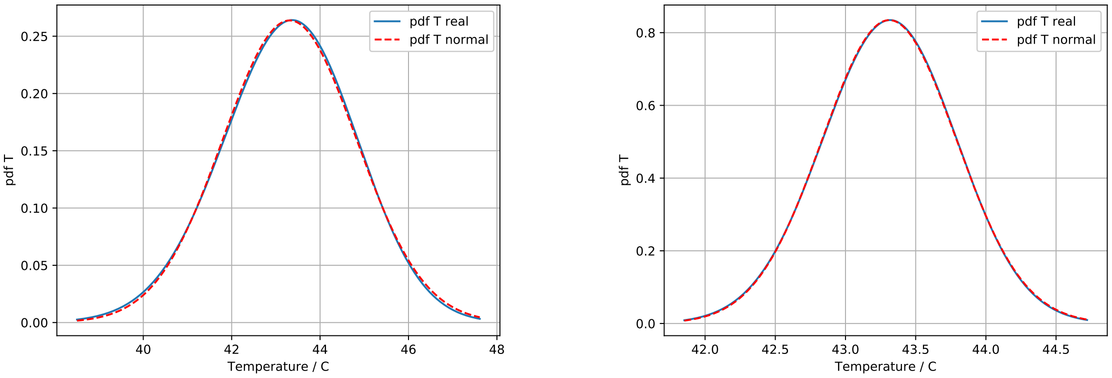
$$\frac{\partial T}{\partial \, \mathrm{ctr}} = \frac{a \cdot b}{\mathrm{ctr}(b + \mathrm{ctr}) \cdot \ln^2 \left(b / \mathrm{ctr} + 1\right)} , \tag{3.11}$$

The mapped distribution is approximated by a Gaussian, based on the temperature obtained from the mean photon flux: $T_0 = T(\Gamma_0)$. The corresponding standard deviation μ_T is calculated based on the first derivative and the mean temperature

$$\mu_T = \sigma_{\mathrm{ctr}} \cdot \left| \frac{\partial T}{\partial \, \mathrm{ctr}} \right|_{T=T_0} . \tag{3.12}$$

Figures 3.3(a) and 3.3(b) show the mapped pdf (3.10) as solid, blue line, and the Gaussian approximation as dashed, red line. The camera parameters corresponds to the best resolving setup available, with a camera discretisation of 45.77 photons per count. The two cases correspond to measurements with of 10 (a) and 100 (b) camera counts, based on 458 and

4577 counted photons. This analysis focuses on the uncertainty due to the photon statistics and ignores additional uncertainties, like electronic noise. Therefore the standard deviation is given as the square root of the counted photons. Note that the temperature $T \approx 43\,^\circ\mathrm{C}$ in this example is about the lowest temperature encountered in experiments. At this temperature range the non-linearity of the system is higher than for any higher temperatures, representing the worst case.



(a) pdfs for 10 counts with $t_{\mathrm{int}} = 1\,\mu\mathrm{s}$. This corresponds to 458 photons respectively a relative uncertainty of 1 / 21.3.

(b) pdf for 100 counts with $t_{\mathrm{int}} = 10\,\mu\mathrm{s}$. This corresponds to 4578 photons respectively a relative uncertainty of 1 / 67.7.

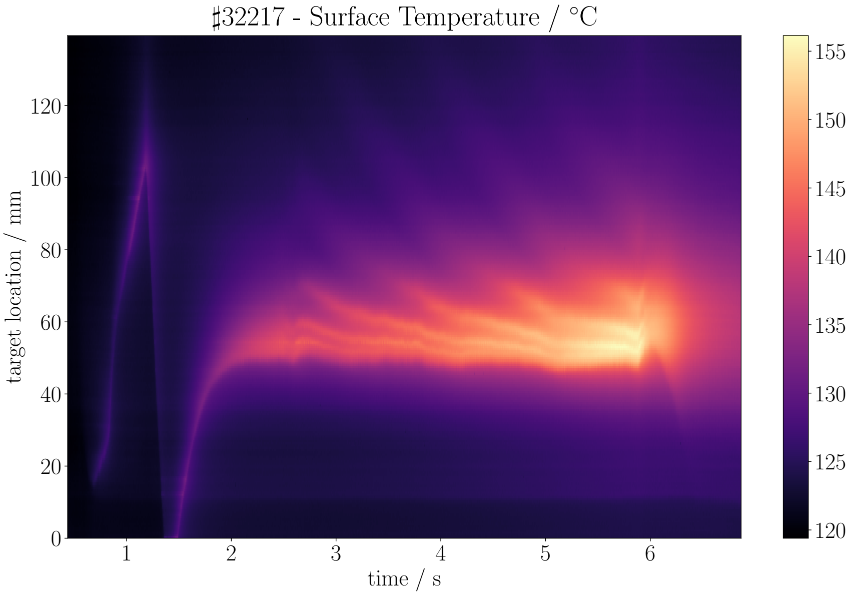
Figure 3.3.: A Gaussian pdf mapped from photon flux space to the temperature space for two integration duration values of the sensor. The temperature is fixed and the integration time t_{int} is varied.

Despite the large uncertainty – compared to the expected values (25 – 100) mK – the non-linearity is apparently negligible. The examples show that a normal distribution in temperature space is a good approximation for the likelihood, for the diagnostic used. However, for such low count values, the electronic noise from the camera can not be neglected. The wider the pdf in the photon flux space, the more it is affected by the non-linearity of the mapping. This result does not necessarily transferred to other diagnostics, which do not use adaptive integration times or can not prevent these low counting statistics. For other wavelengths used, e.g. near IR cameras for machine protection [Her+11], this analysis should be repeated for the relevant parameters.

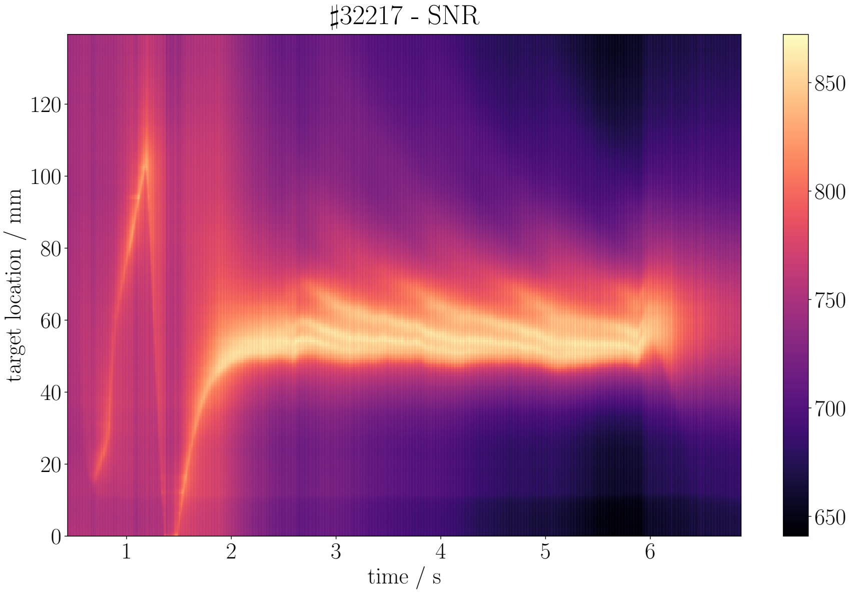
3.4. Signal-to-noise ratio distribution

For a stationary heat load profile on a rectangular tile, which is thermally insulated, the highest surface temperatures appears at the position receiving the highest heat flux densities. Figure 3.4(a) shows the surface temperature trace determined from the camera data for such a discharge. Figure 3.4(b) shows the effective signal to noise ratio. The SNR corresponds to the square root of the detected events of the camera, with the discretisation and electronic noise being well below these values. In the shown case an integrator with a capacity of 1.5 Me is used with the ADC working at 15 bit resolution. The discretisation – the factor from camera counts to photon count – is therefore $1.5 \cdot 10^6$ by $2^{15} \approx 46$ photons per count. This discharged was heated with about 0.7 MW– which is little compared to the total installed 38 MW [Mey+19] heating power. The total energy input to the divertor is therefore small and correspondingly the surface temperature increases only from about 100 °C to about 150 °C.

A similar comparison is presented in figures 3.5(a) and 3.5(b). The plasma is heated stronger – 8 MW instead of 0.7 MW in the case before – leading to a stronger temperature increase of the surface from about 100 °C to close to 700 °C. Here the spread of the SNR within a single frame is much more pronounced. The integrator in this case is larger with 5.8 Me at only 14-bit resolution, leading to about 354 photons per count.

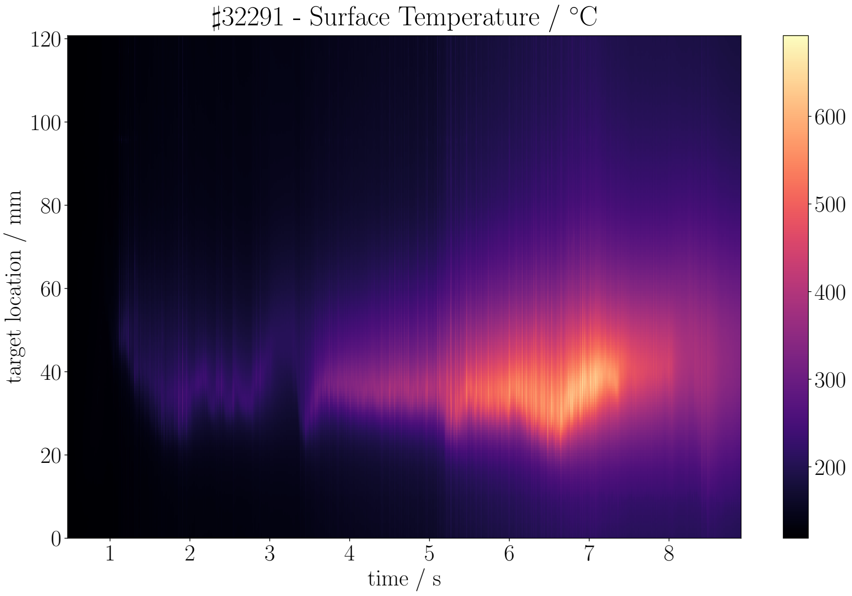


(a) Time trace of surface temperature for discharge ♯32217.

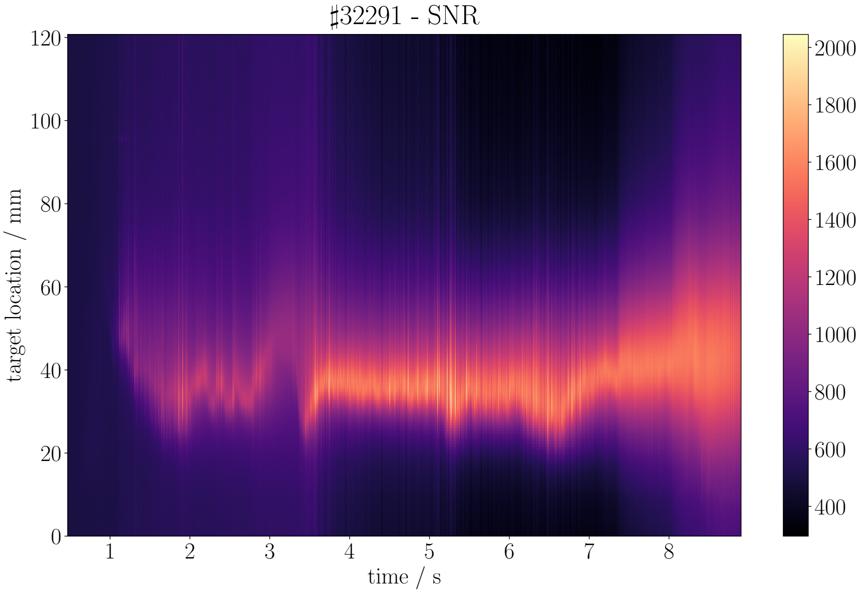


(b) Time trace of signal to noise ratio for discharge ♯32217, which is the relevant quantity in the likelihood.

Figure 3.4.: Temperature and signal to noise ratio for AUG discharge ♯32217.



(a) Time trace of surface temperature for discharge ♯32291.



(b) Time trace of signal to noise ratio for discharge ♯32291, which is the relevant quantity in the likelihood

Figure 3.5.: Temperature and signal to noise ratio for AUG discharge ♯32291.

3.5. Limits of single-colour thermography & potential diagnostic improvements

Under certain conditions the narrow wavelength region used for the thermography diagnostic is not only sensitive to Planck-radiation of solid objects. Radiation from the plasma, like Bremsstrahlung 2.4, is emitted along the line of sight of the camera observing the divertor.

After ELM crashes [Wil+18] – events delivering energy and particles to the volume intersected by the line of sight – the measurement conditions are not fully represented by the forward model presented so far. The additional detected radiation leads to an overestimation of the tile surface temperature [Sie14, ch. 2]. This is interpreted as a large heat influx, needed to reach these temperatures. When the volume radiation decreases, the determined surface temperature drops, leading to an underestimation of the heat load. For a strong perturbation, only a heat out-flux – from the tile to the plasma – could explain the data. This is considered non-physical, and an indicator for problems when using THEODOR. In the statistical approach a priors excluding negative heat loads is used. However, when one profile is inferred at a time – the posterior truncation is described in 7.2.4 – the perturbation leads to the described overestimation of the heat influx. Preventing the tile surface from losing heat, the result are several consecutive profiles with zero heat influx. The perturbation to the heat distribution in the tile is actually larger in this case.

Neither of these scenarios is desirable. An alternative is to infer more than one profile at a time, so that further data limit the overestimation. In addition another likelihood with a heavy tail for the overestimation can be used. While this solution is reasonable from a probabilistic point of view, it still has to be considered incomplete.

A simpler solution would be to ignore the corrupted data and to ask for a simple measure like the mean influx during the affected time period. However, this would require a reliable detection of data corruption by volume radiation, which is not available so far. A second viewing angle with a synchronised camera would allow to detect a disagreement, but this is not available.

The current solution with THEODOR is to see negative heat loads in the model as clear indication for data corruption. Subsequently the negative heat loads together with the heat load profiles before and after are ignored. However, this only works for strong perturbations and relies on the experience of the analysing person.

A better solution is to use spectroscopic methods, to identify other contributions like Bremsstrahlung.

A 2D detector can be turned into a spectrometer with one orientation of the detector representing the wavelength and the second orientation representing the line of sight angle along a plane:

$$\Gamma(x_i, y_j) = \gamma(\Theta_i, \lambda_j) \tag{3.13}$$

This setup can be used to measure along the poloidal orientation on the target element with

spectral resolution. This allows to separate black-body radiation from volume radiation, as was shown in [Nem17].
The new model for the photon flux is then

$$\Gamma_{\text{det}}(T_s, T_e, n_e) = \Gamma_{\text{Planck}}(T_s) + \Gamma_{\text{Brems}}(T_e, n_e) \tag{3.14}$$

with three unknowns. The plasma electron temperature T_e and plasma electron density n_e are additional parameters. Other diagnostics could be used for additional input, or this measurement can be used to obtain a distribution which can be used for further evaluation. By concluding this thesis, the diagnostics will still not be fully completed and therefore no experimental data are yet to be presented.

4. Forward model

As introduced in chapter 2 inference relies on the mapping from surface heat flux to the observed quantity, known as forward modelling. In this chapter considerations about the forward model are made, regarding the solver for the heat diffusion equation (HDE).

4.1. Splitting of the heat load time trace

The time trace of data for one completed experiment can be treated as one data-set, with the entire time trace of heat load patterns as one quantity to infer. This however presents a problem with millions of unknowns, which is hard to track. A separation of the time trace into single *slices* of heat load patterns appears natural. Due to causality any heat source affects the temperature distribution only after its contribution and due to the nature of the diffusion equation any change to the temperature distribution is damped in time. As the temperature distribution is measured at the surface with an IR system, the influence of any heat input is best quantified directly after its arrival.
In section 7.2.4 the truncation to approximate the posterior is derived.

4.2. Analytic solution for the semi-infinite rod

The effective power flux density resolution of the system is estimated using the analytic solution for a special case of the one dimensional HDE. The special case is a for an object starting at the origin $x = 0$ and extending to ∞. The heat load is applied homogeneously to the surface at $x = 0$. This is an approximation for finite sized objects, for times smaller than the diffusion time to other boundaries. The result is an analytic expression for the surface temperature, i.e. $T(x = 0)$, evolving in time t:

$$T(t) = \frac{2}{\sqrt{\pi}} \cdot \sqrt{t} \cdot q\frac{\sqrt{\chi}}{\kappa} + T_0 \qquad [\mathrm{K}] \qquad (4.1)$$

given as constant parameters: the initial temperature T_0, a surface heat flux density q, and the material properties diffusivity χ and conductivity κ. The heat equation was treated analytically by Fourier, using sinusoidal basis functions, leading to the famous Fourier Series. The original work *Theorie Analytique de la Chaleur* (The Analytical Theory of Heat) from 1878 is also available in English translation [Fou09].

The temperature increase in time is accordingly

$$\frac{\partial T}{\partial t} = \frac{1}{\sqrt{\pi t}} q \frac{\sqrt{\chi}}{\kappa} \,. \qquad \text{[K/s]} \quad (4.2)$$

The surface initially reacts fast to incoming heat flux – or a change of it in time – which is notoriously difficult to capture in a spatially and temporally discrete system.

Equation (4.1) connects the surface temperature to a constant heat load in time. Solving for q we can calculate the change in heat load – be it a perturbation or uncertainty – after the duration Δt and a perturbation of the surface temperature δT:

$$\delta q = \frac{\sqrt{\pi}}{2} \frac{\kappa}{\sqrt{\chi}} \frac{\delta T}{\sqrt{\Delta t}} \,. \qquad [\mathrm{MW\,m^{-2}}] \quad (4.3)$$

The precise material properties [Tol17] depend on the alloy used and the temperature. For pure tungsten the heat conductivity is about $\kappa \approx 180\,\mathrm{W/m\,K}$ ($130\,\mathrm{W/m\,K}$) at $T = 300\,\mathrm{K}$ ($800\,\mathrm{K}$). The mass density is about $\rho \approx 19\,\mathrm{kg\,m^{-3}}$ in the range $(300 - 800)\,\mathrm{K}$. The heat capacity is about $c_p \approx 24\,\mathrm{J/mol\,K}$ ($27\,\mathrm{J/mol\,K}$) at $300\,\mathrm{K}$ ($800\,\mathrm{K}$). The resulting diffusivity for tungsten is about $\chi = \frac{\kappa}{\rho c_p} \approx 5.7 \cdot 10^{-5}\mathrm{m^2\,s^{-1}}$.

For this estimate the values $\kappa = 150\,\mathrm{W/K\,m}$ and $\chi = 6 \cdot 10^{-5}\,\mathrm{m^2\,s^{-1}}$ are used together with a time step of $\Delta t = 1\,\mathrm{ms}$ – typical for the IR system at AUG – in equation 4.3:

$$\delta q \approx 1 \cdot \frac{\delta T}{\mathrm{K}} \cdot \mathrm{MW\,m^{-2}} \,. \qquad (4.4)$$

So for an effective temperature uncertainty of 25 mK from the infrared (IR) system, as achieved under ideal conditions, the expected uncertainty in heat load is $25\,\mathrm{kW\,m^{-2}}$. This resolution limit is confirmed by analysing synthetic data in section 8.1.

4.3. THEODOR code

THEODOR is a software tool commonly used at ASDEX Upgrade and other plasma experiments like JET or TCV. The abbreviation stands for THermal Energy Onto DivertOR and expresses the quantity of interest: the heat impinging on divertor elements. This tool was already used in 1994 for thermography data [Her+95] and has been developed since. For the discrete solver for the heat diffusion equation 2.16 only the explicate Euler scheme for the finite difference system was initially computationally affordable. Today an implicit scheme can solve the 2D system in about 1 ms on a single CPU core. This section will outline available options, compare them and argue why the implicit method is used for this thesis.

4.3.1. Finite difference formulation

THEODOR solves the 2D HDE in the divertor tile corresponding to the heat transport into the depth as well as perpendicular to the surface[1]. This is done using finite differences [Pre+07]. The temperature is substituted by the heat potential u, see section 2.1 for details, resulting in the semi-linear partial differential equation (PDE)

$$\frac{\mathrm{d}u}{\mathrm{d}t} = \chi \nabla u \tag{4.5}$$

with the diffusivity χ. The heat flux is expressed as

$$q = -\Delta u \; . \tag{4.6}$$

As a first step the spatial derivative is expressed via finite differences, turning the PDE into a system of ordinary differential equations. For these equations a numerical integration scheme is employed. The distance from cell i with size h_i to the adjacent cell centres is given by average cell size

$$\Delta_+ = \frac{h_i + h_{i+1}}{2} \; , \tag{4.7}$$

$$\Delta_- = \frac{h_i + h_{i-1}}{2} \; . \tag{4.8}$$

To ensure energy conservation the discrete system is at first formulated in terms of the heat flows at the left and right cell boundaries:

$$q_- = -\frac{u_i - u_{i-1}}{\Delta_-} \tag{4.9}$$

$$q_+ = -\frac{u_{i+1} - u_i}{\Delta_+} \tag{4.10}$$

leading to

$$\frac{\mathrm{d}u_i}{\mathrm{d}t} = \chi_i \frac{q_- - q_+}{h_i} \; . \tag{4.11}$$

The ratio q/h represents the rate of change of the heat potential due to the change in the internal energy. By the conservation considerations the weights are already prescribed and only the distance of the cells is to be chosen.

The second step is to approximate the rate of change with the finite difference:

$$\frac{u_i^{k+1} - u_i^k}{\Delta t} = \chi_i \frac{\dfrac{u_{i+1}^j - u_i^j}{\Delta_+} - \dfrac{u_{i-1}^j - u_i^j}{\Delta_-}}{h_i} \; . \tag{4.12}$$

---

[1] As result of this thesis the 3D solver is now available.

On the right hand side the time index is denoted with j for now. Choosing $j = k$ leads to an explicit formulation, $j = k + 1$ to an implicit formulation.
The three components χ, Δt and h_i are typically denoted as one factor $\frac{\chi_i \Delta t}{h_i}$ and can be combined with the discrete distances:

$$\gamma_- \equiv \frac{\chi_i \Delta t}{h_i \Delta_-} \ , \tag{4.13}$$

$$\gamma_+ \equiv \frac{\chi_i \Delta t}{h_i \Delta_+} \ . \tag{4.14}$$

For a general grid this leads to an equation with two resolution dependent coefficients used in conjunction with three field values for one update:

$$u_i^{k+1} - u_i^k = u_i^j \cdot (\gamma_- + \gamma_+) - \left(u_{i-1}^j \cdot \gamma_- + u_{i+1}^j \cdot \gamma_+ \right) \ , \tag{4.15}$$

In the following the explicit and implicit solving schemes are presented for two boundary conditions: the Neumann and the Dirichlet boundary condition (bc). The Neumann bc is defined with a fixed value for u. The Dirichlet boundary condition is described by the gradient of u, which is proportional to the heat flux according to (4.6).

Explicit finite difference scheme - Forward Euler

To obtain the explicit expression from (4.15), the quantities at time index $k + 1$ are solely expressed by the known parameters at time step k [Pre+07]. Setting $j = k$ in (4.15) we find

$$u_i^{k+1} = u_i^k \cdot (\gamma_- + \gamma_+ + 1) - \left(u_{i-1}^k \cdot \gamma_- + u_{i+1}^k \cdot \gamma_+ \right) \tag{4.16}$$

where the single u_i^k was brought from the left to the right hand side to separate the known and unknown quantities.
The Neumann boundary condition (bc) for the given boundary value $u_{1,BC}$ for the cell $i = 1$ – typically at a domain boundary – reduces the equation to

$$u_1^{k+1} \equiv u_{1,BC} \ . \tag{4.17}$$

The Dirichlet bc for the gradient to the left at index i in equation (4.16) yields the expression

$$u_i^{k+1} = u_i^k + \gamma_+ \cdot (u_{i+1}^k - u_i^k) + q_{BC} \cdot \frac{\Delta t}{h_i} \tag{4.18}$$

$$= u_i^k (1 - \gamma_+) + u_{i+1}^k \cdot \gamma_+ + q_{BC} \cdot \frac{\Delta t}{h_i} \ . \tag{4.19}$$

Stability analysis shows [CFL28], that the discrete step sizes in time and space cannot be chosen arbitrarily. The explicit solver is only stable for a Courant-Friedrich-Lewey (CFL) number below a stable limit $C_{\max}$. For the diffusion equation discretised above we find

$$C = \chi \frac{\Delta t}{(\Delta y)^2} \leqslant 0.5 \,. \tag{4.20}$$

With the sample interval of the diagnostic as Δt the spatial resolution is limited to

$$\Delta y \geqslant \sqrt{2 \cdot \chi_{\max} \cdot \Delta t_{\max}} \,. \tag{4.21}$$

For a non-constant diffusivity and time step, the maximal values are relevant. In case of a temperature dependent diffusivity and variations of the sample rate this limits the resolution capabilities for other local CFL numbers. In higher dimensions the effective CFL value is the sum of the CFL values along the main axes. This lowers the stable spatial resolution for the explicit scheme in higher dimensions compared to this 1D derivation.

Implicit finite difference scheme - Backward Euler

An alternative to the explicit formulation shown above is to express the rate of change – r.h.s. in (4.15) – in terms of the target time index $j = k + 1$, [Pre+07]. The result is a linear system, which has to be inverted. Thus, this method can be significantly more computationally expensive, depending on the geometry and dimension of the problem. However, in one dimension the system is tridiagonal, for which efficient solvers exist.

For the Neumann boundary condition two options are viable. The first is to include the boundary in the tridiagonal system via a pseudo-cell outside the computational domain. The second is to use an explicit integration step, increasing the heat potential of the boundary cell according to the heat influx at the outer interface.

By adding another entry in the solver for $i = 0$, we obtain the linear system

$$\begin{pmatrix} u_0^k \\ u_1^k \\ u_2^k \\ \vdots \\ u_{M-1}^k \\ u_M^k \end{pmatrix} = \begin{pmatrix} 1-\gamma_- & \gamma_C & -\gamma_- & 0 & 0 & \dots & 0 \\ -\gamma_- & 1+\gamma_+ + & -\gamma_+ & 0 & \dots & 0 \\ 0 & -\gamma_- & 1+\gamma_C + & -\gamma_+ & 0 & 0 \\ 0 & \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & 0 & 0 & -\gamma_- & 1+\gamma_C & -\gamma_+ \\ 0 & 0 & \dots & 0 & -\gamma_- & 1+\gamma_- \end{pmatrix} \cdot \begin{pmatrix} u_0^{k+1} \\ u_1^{k+1} \\ u_2^{k+1} \\ \vdots \\ u_{M-1}^{k+1} \\ u_M^{k+1} \end{pmatrix}$$

, with $M+1$ elements. By using some linear algebra – subtracting the second line from the first – we find

$$\begin{pmatrix} u_0^k - u_1^k \\ u_1^k \\ u_2^k \\ \vdots \\ u_{M-1}^k \\ u_M^k \end{pmatrix} = \begin{pmatrix} 1 & -1 & 0 & 0 & \dots & 0 \\ -\gamma_- & 1+\gamma_+ + & -\gamma_+ & 0 & \dots & 0 \\ 0 & -\gamma_- & 1+\gamma_C + & -\gamma_+ & 0 & 0 \\ 0 & \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & 0 & 0 & -\gamma_- & 1+\gamma_C & -\gamma_+ \\ 0 & 0 & \dots & 0 & -\gamma_- & 1+\gamma_- \end{pmatrix} \cdot \begin{pmatrix} u_0^{k+1} - u_1^{k+1} \\ u_1^{k+1} \\ u_2^{k+1} \\ \vdots \\ u_{M-1}^{k+1} \\ u_M^{k+1} \end{pmatrix} .$$

The boundary condition is included via $u_0^k - u_1^k = q_{BC} \cdot \Delta_-$.

For including the heat explicitly in the first layer we find

$$\begin{pmatrix} u_1^k + q_{BC}\dfrac{\Delta t}{h_1} \\ u_2^k \\ \vdots \\ u_{M-1}^k \\ u_M^k \end{pmatrix} = \begin{pmatrix} 1+\gamma_+ + & -\gamma_+ & 0 & \dots & 0 \\ -\gamma_- & 1+\gamma_C + & -\gamma_+ & 0 & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & 0 & -\gamma_- & 1+\gamma_C & -\gamma_+ \\ 0 & \dots & 0 & -\gamma_- & 1+\gamma_- \end{pmatrix} \cdot \begin{pmatrix} u_1^{k+1/2} \\ u_2^{k+1/2} \\ \vdots \\ u_{M-1}^{k+1/2} \\ u_M^{k+1/2} \end{pmatrix}$$

which is actually equivalent to including the boundary condition via an additional entry. However, the explicit *implantation* of heat in the surface layer allows to keep the structure of the solver with M elements, independent of the type of boundary condition.

The Dirichlet boundary condition is as simply included as for the explicit scheme, (4.17).

4.3.2. Crank-Nicolson

The Crank-Nicolson (CN) method [CN96] mixes the explicit and implicit solving schemes [Pre+07, sec. 20.2]. A solving step in time for step width Δt is split into a half-step $\Delta t/2$ with the explicit solver, followed by another half-step with the implicit solver. The resulting CN-scheme which is stable, but not a-stable. Oscillations can occur, but are damped in time. Such oscillations can occur from jumps in the boundary condition. One CN step step in time is as costly as one explicit plus one implicit solving step. A benefit is that CN is a second-order method in time [Pre+07, p. 1038], so the accuracy in time increases with Δt^2, in contrast to the linear scaling for the explicit and linear scheme.

4.3.3. Solver comparison

The different solving schemes – explicit, implicit and Crank-Nicolson (CN) – give different results for the same boundary conditions. To illustrate this, all three are compared for different CFL numbers in a one-dimensional test cast. The analytic equation (4.1) is used to obtain the reference temperature for a material with constant heat conductivity and diffusivity, to which a heat flux density of $1\,\mathrm{MW\,m^{-2}}$ is applied. The surface temperature is evaluated in steps of $1\,\mathrm{ms}$, corresponding to $1\,\mathrm{kHz}$ sample rate. This corresponds to an ideal surface temperature measurement, which is then used as input for THEODOR. Figure 4.1 shows the heat load obtained from the three methods, for a range of CFL numbers. On the left the explicit scheme, which takes four iterations to reach 95% of the reference heat load with the ideal CFL$=0.5$. For CFL$=0.25$ it takes 9 iterations to reach 95%. The degree to which the gradient and the surface in the tile can be resolved, is limited by the spatial discretisation. This is further discussed in the next section.

In the middle figure are the results for the Crank-Nicolson (CN) method are shown. At the beginning of the test the boundary condition changes from $q = 0$ to $q = 1\,\mathrm{MW\,m^{-2}}$, resulting in oscillations for CFL> 1. This second order method in time improves the results in this

specific problem only for low CFL numbers.
In the right figure the results of the implicit solver are shown. For CFL$=3$ the reference is reached basically instantaneously, with an overshoot for higher spatial resolutions. This is not to be confused with the oscillations appearing for the explicit or CN method. Reason is the method at hand of setting the target surface temperature as boundary condition, with which the HDE is then solved.

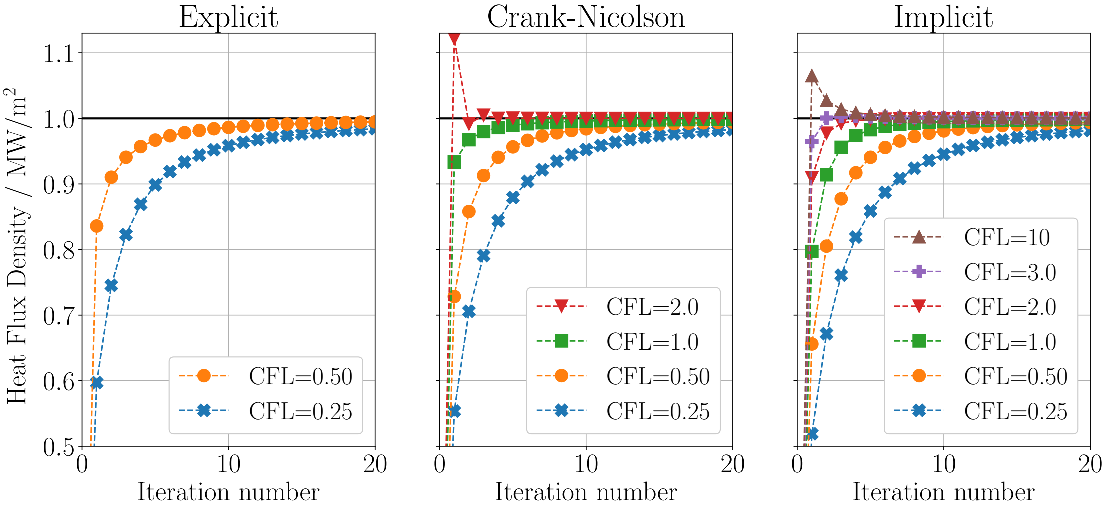


Figure 4.1.: Comparison of the explicit (left) Crank-Nicolson (middle) and implicit (right) solvers schemes for a set of CFL numbers. The reference of 1 MW m^{-2} is marked as black line. The three-layer evaluation method is used for all schemes.

Conclusions on the solving schemes

The implicit solver is considered best for the problem at hand, while the Crank-Nicolson scheme can be preferable in other situations, where higher precision is desired and boundary conditions are well behaved. For more complex geometries finite element or finite volume methods are suggested.

Extension to higher dimensions

Using operator splitting [Pre+07, sec. 20.3.2] the implicit solving scheme is extended efficiently to two or three dimensions. The spatial differential operator is split and solved for sequentially. The computation time of the underlying Thomas-algorithm scales linear with the number of elements N in the discrete system. The operator splitting leads for k dimensions to k sequential solver steps for all elements. Overall the computation time scales with $O(N \cdot k)$. However, the precision of the classic THEODOR suffers from this splitting.
Solving the HDE in 3D increases the problems inherent in the classic 2D THEODOR, which will be discussed in the next section. The use of a statistical analysis is suggested to circumvent

this problem. An example for this using the 2D adaptive kernel (AK) model is given in chapter 10.

Mixing solving schemes

For the introduced operator splitting method, the spatial second derivative operator is split into the second derivatives for different orientations. It is possible to choose different solving schemes for different orientations, depending on what the best scheme is for the given case. E.g. a pure implicit solver for orientations with high resolution, being unconditionally stable, and the Crank-Nicolson scheme for orientations, for which the CFL condition is fulfilled. Throughout this thesis only the implicit method is used.

4.3.4. Deducing surface quantities

Obtaining the heat flux given the heat potential field

The heat flux density q from the surface into the tile can be determined using the heat potential gradient in the tile. In the most simple case this is done using the two top-most values from the surface at y_0 and the first layer below at y_1, at lateral position x_i :

$$q(x_i) \approx \frac{u(x_i, y_0) - u(x_i, y_1)}{y_1 - y_0} \tag{4.22}$$

A more robust method is to use the first derivative including the first three layers with distances $d_1 = x_1 - x_0$ and $d_2 = x_2 - x_0$:

$$q(x_i) \approx \frac{u(x_0) \cdot \left(1 - \frac{d_2^2}{d_1^2}\right) + u(x_1) \cdot \frac{d_2^2}{d_1^2} - u(x_2)}{\frac{d_2^2}{d_1^2} - \frac{d_2}{d_1}} \,. \tag{4.23}$$

This method corrects for the curvature of the profile and allows a more precise heat load determination. However, diffusion perpendicular to the surface distorts the depths profile. This three-layer method is used in the evaluations of figure 4.1. Applying the simple two-layer method leads to figure 4.2. Even for high CFL numbers in the implicit solver, the reference heat load cannot reliably be reproduced.

Obtaining the surface temperature given the heat potential field

For the forward modelling not the deduced heat load, but the precision of the modelled surface temperature for a given surface heat load is relevant. Similar to the test before, a material with constant parameters is used and a heat load of $1\,\mathrm{MW\,m^{-2}}$ is applied. Reference is the same analytic surface temperature trace as before.

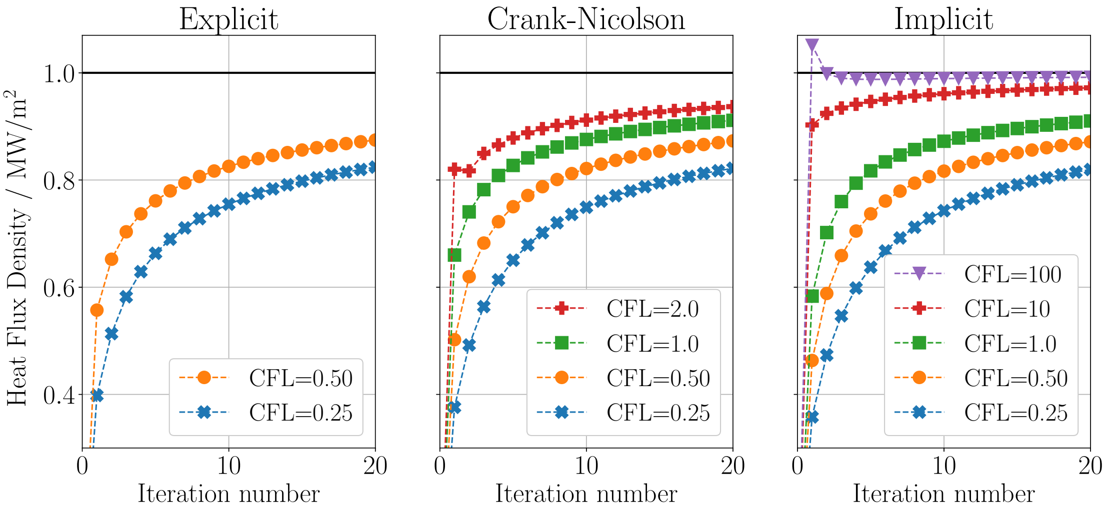


Figure 4.2.: Comparison of the explicit (left) Crank-Nicolson (middle) and implicit (right) solvers schemes for a set of CFL numbers. The reference of $1\,\mathrm{MW\,m^{-2}}$ is marked as black line. The two-layer evaluation method is used for all schemes.

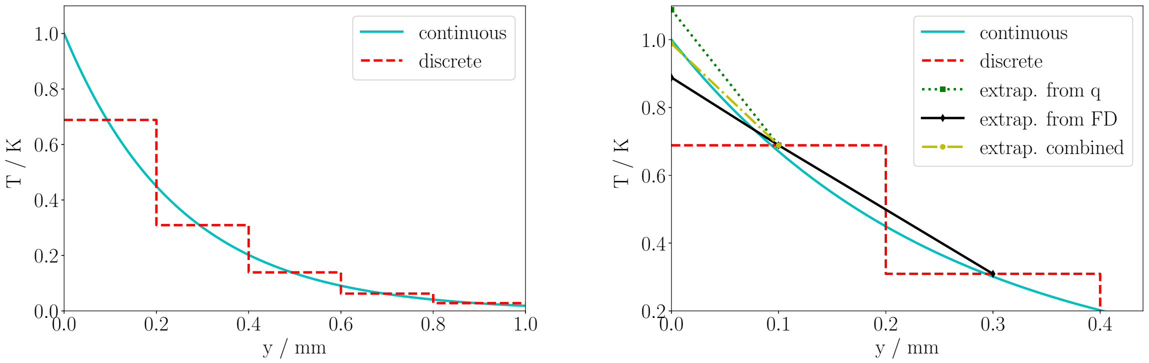
The easiest way to determine the surface temperature – which enters the likelihood in the statistical evaluation – is to interpret the value in the top layer of the grid as surface temperature. However, the finite discretisation results in spatial averaged values for the temperature. As a temperature gradient is to be expected, unless there is no heat exchange at the boundary, this leads to an underestimation of the surface temperature for impinging heat. This is illustrated in figure 4.3(a), with the discrete levels determined as mean of the continuous function in the corresponding interval. The blue solid line shows the continuous scalar field, the red, dashed line the discrete representation. Basis is an artificial profile defined as

$$T(y) = \exp(y/0.25\,\mathrm{mm}) \;. \tag{4.24}$$

An alternative is to extrapolate from the cell centre of the surface cell to the *actual* surface position, using gradient information. Similarly to equations (4.22) and (4.23) this gradient can be determined from the discrete gradient into the tile.

An improvement in case of the heat flux density given as boundary condition is to use the corresponding gradient for the extrapolation. This is illustrated in figure 4.3(b). The abscissa represents the distance to the surface – denoted y. The ordinate represents the temperature – or analogous heat potential. Starting at the centre of the top layer and extrapolating with the known gradient yields the dotted, green line denoted *extrap. from q*. The black, solid line is an extrapolation from the two top most-layers. The yellow, dash-dotted line is the mean of the two extrapolation methods.

The extrapolation relies on the first layer being thin enough, to justify the assumption of a constant diffusivity and conductivity as well as a constant gradient.



(a) Temperature profile over four decay lengths of the exponential and five discrete layers.

(b) Close up of the left figure, showing three extrapolation options: from boundary condition (q), from finite difference (FD) and combined.

Figure 4.3.: Example for a temperature (or analogous heat potential) distribution from the surface at $y = 0$ into the depth of the tile. The cyan, solid line shows the continuous function. The discrete approximation is shown as red, dashed line. a) shows on overview of the function over five decay lengths of the exponential. b) shows the interval close to $y = 0$ with three extrapolation schemes.

The best result in this test case is obtained by averaging the finite difference gradient and the boundary condition to extrapolate from the top cell value to the surface. This however works only for constant heat loads in between samples and evaluation of the final state. Ideally the discretisation is sufficiently fine to resolve the surface temperature within the measurement uncertainty. Only the implicit solver allows a sufficiently fine grid at the surface without spurious oscillations, so that no extrapolation techniques have to be used.

Figure 4.4 shows a comparison for the inferred surface temperature from the implicit solver using 10 sub-steps with CFL=1 (left), CFL=100 (middle, a tenth of the layer thickness) and CFL=10000 (right, a hundredth of the layer thickness). With increasing resolution the results for all methods improve and converge. The implicit method without extrapolation is therefore suggested and used in this thesis.

The depths resolution does not have to be chosen uniform, as with increasing distance to the surface the heat potential distribution is smoothed by the diffusion operator. The best choice of the discretisation depends on the tile dimensions and sample rate. However, a good rule of thumb – based on a numerical study optimising RMS of the temperature deviation – suggests an exponential increase in the layer thickness with an exponent of $1.05 \cdot i$ for the i'th layer.

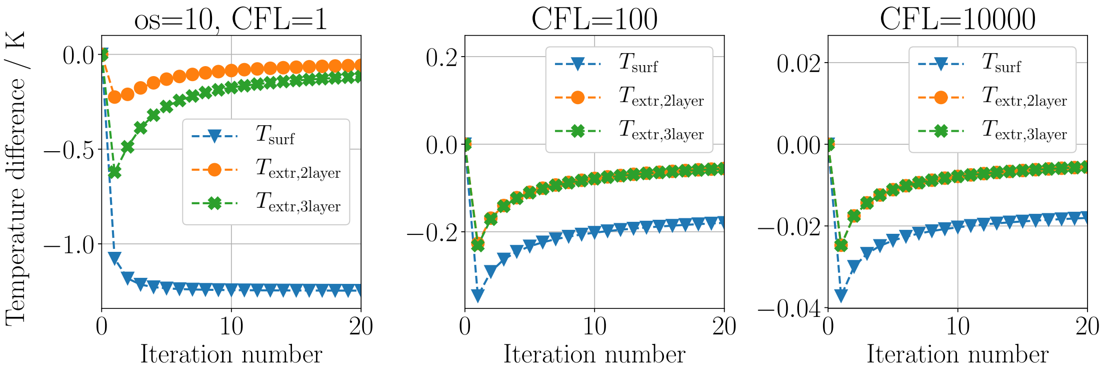


Figure 4.4.: Comparison of deduced surface temperature with increasing CFL number. Implicit solver with 10 times oversampling. CFL= 1 (left), CFL= 100 (middle), CFL= 10000 (right).

4.4. THEODOR from a Bayesian point of view

From a statistical perspective, THEODOR delivers the wrong answer to the question about the surface heat flux density. However, given some conditions and knowing about when it fails, the use of this tool can be justified for a fast analysis, when needed. This section compares the classic and the Bayesian THEODOR.

4.4.1. What heat flux is obtained?

THEODOR uses a time trace of temperature profiles as input and returns a time trace of heat flux densities. The relevant question of the user usually is:

What is the heat flux distribution impinging on the surface of the tile, given observations about the surface temperature?

The result obtained from THEODOR answers a different *question*: what is the distribution of the diffusive heat transport perpendicular from the surface of the tile into the bulk of the tile, given a new temperature profile on the surface as a boundary condition?

The difference is, that the heat flux obtained from THEODOR contains a low-pass filter, intrinsic to the sequence of steps: the boundary condition is set, the HDE is solved and then from the first two or three layers the gradient is determined, which is proportional to the heat flux density.

When using inference and forward modelling, the heat flux onto the tile actually is the boundary condition and matching the model result to the measurement is the goal.

4.4.2. Which information is used and how?

1) Which information is available?

The IR system integrates a photon flux from the tile surface in time. This can be translated into a surface temperature using Planck's law. This is however only valid, when the temperature is constant during the integration of the system. Otherwise the photon flux is averaged, leading to a sort of mean temperature, depending on the functional dependence between photon flux and temperature. A normal distributed uncertainty in the photon flux does not in general translate to a normal distributed uncertainty in the temperature. In addition, when the sample rate or integration time chances, it is not trivial to say, when the deduced temperature should be used as boundary condition.

2) How is the information used?

THEODOR uses the mentioned evaluated temperature as input, and uses it as boundary condition at the surface of the tile. The temperature is assumed to represent the temperature at a specific point in time and the heat flux is interpreted as constant between samples. A change in sample rate or integration time, significant temperature changes during integration and non-constant heat loads cannot be addressed. Using temporal oversampling is also not trivial, as assumptions about the time evolution of the boundary conditions have to be made.

THEODOR sets the determined temperature as boundary condition on the surface. Based on this boundary condition the HDE is solved. However, the obtained signal actually is the result of heat impinging onto the tile and heat diffusion in the tile. In short: the diffusion operator is applied twice.

Starting at a tile in thermal equilibrium – with a constant temperature distribution - a delta peak in the heat flux in time and space will result in a Gaussian temperature profile on the surface, with the width corresponding to the diffusion time and coefficient. The initial delta-peak in the surface temperature cannot be observed with a finite integration time. THEODOR uses this broadened temperature profile as boundary condition for solving the HDE, which introduces further broadening. From the resulting temperature field in the tile the heat flux is deduced. This leads to artificial broadening.

For the forward modelling the characteristic of the diagnostic is taken into account, and the measurement process can be modelled properly. The model output is an integration of the photon flux from the surface and captures any non-linearity. When the heat load cannot be assumed constant between samples, this can be modelled. In short: as long as the system can be described accurately, there are no limitations as for the classic interpretation.

In the statistical approach the heat flux profile best describing the data is inferred. Ambiguities due to the smoothing property of the diffusion operator are resolved by regularisation, as far as necessary. If the data quality allows to infer the before mentioned delta-peak, it is resolved. In addition, uncertainties can be quantified.

3) Can the classic THEODOR be used anyway?

The criticism about the wrong interpretation of the deduced temperature can be overcome by using systems with integration times short compared to the time interval between samples and also short enough to justify t he constant temperature assumption. However, reducing the integration time reduces the signal-to-noise ratio (SNR). Not using a variable integration time in order to optimise the SNR is an additional detrimental consequence.
The criticism regarding the double application of the diffusion operator becomes more important, as the ratio of the spatial resolution to the sample rate increases. For sufficient sample rates $1/\Delta t$ the perpendicular diffusion length x_D is negligible compared to the spatial resolution Δx of the measurement system:

$$x_D = \sqrt{\chi \cdot \Delta t} \gg \Delta x \ . \tag{4.25}$$

For usual values of the diffusivity of $\chi \approx 6 \cdot 10^{-5}\,\mathrm{m^2\,s^{-1}}$ – see section 4.2 – for tungsten at 1 kHz sample rate ($\Delta t = 1 \cdot 10^{-3}\,\mathrm{s}$) the diffusion length is

$$x_{D,\mathrm{AUG}} = \sqrt{\chi \ \cdot \ \Delta t} \approx 0.245\,\mathrm{mm} \ . \tag{4.26}$$

At ASDEX Upgrade the current system has a spatial resolution of about 0.7 mm, which is about three times the diffusion length. The diffusion length decreases with the square root of the sample rate. For 10 kHz it is about 0.07 mm, for 50 Hz it is 1 mm.
When the explicit method is stable, the spatial resolution is less than the diffusive broadening between two samples. Using faster systems or lower resolutions mitigates this detrimental effect.

4.4.3. Conclusions

The *classic* approach, using measured data as boundary condition, delivers a fast estimation of the surface heat flux, but is logically different from solving the inverse problem at hand. To quantify short events as well as features on the level of the SNR inference is necessary to obtain the actual heat load profile to the surface with an uncertainty estimation.

4.4.4. Lateral discretisation

Another topic is the lateral discretisation and its effect on the measurement. Typical IR systems have spatial resolutions and sample times which are stable with regard to the CFL condition, section 4.3.1. This surface resolution is a reasonable choice for the lateral discretisation. However, for thermography the gradient in the surface temperature has to be considered. A strong gradient together with the non-linearity of Planck's law leads to a systematic overestimation of the mean temperature within a cell. For the presented work, the IR camera resolution is sufficient to keep this error within the bounds of the measurement uncertainty. In this regard the tools are equivalent at the moment, while only forward modelling allows a higher spatial resolution if the approximation of constant temperature observed by one camera pixel does not hold.

4.5. Toroidal heat transport

In general the 2D surface heat flux distribution has to be taken into account. For tokamaks the assumption of toroidal symmetry – for the plasma and the tile it is in contact – can break for various reasons:

- Polygon shaped tile: The target tiles are often planar and therefore form a polygon structure along the toroidal axis. Therefore the angle between the magnetic field and the tile is a function of the toroidal position.
- Mounting uncertainties: the orientation of the tiles inhibits uncertainties in mounting and due to movement during operation. Thermomechanical stress can lead to curvature of the surface [JHG14].
- Shallow angles: The incidence angle is typically on the order of $\simeq 3$ degree, due to the much stronger toroidal than poloidal field, especially near the X-point and therefore divertor target. This results in a high sensitivity of the projected heat flux on the incidence angle.
- Shadowing: not the entire surface receives heat loads, due to the toroidally wetted fraction – area receiving heat load over total surface area – is smaller than one [Rap+10]. A part of the tile is shadowed by its neighbouring tile to prevent damage on the edges.

In addition, other plasma experiments like Stellarators or linear devices show no exploitable symmetry. Numerical tools for the general 2D case exist [Kan+16], but only use a deterministic calculation like THEODOR.

4.6. Other tools

The problems of inferring surface heat loads discussed in this section do not only hold for THEODOR, but also other numerical tools. E.g. the Augmented Lagrangian Implicit Constraint Inverse Analysis tool *ALICIA* [Igl+17], which uses an iterative scheme to determine the given surface temperature as close as possible. The allowed temperature mismatch is a manually chosen value and the result does not necessarily reflect the true incoming heat-load. From the cited paper: ”The augmented Lagrangian scheme detailed in [...] adds a loop to the numerical procedure reducing the temperature difference until it is very small, $\Delta T < \epsilon$. This imposes the constraint without modifying the power balance, therefore increasing the accuracy.” The last sentence focuses on numerical accuracy, which is only part of the story.

For some studies [Igl+18] ”ALICIA applies augmented Lagrangian and implicit integration schemes to bound the error to 5 °C on the surface of the lamella.” This can be seen as a flat probability distribution for the likelihood, where the widths is not connected to the uncertainty of the observation.

Dedicating more numerical resources to solving the PDE can be viable for some studies, e.g. regarding thermomechanical stress or erosion, but will not lead to a good answer for the inverse problem. ANSYS is another well known tool for FEM analysis, but suffers from the same drawbacks when using measured – therefore uncertain – values as exact boundary condition.

The typical numerical error of the presented implicit solver with 10 sub-steps in time and the refined grid near the surface is on the order of $1\,\mathrm{mK}$ for fluctuations of $1\,\mathrm{MW\,m^{-2}}$. Even for the highest resolved discharges at AUG with $25\,\mathrm{mK}$ uncertainty this error is considered negligible. The solving scheme should be chosen with the data quality in mind.

5. Inference Methods

Given a model for the quantity of interest and the forward model, introduced in chapters 6 and 4, the aim is to obtain information from data. For this the parameter space is investigated and for practical applications a *best* solution is to be found. A typical method is optimisation, looking for the maximum a posterior (MAP) solution – the parameter set with the highest probability density. A second order expansion in negative log-space around the MAP point corresponds to a Gaussian approximation of the posterior, with its standard deviation as a quantifier for the uncertainty of the solution.

For non-Gaussian and especially multi-modal distributions the MAP solution can be meaningless, in which case the posterior distribution needs to be explored otherwise, e.g. using Monte Carlo methods. The trace of a Markov Chain can be used to inspect marginal distributions of the parameters, to reveal multi-modality and decide which parameters are *best* for the question asked.

5.1. Monte Carlo

5.1.1. Metropolis Hastings Monte Carlo

The probably most used numerical integration method for distributions in 5 or more dimensions is Markov chain Monte Carlo (MCMC) [J G11]. One of its drawbacks is that the underlying random walk is not efficient in high dimensions. The correlation between subsequent points tends to be significant, so that many function evaluation are needed. Also fine tuning of the step width is not trivial, but affects the efficiency of the method via the resulting acceptance rate – the relative amount of accepted steps [J G11]. A usual value for high dimensional problems is an acceptance rate of ≈ 0.25 [Béd08]. Though there are more advanced variants to improve its performance, the simplicity of the standard implementation allows for reliable exploration to check the results of other methods.

5.1.2. Hamiltonian MC

An alternative way to perform MC integration is to view the model parameters as position of a particle in a potential [Nea93]. The potential itself is the posterior evaluated for those parameters. To traverse the parameter space a momentum component for every parameter is introduced [BS11]. At the beginning of each trajectory the velocity distribution is drawn randomly from a suited distribution. Several steps are made to traverse the Hamiltonian space and obtain a new sample.

To obtain meaningful samples the integrator has to be time-reversible and preserve phase volume, [Bet+17; Bar+18], which is achieved using a symplectic integrator.

Beneficial is that instead of a random walk, like in MCMC, the exploration in phase space naturally incorporates gradient information, to direct the exploration. A number of steps along a trajectory on the surface of constant energy allows to move further than the same number of random walk steps.

Selecting the integration time step

The rejection of samples is only necessary in Hamiltonian MC because the discrete numerical integration does not follow the analytic trajectory exactly. Therefore the step width ϵ is chosen in order to make as large steps as possible, in order to traverse the function quickly, while keeping the error low.

One way to select ϵ is to make an ad hoc estimation, based on a few random momentum vectors. By using those vectors for a set of few integration steps – say 1 and 2 – the change in the Hamiltonian ΔH can be determined. The acceptance rate is $\min(1, H_{i+1}/H_i)$. Instead of drawing a sample and applying the Metropolis algorithm, the ratio of the Hamiltonian can be used directly as measure for the acceptance probability.

The general relationship between step size ϵ and solver precision (deviation ΔH) can be described by a power low with exponent β

$$\Delta H \propto \epsilon^{\beta} \ . \tag{5.1}$$

A set of random initial momentum vectors is used to determine the average ΔH for a value of ϵ. Iterative methods can be used to adept ϵ until the error ΔH is in a reasonable range. The warm up is started with this educated guess for ϵ.

This method can be extended – by comparing ΔH for different values of ϵ – to determine the effective power

$$\beta = \ln\left(\Delta H_{i+1}/\Delta H_i\right) / \left(\epsilon_1/\epsilon_0\right) \tag{5.2}$$

given the used solving scheme and geometry of the problem.

Any adaption of the step width is only used during the initial *warm-up* phase.

5.1.3. Implementation details

For both MC schemes the *hard* parameter limits – forming a hyperrectangle in parameter space – are treated as reflective boundaries. While this keeps detailed balance intact, the number of rejected (in this sense obviously *lost*) samples is reduced. For the Hamiltonian implementation, the preservation of the volume of the phase space can be shown [MD15] also for deflection from finite steps in the potential field.

5.2. Optimisation

For production work optimisation routines from NAG [NAG16] are used, searching for the mode of the posterior distribution. The C++ code uses the *adept* library [Hog14] to efficiently determine the Jacobian of the posterior with respect to all input parameters.

For this thesis the NAG methods e04kbc and e04wdc were both used and compared. They find the mode with a similar number of function and gradient evaluations with similar reliability. For high-dimensional problems – more than 500 parameters – the kbc function – using a quasi-Newton method – is found to require less memory. The wdc function proved more resilient against local minima.

5.3. Automatic differentiation

For optimisation as well as integration with Hamiltonian Monte Carlo gradient information is valuable. More specifically one may use the gradient of the posterior with respect to the input parameters. The most simple option is to use finite difference, which requires at least one additional function call per parameter. An alternative is to use the chain rule on the known derivatives of algebraic operations, to construct the analytic derivative. While the gradients can be implemented manually for many models, this becomes a tedious task, as the models become more complex. Also it increases the amount of necessary testing and maintaining, while reducing the reliability of code.

An alternative is given by automatic differentiation, which is available for C++ and other languages. By using a suitable library and *marking* the variables of interest, the compiler constructs the derivatives, respecting conditionals and branches in the code. After a forward evaluation – propagating the input parameters until the result is found – the library can perform the backtracking and returns the gradients as a vector or an array. Finding the Jacobian takes about 3-10 times the computation time of the simple function call. However, for many parameters the evaluation is faster, with the additional benefit of more precise derivatives. Also, no finite step width like for finite difference method, has to be chosen to find an optimum between numerical cutoff and approximation of the slope.

A drawback is the increased memory consumption, as most libraries create what is called a *tape* to store the path from input variables to the cost function.

In this project the library *adept* for C++ is used [Hog14].

5.4. Matrix determinant algorithms

In this section two schemes for determining the determinant of a matrix are presented. Namely the singular value decomposition (SVD) and the stochastic trace estimation (STE). An application is to evaluate the integral of a distribution, when using the saddle point approximation [LDT14, p. 352]. This will be used for the adaptive kernel (AK) model, see section 6.1.2.

5.4.1. Singular value decomposition (SVD)

The integral of a Gaussian distribution can be expressed in terms of the determinant of the representing Hessian matrix, which is equivalent to the product of the Hessians eigenvalues. However, for numerical reasons it is beneficial to use the singular value decomposition (SVD) instead. Typically the algorithms are faster and less sensitive to (close-to) degenerated eigenvalues.

Let $\mathbf{X}$ denote a real valued $M \times N$ matrix with $M \geqslant N$ with rank $r \leqslant N$. The condition $M \geqslant N$ is without loss of generality, as otherwise the transpose of the matrix can be considered. The SVD is expressed in terms of three matrices:

$$\mathbf{X} = \mathbf{U}\mathbf{S}\mathbf{V}^T \ . \tag{5.3}$$

$\mathbf{U}$ is an $M \times M$ matrix, storing the left singular vectors column wise. The $N \times N$ Matrix $\mathbf{V}^T$ keeps the right singular vectors row wise. The $N \times M$ matrix $\mathbf{S}$ keeps the singular values on the diagonal, while all other elements are zero.

For a square, symmetric matrix, the SVD is equivalent to the eigenvalue decomposition. For a square, but not symmetric matrix the singular values correspond to the eigenvalues of the squared matrix $\mathbf{X}^T\mathbf{X}$.

The singular values are of interest in this work because the determinant of a square $N \times N$ matrix $\mathbf{A}$ is equivalent to the product of all its eigenvalues λ_i:

$$\det(\mathbf{A}) = \prod_{i=1}^{N} \lambda_i \ . \tag{5.4}$$

The SVD is a reliable method to obtain this value, especially when a lower threshold to vanishing eigenvalues is desired. With it the logarithm of the determinant is obtained by the sum

$$\ln(\det(\mathbf{A})) = \sum_{i=1}^{N} \ln(\lambda_i) \ . \tag{5.5}$$

The logarithm of the determinant is of interest, as numerical exploration tools for the posterior work on the negative logarithm of the probability.

5.4.2. Stochastic trace estimation (STE)

An alternative to the SVD is the stochastic trace estimation (STE), which is used to find an estimation of the logarithm of the determinant (log-det). This is beneficial as the cost of the SVD increases with about N^3 and the memory demand also increases strongly with N, especially when automatic differentiation is used.

The background is, that the log-det can be expressed as the trace of the matrix logarithm

$$\ln(\det(\mathbf{A})) = \operatorname{tr}(\ln(\mathbf{A})) \ . \tag{5.6}$$

Typically the matrix-logarithm is not a trivial quantity to obtain, but it can be approximated. Using Chebyshev polynomials $\mathbf{T}_k$ up to degree n we find

$$\ln(\mathbf{A}) \approx \sum_{k=0}^{n} c_k \mathbf{T}_k(x) \tag{5.7}$$

with coefficients c_k weighting the k'th polynomial $\mathbf{T}_k$.

In order to estimate the trace, a set of vectors $\vec{v}_i$ is sampled from a distribution such that the expectation value is

$$E(\vec{v}^T \mathbf{M} \vec{v}) = \operatorname{tr}(\mathbf{M}) \ . \tag{5.8}$$

An example is a element-wise normal distribution with $E(\vec{v}_i) = 0$, $E(\vec{v}_i^2) = 1$. Alternatives are the Hutchinson estimator [Hut90] with elements having a 50/50 chance to be +1 or −1, or mutually unbiased bases [Fit+18] inspired from quantum mechanics.

An estimate of the trace is then obtained via applying m such vectors $\vec{v}^{(j)}$:

$$\operatorname{tr}(\mathbf{A}) \approx \frac{1}{m} \sum_{j=1}^{m} \vec{v}^{(j)T} \mathbf{A} \vec{v}^{(j)} \ . \tag{5.9}$$

A direct approach to the log-determinant is then

$$\operatorname{tr}(\ln(\mathbf{A})) \approx \frac{1}{m} \sum_{j=1}^{m} \sum_{k=1}^{n} c_k \vec{v}^{(j)T} \mathbf{T}_k \vec{v}^{(j)} \ . \tag{5.10}$$

An introduction and examples can be found in [Han+16].

Probabilistic methods to obtain the log-det are available, e.g. Skilling [Ski89b] and Fitzsimons [Fit+17], giving access to an uncertainty estimation. In these probabilistic cases obtaining the eigenvalues is seen as inverse problem searching for $p(\lambda)$. However, for an optimisation depending on the log-det, the direct approach does work. A benefit of this direct STE compared to iterative schemes like in [Ski89b; Fit+17] or Conjugate Gradient methods like [Ski09] is that the method produces a smooth output.

5.5. Considerations regarding probability densities

Two examples are used to show how the choice of the parameterisation affects the MAP result and why multimodality is a problem for the MAP approach. More examples in more details can be found in statistics books like [Jay03; LDT14].

Parameter transformation

One difficulty is, that the probability density considered for the MAP depends on the chosen scale for a parameter. The mode can shift due to parameter transformations, while the median is conserved for a proper transform. Figure 5.1 A) shows a normal distribution – $N(0,1)$ on linear scale in x. Changing to logarithmic scale y:

$$f\colon \mathbb{R} \to \mathbb{R}^{+}, x \mapsto f(x) = exp(x) \tag{5.11}$$

with the relations of x, y and their differentials

$$y = \exp(x) \iff x = \ln(y) \tag{5.12}$$

$$\mathrm{d}y = \exp(x)\mathrm{d}x \iff \mathrm{d}x = \frac{\mathrm{d}y}{y}\ , \tag{5.13}$$

the PDF is transformed according to

$$p_y(y) = p_x(x) \cdot \left|\frac{dx}{dy}\right| \tag{5.14}$$

$$= p_x(f^{-1}(y)) \cdot \left|\frac{dx}{dy}\right| \tag{5.15}$$

$$= p_x(\ln(y)) \cdot 1/y\ . \tag{5.16}$$

These rules can be generalised to higher dimensions and ensure that the new density is normalised. This, however, changes not only the shape of the PDF, but also the position of the mode, as shown in 5.1 C). The peak of this log-normal function is analytically known to be at $\mu - \sigma^2$, so for the original $\mathcal{N}(0,1)$ it is at -1 in x-space. Note that the cumulative distribution functions (CDF) – shown on the right hand side of the PDFs – change their shape, but the median is translated according to the transformation: $f(0) = 1$.

When changing the variable, e.g. due to numerical considerations, without actually transforming the pdf, the MAP position is preserved, but the PDF is not in general normalised. This is shown in the lowest row 5.1 E) and F). The MAP position is translated as was the median position for the proper transformation: $f^{-1}(1) = 0$.

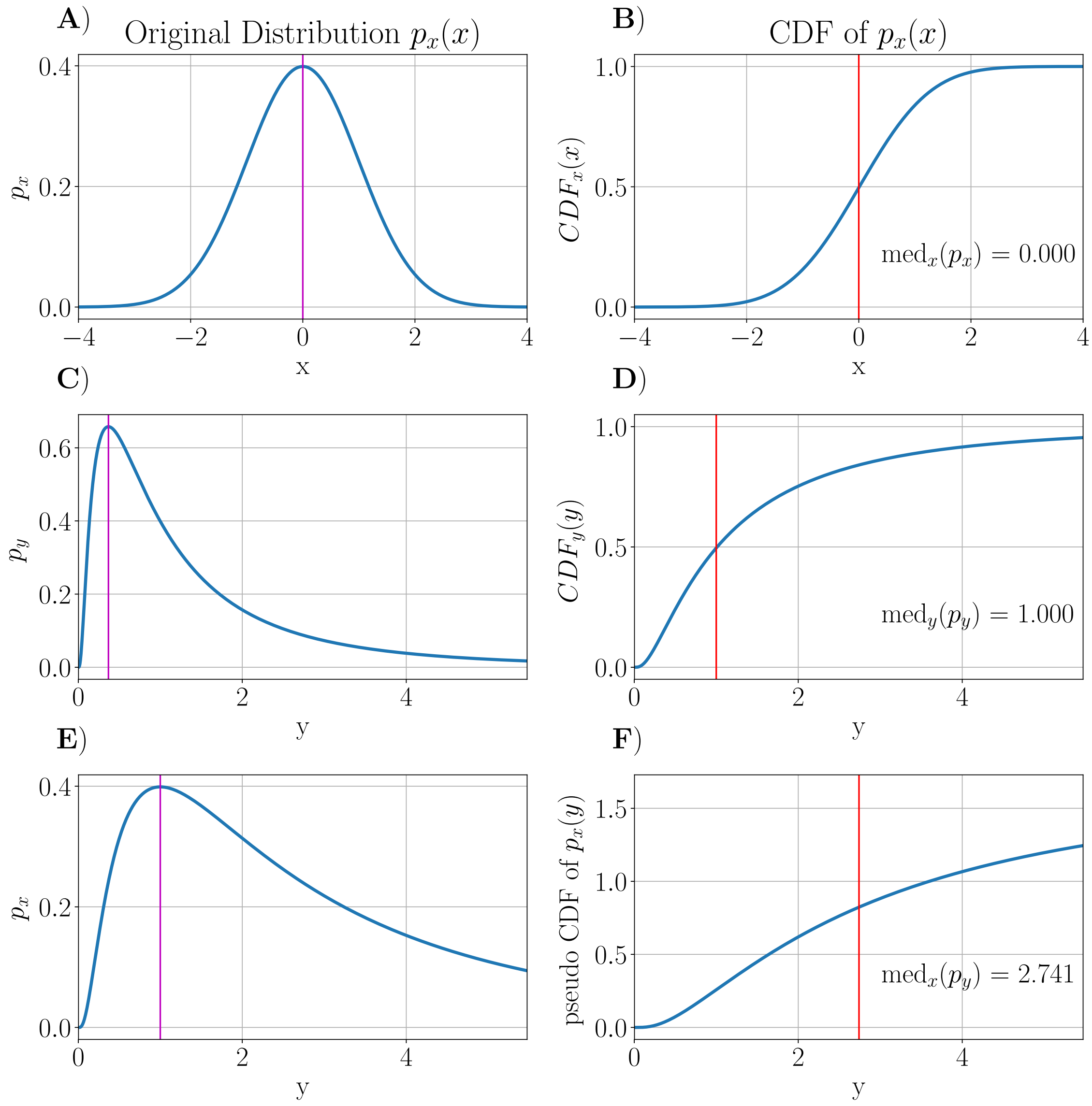


Figure 5.1.: In the left column three probability distributions are shown: A) a normal Distribution $\mathcal{N}(0,1)$ in x: $p_x(x)$. C) The normal distribution from A) transformed with $y = \exp(x)$, yielding the log-normal distribution $p_y(y)$. E) The distribution from A) p_x in y-space: $p_x(y)$. The right column shows the corresponding cumulative distributions. Note that the median position in D) is the mapped value from B): $1.000 = \exp(0.000)$ – meaning it is preserved. The same applies to the MAP position from A) to E), when the density is not adapted.

Multimodality and stability

As described before, the *best* model parameters can be quantified via the highest posterior density or the mode of the distribution. However, often the global maximum of the posterior density is not a good quantifier. For multimodal distributions a higher probability density indicates a *better* result for a specific parameter set. However, the generality or stability of the solution also depends on the width or the overall weight of the mode. Similar considerations apply to non-symmetric distributions.

Figure 5.2 A) shows a PDF composed from the sum of two normal distributions with different weights and widths. Although the right peak has an 11% higher peak density, the left mode is 10 times wider and carries 90% of the probability *weight*, as seen in graph B). For many cases, the left mode is the better choice, since it is more general in the sense, that a wider range of parameters describes the data well. For a given perturbation this corresponds to a higher stability of the solution. This is a core element of model selection, where different models are compared in order to find the most simple solution – Ockham's razor.

As in the section before, C) and D) show the PDF after the parameter transformation $y = \exp(x)$. Applying the transformation correctly – as in C) – the MAP changes from one mode to the other, while the median is unaffected.

Summary

Reparametrisation can effectively move the mode with respect to the initial space and change the relative height of different modes in the new space. The median is not affected by a proper transformation of the parameter space. For a MAP analysis, a simple replacement of a parameter by a differently scaled one will preserve the MAP position, but will offset any integration scheme or attempt to find the median.

In principal non-linear transforms can be used to introduce more or *compensate* existing modes. Similarly to how the p-value can be manipulated, the density is subject to the measure used. This MAP position actually is well defined, if measure theory is respected.

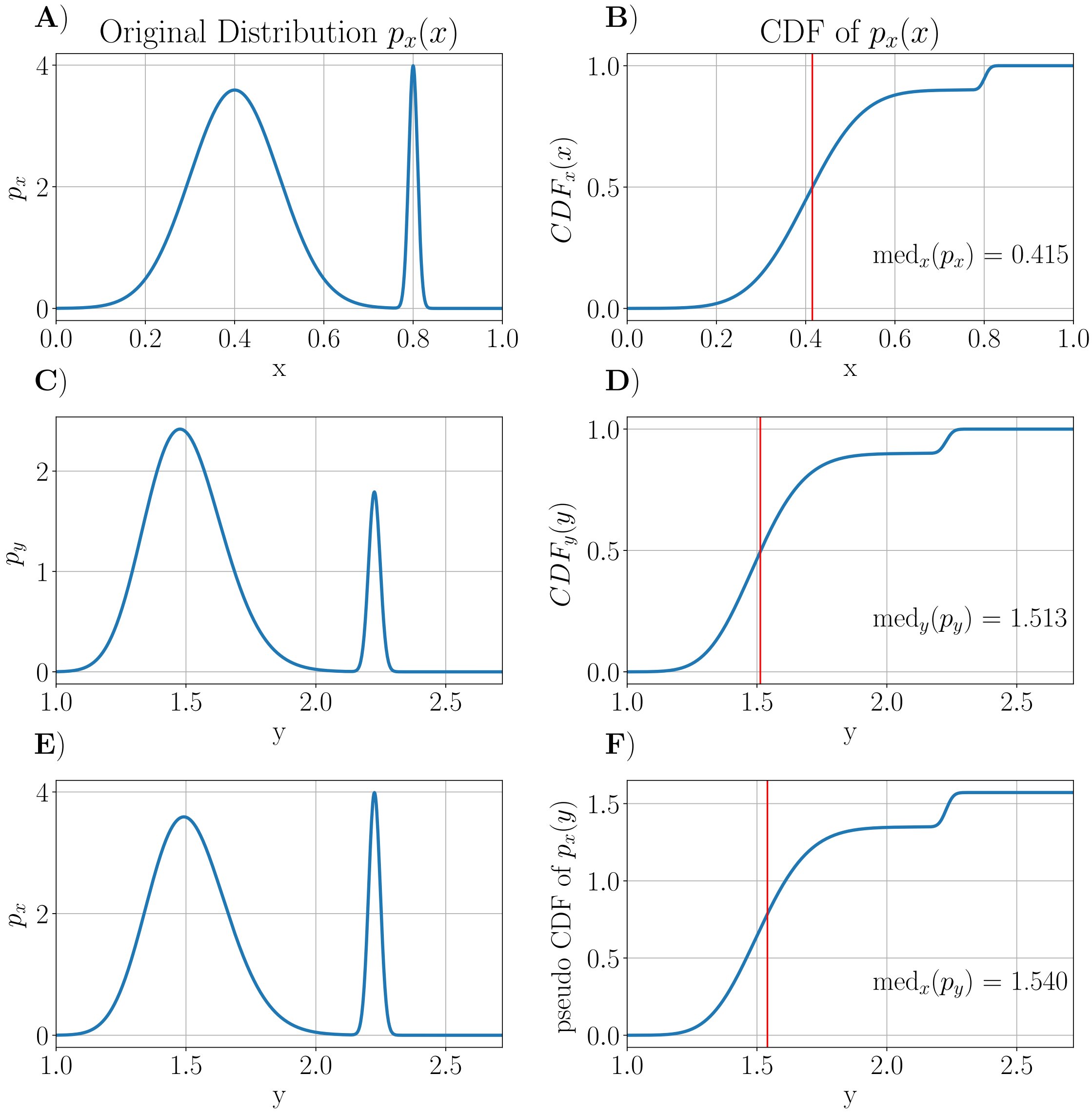


Figure 5.2.: A) a multi modal test distribution. C) The distribution from A) transformed with $y = \exp(x)$. E) The distribution from A) over the transformed space: $p_x(y)$. The right column shows the corresponding cumulative distribution functions. Note that the mode position from B) is moved according to the parameter transformation $\exp(0.4) \approx 1.5$, though the new CDF is not normalised. In contrast the modes from A) are shifted in C): $\exp(0.8) \approx 2.2$, $exp(0.4) \approx 1.5$.

5.6. Performance and scaling

AKs are a powerful tool, but the number of kernels as well as the number of evaluated cells play a major role in performance considerations. The used SVD scales typically with $O(N^3)$ with N resembling the number of rows and columns. For the AK model introduced in section 6.1.2 N corresponds to the number of kernel. In additional there are some sparse matrix operations used to form the precision matrix on which the SVD is applied.

For the BayTh evaluation on 1D profiles using the adaptive kernel model with the SVD the evaluation time is about $(100 - 500)$ ms for a profile width of about 200 pixel. The MAP is found for random start parameters after some 1000 iterations with the optimiser, at best after about 100 – e.g. when the preceding profile was very similar. This leads to an optimisation time of about 10 s to 5 min. In addition, about 30 s are spend to determine the Hessian and from it the covariance matrix. The total average for longer sequences is about 1 min. For a series of 1000 profiles the single-threaded program needs about 15 hours. Longer sequences can be separated with an overlap of about 20 profiles and combined, allowing for efficient parallelisation.

With classic MCMC the number of steps depends on the correlation of the parameters. For sequential updates, where one parameter is changed at a time as they are ordered, at least 10^4 sweeps are needed for a quantitative result. For 400 parameters this amounts to $4 \cdot 10^6$ evaluations. As the SVD is not evaluated and no gradients are used, the time for a single step is about 5 ms. This leads to about 5 hours of sampling time per time index.

With Hamiltonian Monte Carlo a step takes about 20 ms, and on average 50 are made for one sample. At least 1000 samples are needed to explore the kernel width distribution sufficiently, leading to a time of about 15 minutes per index. With 10^4 samples the exploration can be considered converged, after 1.5 hours.

6. Modelling

This chapter presents models for our quantity of interest – the surface heat flux density distribution. As discussed in chapter 5 inference on an ill-posed problem requires to formulate a model for the quantity of interest. This model is then propagated to the measured quantity using the forward model described in chapter 4. The first section presents the models used in this work for a single heat flux profile. The second section discusses how those models are used for the time dependent problem.

6.1. Representations for the quantity of interest

6.1.1. Form free model

Lets start with a simple model, from here on called form free model. A discrete profile m along a one dimensional axis x can be described by a number of amplitudes:

$$m(x_i) = h_i \; . \tag{6.1}$$

In the most simple case the profile shape is not restricted and there are no correlations in the model, therefore the term is form free. A reasonable prior for positive distributions is the entropic prior [Ski89a], which does not introduce correlation between the amplitudes. More details can be found in section 2.7.2. This combination will lead to a Maximum Entropy solution.

The form free model is treated as dummy model, to show the basic difference between the classic THEODOR and the probabilistic approach via forward modelling with the most simple model in terms of assumptions.

Remark: this model corresponds to the adaptive kernel (AK) model, introduced in the next section, with a set of Dirac delta kernels – or Gaussians with a fixed, very narrow standard deviation.

Several extensions can be though of, resulting in more or less common models. By including a penalty on the second derivative of neighbouring values, cubic splines are recovered – which minimise the second derivative. For each of these additions, new hyperparameters are introduced, for which a prior has to be specified and which has to be marginalised. For this work the simple formulation without correlation is used, as the work horse for the data analysis is the AK model.

6.1.2. Adaptive kernel

Basic idea

While the form free model – introduced in the section before – has no correlation in model space, inference approaches to ill-posed problems often rely on some mean of regularisation. On one hand, the model function should be able to represent all features supported by the data. On the other hand, the data often provide not enough information define all the degrees of freedom in such a model. In these cases a correlation in model space is desirable, to prevent over-fitting of the data. This can be achieved by treating the form free model as hidden image h and applying a smoothing operator $\mathbf{B}$ to obtain a smooth model:

$$f = \mathbf{B}h \tag{6.2}$$

In this case the degree of smoothing, e.g. the widths for a Gaussian filter, determines the effective resolution of the model and should be part of the inference. In the AK approach, the smoothing applied to each amplitude is treated as additional hyperparameter, allowing for an adaptive choice of the resolution. The kernel widths can be uncorrelated to separate a constant or slowly varying background from narrow spikes in the signal. If the model is expected to continuously vary the length scale, the kernel widths can be correlated. The ideal resolution is then obtained by the inference process. It depends on the data and the system connecting the model space and the data space. We find a multi-resolution model, offering a complete function space – like in the form free formulation – if the data are informative enough. This model was also called pixon-based in [Pue96] and as AK applied for a set of one dimensional inverse problems [FvD96; Fis02].

For a model based on splines, with a discrete number of knots, the discrete number has to be marginalised – solving a number of sub problems. In the Bayesian sense this comparison relies on the evidence of each model, which can be hard to obtain [Ski06]. The AK model in contrast is defined on a continuous resolution spectrum. Benefit is that the cumbersome work of comparing various models is not necessary. However, the kernel widths have to be marginalised or a maximum a posterior (MAP) approach has to be employed.

Definition

The model function $f(x)$ on a continuous domain is defined as

$$f(x) = \int \mathrm{d}y \, \mathbf{B}\left(\frac{x-y}{b(y)}\right) \cdot h(y) \tag{6.3}$$

for the continuous hidden or latent image $h(x)$ and smoothing function $b(x)$. For our purpose, we are interested in discrete distributions for N kernels at positions x_i:

$$f(x_0) = \sum_i B\left(\frac{x_0 - x_i}{b_i}\right) \cdot h_i \tag{6.4}$$

Maximum Entropy principles lead to a Gaussian as the smoothing function, as it is the least informative function. This happens to be a symmetric and normalised kernel function. Other functions can be used if the problem at hand justifies it.

For a Gaussian kernel we obtain

$$f(x_0) = \sum_i \frac{h_i}{\sqrt{2\pi}\, b_i} \exp\left(-\frac{1}{2}\left(\frac{x_i - x_0}{b_i}\right)^2\right) \ . \tag{6.5}$$

For evaluation on a discrete grid – from x_a to x_b including x_0 – the contribution of the exponential function is given by integration, yielding the error function

$$\hat{f}(x_0) = \int\limits_{x_b}^{x_b} f(x) = \sum_i \frac{h_i}{2}\left(\mathrm{erf}(\frac{x_b}{\sqrt{2}\, b_i}) - \mathrm{erf}(\frac{x_a}{\sqrt{2}\, b_i})\right) \ . \tag{6.6}$$

The probability distribution for this discrete model f is represented by the model parameters b and h, so that

$$p(f) = \int d^N b\, d^N h\, p(f, b, h) \tag{6.7}$$

$$= \int d^N b\, d^N h\, p(f \mid b, h) \times p(b, h) \ , \tag{6.8}$$

with the relation

$$p(f \mid b, h) = \delta(f - \mathbf{B}h) p(b, h) \ . \tag{6.9}$$

Or as posterior in Bayes' Theorem

$$p(f \mid D) = \int d^N b\, d^N h\, p(f, b, h \mid D) \ , \tag{6.10}$$

$$= \int d^N b\, d^N h\, p(f \mid b, h, D) \times p(b, h) \ , \tag{6.11}$$

$$= \int d^N b\, d^N h\, p(b, h \mid D) \times p(b, h) \times \delta(f - \mathbf{B}h) \ . \tag{6.12}$$

With the likelihood

$$
\begin{aligned}
p(D \mid f) \times p(f) &= \int d^N b \, d^N h \, p(D, b, h \mid f) \times p(f, b, h) \,, && (6.13)\\
&= \int d^N b \, d^N h \, p(D \mid b, h, f) \times p(f \mid b, h) \,, && (6.14)\\
&= \int d^N b \, d^N h \, p(D \mid b, h) \times p(b, h) \delta(f - \mathbf{B} h) \,. && (6.15)
\end{aligned}
$$

The Dirac delta ensures, that only the set of kernel widths and amplitudes which exactly match to the model function f are weighted. However, the volume in the subspace is essentially zero, as only a lower dimensional subset contributes. An exception is an amplitude with value equal to zero, implying that the corresponding width does not affect the resulting f. Even besides this trivial case the function space is overcomplete. However, the model function is not perfectly defined, but inherits some uncertainty from the data. The exploration in h and b, implying each set represents an unique model, is therefore meaningful in a MAP approach or integration.

Bayesian framework

The aim of the exploration is to find the posterior probability, given by

$$
p(f \mid d, \sigma, I) = \frac{p(d \mid f, \sigma, I) \cdot p(f \mid I)}{p(d, I)} \tag{6.16}
$$

with the product of the likelihood $p(d \mid f)$ and the prior $p(f)$ in the numerator and the evidence $p(d)$ in the denominator.

The evidence is of importance when two or more models are to be compared, being an objective measure for Ockham's razor. For the AK model each distinct set of kernel widths can be seen as a separate model.

The kernel widths are hyperparameters. For the MAP estimation via optimisation, they are treated in the Gaussian approximation. When integrating via Monte Carlo methods larger kernel widths are naturally favoured, unless the data demand details in the model via the likelihood function. While small kernel widths allow to represent many features, larger widths lead to correlation between neighbouring kernels. For small widths a specific set of amplitudes describes a certain profile, while for larger widths a larger volume in the amplitude prior space leads to similar profiles. The evidence can be interpreted as the volume of the prior in which the likelihood is large, which favours larger kernel widths. At the same time the likelihood sets an upper limit for the kernel widths.

As the evidence is a constant for a given model, the optimal model parameters can be found by maximising

$$
p(f \mid d, \sigma, I) \propto p(d \mid f, \sigma, I) \cdot p(f \mid I) \,. \tag{6.17}
$$

The uninformative prior for a positive additive distribution (PAD) is the entropic prior [Ski89a]

$$p(h \mid \alpha, I) = \frac{1}{Z} \exp(\alpha S) \ . \tag{6.18}$$

with the entropy

$$S = \sum_{j=1}^{N} h_j - m_j - h_j \ln(\frac{h_j}{m_j}) \tag{6.19}$$

based on the default model m_j. The factor Z ensures normalisation for integrals over the entire parameter space

$$Z = \int d^N h \, p(h \mid \alpha, I) \ . \tag{6.20}$$

Introducing the AK as model to describe the data, the posterior is written – with f depending on parameters h and b – as

$$p(f \mid d, \sigma) = \int d^N h \ d^N b \ p(f, h, b \mid d, \sigma) \ , \tag{6.21}$$

$$p(f \mid d, \sigma) \propto \int d^N h \, d^N b \, \delta(f - B \times h) p(d \mid h, b, \sigma) \cdot p(h \mid b) \cdot p(b) \ . \tag{6.22}$$

Effective number of degrees of freedom (eDoF)

The degrees of freedom (DoF) are an important quantity for model comparison, as additional DoF's typically improve the likelihood, while not necessarily gaining more information about the system. For the AK model, the complexity is described by the transfer matrix, mapping the hidden image into the model space. The most simple example is a set of delta distributions as kernel, for which **B** is the unit matrix, corresponding to a 1 to 1 map from hidden image to the model function f, without smoothing. Note that **B** is not in general square, e.g. for more or less kernels than cells in the model space for over- or under-sampling.
For the limit of the kernel widths being much larger than the system size

$$b_i \rightarrow \infty \ \forall i \tag{6.23}$$

the effective degree of freedom (eDoF) is approach 1, because the resulting function is the sum of constant distributions in the model space. The other limit $b_i \rightarrow 0$ represents a collection of independent delta peaks mentioned before. The eDoF can be determined for the normalised kernel by the eigenvalues λ_i of the squared transfer matrix **B**:

$$\text{eDoF} = \sum_{i=1}^{N} \sqrt{\lambda_i(\mathbf{B}^T \cdot \mathbf{B})} \ . \tag{6.24}$$

This value corresponds to the rank of **B**, which in turn is equal to the number of independent parameters of the model [SM06].

MAP estimation for kernel widths

Related to this is expression is the evaluation for the posterior contribution. The principle of Ockham's razor is expressed by the prior volume covered by the high-likelihood area

$$p(b \mid d, I) = \int d^N h \, p(d \mid h, b, I) \cdot p(h, b \mid I) \tag{6.25}$$

for which a Gaussian approximation with expansion around the maximum $\hat{h}$

$$p(b \mid d, I) \approx p(b) \cdot p(d \mid \hat{h}, b) \cdot p(\hat{h}) \cdot \det{}^{-1/2}(\mathbf{H}) \tag{6.26}$$

is used. The Hessian is expressed for a linearised system by

$$\mathbf{H} = \mathbf{B}^T \, \mathbf{A}^T \, \mathbf{diag}(1/\sigma^2) \, \mathbf{A} \, \mathbf{B} \tag{6.27}$$

for which the determinant has to be determined.

Here $\mathbf{A}$ corresponds to the linear system mapping the AK profile into the data space and the diagonal matrix $diag(1/\sigma^2)$ represents the fidelity of the single data points. For a linear system, like an aperture with a point spread function acting on the kernel profile, A is just the matrix representing this property of the observing system.

When using the entropic prior, an additional element appears in the Hessian. By taking the derivative for a single amplitude h_i we find

$$\frac{\partial p(h \mid \alpha)}{\partial h_i} = \alpha \cdot \ln\left(\frac{h_i}{m}\right) \tag{6.28}$$

and

$$\frac{\partial^2 p(h \mid \alpha)}{\partial {h_i}^2} = \left(\frac{\alpha}{h_i}\right) \tag{6.29}$$

effectively adding the diagonal elements

$$\mathrm{diag}(\alpha/\vec{h}) \tag{6.30}$$

to $\mathbf{H}$. In this case the Gaussian approximation relies on the determinant

$$p(b \mid d, I) \propto \det{}^{-1/2}\left(\mathbf{B}^T \, \mathbf{A}^T \, \mathbf{diag}(1/\sigma^2) \, \mathbf{A} \, \mathbf{B} + \mathrm{diag}(\alpha/\vec{h})\right) \tag{6.31}$$

For the determination of heat fluxes, the diffusive transport in the tile has to be taken into account. A representation mapping the heat flux vector to the resulting change of temperature the matrix $\mathbf{T}$ is chosen. As the diffusion length between two frames is smaller than the separation of the camera pixels, this is a weakly banded, close to a diagonal, matrix representing

$$T_{i,j} = \frac{\partial \mathbf{T}_i}{\partial q_j} \ . \tag{6.32}$$

In addition the sensitivity of the data to the surface temperature is expressed by the diagonal matrix

$$I_{i,i} = \frac{\partial \Gamma_i}{\partial T_i} \ . \tag{6.33}$$

Non-negligible interactions of pixels enter this matrix as off-diagonal elements. The two components $\mathbf{T}$ and $\mathbf{I}$ define the linearised model from the incoming heat load to the infrared (IR) observation:

$$\mathbf{A} = \mathbf{I} \cdot \mathbf{T} \tag{6.34}$$

and hence

$$\mathbf{H} = \mathbf{B}^T \, \mathbf{T}^T \, \mathbf{I}^T \, \mathbf{diag}(1/\sigma^2) \, \mathbf{I} \, \mathbf{T} \, \mathbf{B} \tag{6.35}$$

plus the addition from the amplitude prior. The matrix determinant is obtained as product of the eigenvalues λ_i 5.4.

Note on $\mathbf{T}$: depending on the *tension* in the tile, i.e. the temperature gradient at the surface into the depth, zero incoming surface heat flux does not result in zero change of the surface temperature. However, for the quantification of the uncertainty the default behaviour when no heat impinges on the surface is not of relevance. The formulation stems from a second order approximation around the current surface temperature.

Overall the precision matrix $\mathbf{H}$ tracks how the parameter representing the heat flux in the hidden image turns to a smooth heat flux in terms $\mathrm{MW\,m^{-2}}$ ($\mathbf{B}$), which interacts with the tile to a change in surface temperature in units of $\mathrm{K/MW\,m^{-2}}$ ($\mathbf{T}$), which then is observed by the camera counting electrons proportional to the photon flux from Planck's law in units of counts/K ($\mathbf{I}$) and finally to a dimensionless scalar by dividing by the standard deviation in counts.

An example: the hidden image is scaled by a factor of 10^5, representing a heat flux of $0.1\,\mathrm{MW\,m^{-2}}$. The conductivity of tungsten at a sample rate of 1 kHz corresponds to a temperature change of $0.17\,\mathrm{K}$ / $0.1\,\mathrm{MW\,m^{-2}}$. The camera sensitivity is about $1 \cdot 10^4\,\mathrm{cts\,K^{-1}}$ leading to an overall sensitivity of the forward model of 17 kcts / $\mathrm{MW\,m^{-2}}$, or 170 counts per unit amplitude. Together with a standard deviation of about 500 cts – corresponding to a quarter million counts and Gaussian noise – the result is on the order of 1 in parameter space. This result justifies the chosen scaling of the heat flux and implies the largest eigenvalues to be around 1 as well, which is numerically favourable.

Trace approximation

As presented in section 5.4.2 the trace of the matrix can be used to determine the logarithm of the determinant (log-det) of a matrix. The first order expansion of the stochastic trace estimation (STE) – the diagonal of **H** – can be used as simple and rough approximation to the log-det. This is meant as a cheap way to find a good starting point for large problems.

The results can differ significantly from other approaches, especially when the kernel width is growing. However, the *high-detail* areas with the lowest kernel widths are comparable, while the largest kernel widths are above the results from optimisation using SVD or integration. After all, it seems to be a good approximation to set the kernel width to reasonable starting points for the more time consuming approaches shown above.

Confidence interval

For the confidence interval the derivatives of (6.5) for cell j

$$f(x_j) = \sum_i \frac{h_i}{\sqrt{2\pi}\, b_i} \exp\left(-\frac{1}{2}\left(\frac{x_i - x_j}{b_i}\right)^2\right) \tag{6.36}$$

with respect to the amplitudes h_i and kernel widths b_i are calculated.

The hidden image contribution to a cell is given by the transfer matrix element $B_{i,j}$, or explicitly written

$$\frac{\partial f(x_j)}{\partial h_i} = \frac{1}{\sqrt{2\pi}\, b_i} \exp\left(-\frac{1}{2}\left(\frac{x_i - x_j}{b_i}\right)^2\right) = T_{i,j}\ . \tag{6.37}$$

For the kernel widths the derivative is shown step wise, as two terms contain b_i:

$$\frac{\partial f(x_j)}{\partial b_i} = \frac{h_i}{\sqrt{2\pi}} \frac{\partial}{\partial b_i}\left[b_i^{-1} \cdot \exp\left(-\frac{1}{2}\left(\frac{x_i - x_j}{b_i}\right)^2\right)\right]\ . \tag{6.38}$$

Applying the product rule and extracting the exponential from the bracket gives us

$$\frac{\partial f(x_j)}{\partial b_i} = \frac{h_i}{\sqrt{2\pi}} \cdot \exp\left(-\frac{1}{2}\left(\frac{x_i - x_j}{b_i}\right)^2\right) \cdot \left[-b_i^{-2} + b_i^{-1} \cdot \left(-\frac{1}{2}\frac{(x_i - x_j)^2}{b_i^2} \cdot -\frac{2}{b_i}\right)\right]\ . \tag{6.39}$$

Rearranging the terms gives the expression

$$\frac{\partial f(x_j)}{\partial b_i} = \frac{h_i}{\sqrt{2\pi}} \cdot \exp\left(-\frac{1}{2}\left(\frac{x_i - x_j}{b_i}\right)^2\right) \cdot \left[\frac{(x_i - x_j)^2 - b_i^2}{b_i^4}\right]\ . \tag{6.40}$$

In terms of the transfer **B** matrix we thus find

$$\frac{\partial f(x_j)}{\partial b_i} = B_{i,j} \cdot h_i \cdot \left[\frac{(x_i - x_j)^2 - b_i^2}{b_i^4} \right] . \tag{6.41}$$

The confidence interval for the spatial integral of the heat load profile

$$F = \int f(x) dx \tag{6.42}$$

can be of interest as well. For normalised kernels, taking truncation of the kernel at the boundaries into account, this is equivalent to the integral of the hidden image respective the sum of the kernel amplitudes. The uncertainty of F is reduced by the anti-correlation of the single amplitudes. Assuming fixed kernel widths and an equidistant grid – with discretisation $\Delta x \equiv 1$ – the uncertainty is

$$\sigma_F = \sum_{i,j,k} \frac{\partial f_i}{\partial h_j} \cdot \frac{\partial f_i}{\partial h_k} \operatorname{Cov}(h_j, h_k) \tag{6.43}$$

For normalised kernels this can be written as

$$\sigma_F = \sum_{i,j} \operatorname{Cov}(h_j, h_k) \tag{6.44}$$

6.1.3. Extension to 2D kernel

For the case of a two dimensional model image the kernel have to be generalised from 1D to 2D distributions. The most simple approach for any 1D distributions is to use the outer product on two independent distributions along the two main axis. This is justified, when the quantity of interest shows uncorrelated behaviour along these two main axis. A more general treatment can be desirable and would include an angle between the elliptical kernel and the image or the kernel expressed via a covariance matrix.

For this work the outer product of two Gaussian kernels is used, which allows faster computation and normalisation than a rotated kernel system. Therefore per kernel there are three parameters: amplitude $h_{i,j}$, kernel widths $\sigma_{x,i,j}$ and $\sigma_{y,i,j}$. In addition to these $3N_k$ parameters there is a weight for the entropic prior α, for the hidden image [Ski89a].

Figure 6.1 shows an example for two perpendicular, independent normal distributions located at different positions along their axis. For this 2D formulation the matrix expressions are similar to the 1D case: a 10×10 2D system corresponds in size to a 1D system with 100 cells.

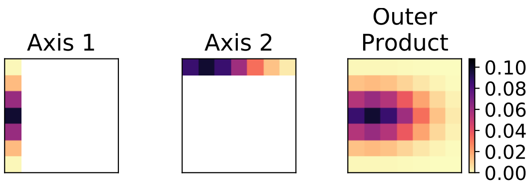


Figure 6.1.: Example for the 2D kernel function as outer product of two Gaussians.

6.1.4. Potential problems

For a system with two kernel – and two eigenvalues – with different widths, the narrower kernel contributes most to the highest eigenvalues, as it has the strongest influence on the result. If the two kernel widths approach and pass each other, the effective degree of freedom actually drops, as the contribution is close to identical, and the role of the kernel width is exchanged. This results in a jump – or at least strong change – in the gradient of $\partial \sum \ln(ev_i)/\partial b_j$.

This is mitigated by the contribution of the entropic prior, which effectively adds positive values to the diagonal of $\mathbf{B}^T \cdot \mathbf{B}$. These values set a lower limit to the eigenvalues obtained.

Alternatively, especially for larger systems, kernels can be thinned, when their widths are similar and much larger than their distance. For this work this was not further investigated, as the change of kernel corresponds to a change in parameter count, adding to the complexity of the problem. For a sufficiently stable implementation and sufficiently small system sizes a dense grid works fine. In higher dimension the rules for thinning have to be well thought of.

6.1.5. Adaptive diffusion

An alternative to place kernels of variable widths, is the application of the diffusion operator with variable, spatially varying conductivities. Especially for higher dimensional systems, the evaluation of this model can be significantly faster than the AK.

Basis is a hidden image, like for the AK model, which is smoothed by using the diffusion operator:

$$h = u(t = 0) \ , \tag{6.45}$$

$$m = u(t = T) \ , \tag{6.46}$$

$$\frac{\mathrm{d}u}{\mathrm{d}t} = \nabla(\kappa(\vec{r})\nabla u) \ , \tag{6.47}$$

with the field $u(t = 0)$ defined by the hidden image and the final model m obtained after time T – e.g. unit time $T = 1$. Like with the AK the regularisation can be defined spatially resolved via the conductivity κ. The conductivity sets the flow j for a given gradient

$$j = \kappa \nabla u \tag{6.48}$$

or in the discrete system

$$j_{i\to j} = \kappa_{i,j}\frac{u_i - u_j}{\Delta x} \ , \tag{6.49}$$

so $\kappa_{i,j}$ is a property at the interface of two cells. Either it is treated directly as a free parameter of the model, or it is defined for each cell on which the interface value has a functional dependence.

Two variants of this functional dependence were tested, the simple average

$$\kappa_{i,j} = (\kappa_i + \kappa_j)/2 \tag{6.50}$$

and the harmonic mean

$$\kappa_{i,j} = \begin{cases} \left(\frac{\kappa_i^{-1}+\kappa_j^{-1}}{2}\right)^{-1} & : \kappa_i > 0 \wedge \kappa_j > 0 \\ 0 & : \text{else} \end{cases} \tag{6.51}$$

While the first results in a more smooth behaviour, the latter results in a faster drop towards zero of the flux, for a low conductivity of a single cell. With the harmonic mean a cell can be entirely shielded from the surrounding, independent of the surroundings conductivities.

The model can be treated similarly to the AK model. However, the Hessian cannot be obtained as efficiently, as the smoothing operator is applied. Building the Hessian is about as computationally expensive as for the AK model. As the faster evaluation is considered a benefit, other ways of inference for the conductivities are needed to justify the use of this model.

Scaling with system size

The numerical cost to evaluate the model depends on the method used. Here we consider an implicit solver using operator splitting – treating the directions as independent. The numerical core is equivalent to the heat diffusion equation (HDE) solver of the forward model, see section 4.3.1. Consider a domain in k dimensions with n cells along each orientation, leading to $N \equiv n^k$ cells. The operator splitting leads to k independent solving steps on each cell, represented by tridiagonal systems of size n. The total number of operations is about $O(k \cdot n^k) = O(k \cdot N)$. Due to the operator splitting, there is a linear increase in cost with each dimension and the total cell count.

However, for the AK model, each kernel has to be evaluated for each cell. This might be mitigated by truncating a kernel of infinite range at a given threshold or using kernel functions with limited range. In the case of a dense set of kernel – one kernel per cell in our field – with infinite range the scaling is $O(N^2) = O\left(n^k\right)^2$. Note that $\left(n^k\right)^2 = n^k \cdot n^k = n^{2k}$. The 1D AK model scales like a 2D DK model and a 2D AK model scale like a 4D DK model. That is except for the linear increase of the adaptive diffusion (AD) model for each extra dimension.

For the diffusive approach an implicit solver is suggested, as it is unconditionally stable, allowing for long range diffusion in one step. For the explicit variant – not using the operator splitting – it becomes more simple and efficient to treat diagonal diffusion terms. The diffusivity can be given as a tensor, describing an anisotropy which is not aligned to the chosen axes describing the system.

6.2. Modelling the physical process

This section outlines the expected temporal behaviour of our quantity of interest. On one hand there are cases where the heat load changes slow in time compared to the sample rate, allowing to correlate consecutive profiles in the prior. On the other hand, there are fast events, shorter than the sample interval, for which several samples in time can be combined to disentangle the fast event from the background. And finally there is a middle ground, where a constant heat load between samples is a reasonable representation, but the correlation in time is weak. How the time-dependence in the posterior is treated is described in section 7.2.4.

6.2.1. One time-step, one profile

Starting at the middle ground is the easiest, when no beneficial correlation for consecutive profiles can be used. In this case the profiles are treated as independent. Using the truncation including two profiles presented in 7.2.4 is suggested anyway. Reason is that a single measurement can indicate a positive heat input at one position when actually no heat is arriving, due to the measurement uncertainty. Including two data sets reduces the probability of on overestimation of heat loads at the level of the signal-to-noise ratio (SNR), where THEODOR would return heat loads alternating around zero.

6.2.2. Slow changes in time – correlated profiles

For slow changes in the heat load, compared to the sample interval, consecutive profiles can be coupled. However, for the used measurement system with sample intervals of about 1 ms available models of the scrape-off-layer (SOL) do not support correlation times of several ms. At the same time, rare intermittent events can change the signal significantly from one frame to the other.

For this reason only a weak coupling in time is used on the parameters defining the heat flux density – via a Cauchy prior with the local SNR as width.

6.2.3. Fast events

At AUG there are fast events, lasting about 10 µs at the outer mid-plane of the tokamak [Man+17] and expected to have a duration of about 100 µs at the divertor. With the current

sample rate of up to 3 kHz this poses an obvious problem. The question arising is how to disentangle these intermittent events from the background signal.
Possible solutions or improvements are

1. Increase the sample rate of the diagnostic to resolve the fast events.
   The increasing the sample rate to fully resolve each event comes at the cost of a reduced signal-to-noise ratio, as the integration time becomes shorter for the same optical setup. Also, the speed of the available systems is limited and coupled to the spatial resolution and field of view. Due to a lack of the needed hardware, this is not further considered in this thesis. From the inference point – working with what is available – point one and two can be combined, when the model is extended.

2. Use interleaved cameras: combine data of cameras recording at different times.
   This option is similarly problematic as the first, as the space for diagnostics in the vicinity of the machine is limited. Especially the number of access ports to the vacuum vessel. However, more space is available outside of the vacuum vessel. A beam splitter could be used to distribute the signal to several cameras, integrating out of sync. This still reduces the photon flux for each camera. To combine high temperature and high temporal resolution, the use of two different different cameras appears appealing. Similarly to the first point, the hardware for this recording technique is not available at the moment. Also the alignment is not trivial.

3. Use several data points of one IR-diagnostic to disentangle several signal contributions between two samples.

4. Add information from other diagnostics.
   More information – not IR based – like the onset-time – when does the burst start – or the burst duration can be used.

The models explained so for in this chapter allow to represent our quantity of interest: the heat flux density distribution. For the physical problem, the heat flux as function of space and time is of interest. The time-dependent problem is solved in time, with the finite difference solver introduced in section 4.3. For most of the analysis, the heat flux is assumed to be constant between two samples. While a linear interpolation in time at first appears more natural, it leads to the problem of further analysis and visualisation. To disentangle the fast events from the background, this model is extended.
For the proof of principle a separation is introduced into constant background and an additional burst heat load. To disentangle the two, three measurements are included:

1. The last unaffected frame – to give a good estimate for the background

2. The first affected frame – here the sharp increase in temperature may be recorded, but with an uncertainty of arrival time and duration, the heat input cannot be deduced unambiguously.
3. The frame after – to give information about the decay of the surface temperature, as the heat is transported into the tile.

An example is presented in section 9.1, where a synthetic test case is used as benchmark. As it is not obvious from a physical point of view, why the background should stay constant, this is considered a dummy model.

7. Bayesian THEODOR (BayTh)

The name BayTh refers to the deterministic tool THEODOR – chapter 4 – used in a probabilistic framework. This chapter explains how the single elements explained in the pervious chapters come together for the analysis, plus some technical details.

7.1. Evaluation chain

First the initial state of the tile has to be chosen. The heat potential field – being a function of the temperature and the conductivity, see section 2.1 – is set to a uniformly distributed, with the value chosen according to the first deduced temperature profile of the measurement series. An uniform distribution is justified, as the time between experiments is long enough to ensure that a thermal steady state is reached. Using the first measurement vector is a reasonable choice, as a small offset in the starting temperature will not affect the inferred heat loads within the measurement uncertainty. For an uncertainty of $(25 - 100)\,\mathrm{kW\,m^{-2}}$ per measurement and a profile with 100 pixels – typically 200-300 are used – the effective uncertainty is $(2.5 - 10)\,\mathrm{kW\,m^{-2}}$. While there is an effect on the inferred heat load, it is negligible compared to the other uncertainties.

The model function m for our quantity of interest f – see chapter 6 – is used to obtain a spatially discrete heat flux density profile $\vec{f^k}$ in the time interval $(t^{k-1}, t^k]$

$$\vec{f^k} = m(p^k)\ . \tag{7.1}$$

In the most simple case – with one profile modelled between two measurements, so no temporal over- or undersampling – this heat load profile leads from measurement d^{k-1} to d^k. The choice of this convention will become clear later in this chapter, when parameter set k is used in the likelihood for data set k. This heat flux is used as the surface boundary condition for the diffusion solver – see chapter 4 – denoted here as operator $\mathbf{A}$ to model the evolution of the heat potential distribution u in the tile:

$$u^k = \mathbf{A}(u^{k-1}, f^k) = \mathbf{A}(u^{k-1}, m(p^k))\ . \tag{7.2}$$

This numerical integration is divided into several sub-steps, not explicitly denoted here. The data consist of the counts of the diagnostic – being spatially resolved – and the integration time. The resulting surface temperature of the modelling is used to calculate the corresponding

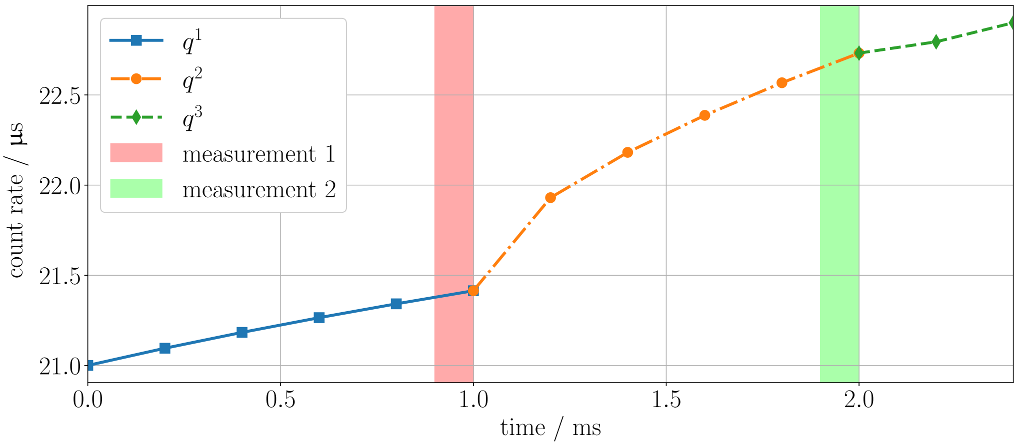


Figure 7.1.: Example of the numerical integration results in time, for a the five fold over-sampled forward model. The vertical bars mark the time intervals, when the modelled count rate is integrated.

photon rate γ of the grey body radiator, given its emissivity and Plank's law – see section 2.3. The obtained photon flux is translated into a count rate, given the calibration of the diagnostic.

Figure 7.1 shows an example of the predicted count rate from the forward model. Different values for the heat load are defined from the end of one measurement until the end of the next measurement – in this case in intervals between integer time in ms. This choice takes into account that any heat exchange does not affect the state of the system before it starts and the influence on the system state is best determined at the end of the interaction interval. A typical sample rate of 1 kHz and an integration time of the diagnostic of 100 µs is used in the present example. The count rate is interpolated linearly in between the solver steps, as depicted.

The modelled counts – given the known integration time of the diagnostic – are compared to the data d via a Gaussian likelihood, with the standard deviation derived from the experimental photon count. An alternative is to use a Cauchy distribution for the likelihood, providing resilience against outliers.

Probabilities are evaluated in negative log-space. Therefore the log-posterior is the sum of the log-likelihood and the log-prior associated with the parameters. Either the maximum a posterior (MAP) of the posterior is searched for, or a Markov chain is used to obtain the mean of the parameters, see chapter 5.

7.2. About the parameters

7.2.1. Scaling of the parameters

Heat load amplitudes

The heat load density amplitudes are given in units of $100\,\mathrm{kW\,m^{-2}}$. A typical peak heat load is about $1\,\mathrm{MW\,m^{-2}}$, with the lowest experimentally meaningful heat load on the order of $10\,\mathrm{kW\,m^{-2}}$. This choice depends on the problem at hand, but for the analysed data sets numerical considerations – regarding cut-off and rounding errors – make this value a reasonable choice.

Adaptive resolution – kernel width

The kernel widths are defined in terms of the pixel width – the grid is equidistant in the task at hand. The lowest value is 0.5 pixel, which can be necessary to resolve features affecting only a single cell. A single amplitude then attributes with 95% to the cell it is positioned at, with only 5% contribution to the neighbour. The upper limit is chosen to be 40 pixel, as higher values show no benefit to the smoothness of the signal.

7.2.2. Entropic prior

For the entropic prior – see section 2.7.2 – a standard model m has to be selected. The deterministic approach from THEODOR presents a simple way of how to get a proxy for the heat load profile. The average of this profile – technically using the measured data already once – is then used as a constant standard model $m = \langle q_{classic} \rangle$ for all cells. An alternative is to use the mean heat load of the profile inferred before. As the simple evaluation is numerically cheap and strong jumps in the signal appear under certain plasma conditions, using the deterministic approach is preferred.

The *weight* α is not of interest to our inference and should be marginalised. An alternative is to determine a single *best* value, in terms of the MAP solution. This is justified, when the posterior for α is strongly peaked for this value:

$$
\begin{aligned}
p(h \mid D) &= \int \mathrm{d}\alpha \, p(h \mid \alpha, D) \, p(\alpha) && (7.3) \\
&\approx \int \mathrm{d}\alpha \, p(h \mid \alpha, D) \, p(\alpha) \, \delta(\alpha - \alpha^*) && (7.4) \\
&= p(h \mid \alpha^*, D) p(\alpha *) \; . && (7.5)
\end{aligned}
$$

This was tested for a set of MAP solutions, obtained from the analysis of experimental data. For AUG discharge 32217 – presented in section 8.2 – the posterior for α was scanned, resulting in figure 7.2. The posterior is presented on linear scale in the top plot and its negative z

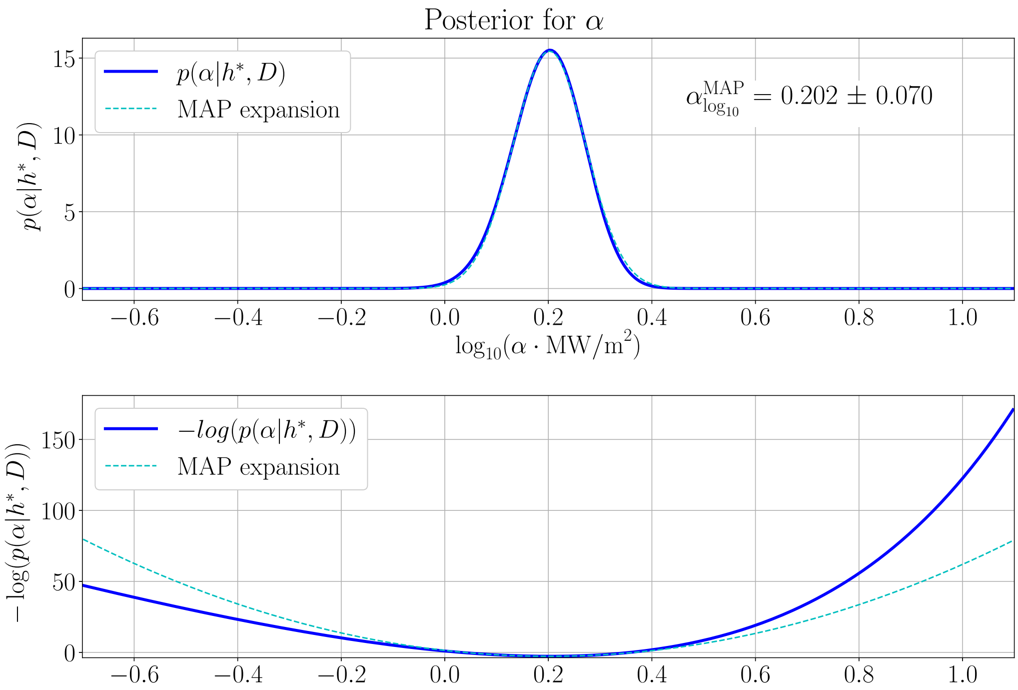


Figure 7.2.: Posterior distribution for α plotted on a linear scale (top) and a logarithmic scale (bottom) for the probability. α is scanned on plotted on the $\log_{10}$ scale.

logarithm in the bottom plot. In both cases the Gaussian of the second order expansion – in log-space for α – is shown as a dashed line. The position of the MAP is at $\alpha^{MAP} \approx 1.6\,\mathrm{MW\,m^{-2}}^{-1}$. This value and standard deviation – denoted $\alpha^{\mathrm{MAP}}_{log_{10}}$ – refer to the MAP position in log-space $\log_{10}(\alpha^{\mathrm{MAP}}) = 0.202$ and the standard deviation of the distribution determined in log-space $\mathrm{std}(p(\log_{10}(\alpha \mid h^*, D))) = 0.070$. In negative log-space the asymmetry of the posterior is apparent, but the weight is essentially zero after three sigmas, where the agreement is good. This justifies the use of the MAP approach. Note, that for the amplitude parameter in units of $\mathrm{MW\,m^{-2}}$ the unit for α is its inverse. For convenience the scale of the amplitudes was changed from $100\,\mathrm{kW\,m^{-2}}$ to $1\,\mathrm{MW\,m^{-2}}$. The MAP position of 0.2 corresponds to $1.6\,\mathrm{MW\,m^{-2}}^{-1}$ which multiplies to unity for an amplitude of $0.6\,\mathrm{MW\,m^{-2}}$.

From

$$
\begin{aligned}
p(\alpha, D, h) &= p(\alpha \mid D, h) \cdot p(h \mid D) \cdot p(D) && (7.6)\\
&= p(h \mid D, \alpha) \cdot p(D) \cdot p(\alpha) && (7.7)
\end{aligned}
$$

it follows the posterior for α

$$
p(\alpha \mid D, h) = \frac{p(h \mid D, \alpha) \cdot p(\alpha)}{p(h \mid D)} \tag{7.8}
$$

presented in figure 7.2.

For the adaptive kernel (AK) model the MAP approach is preferred, as the additional diagonal entries on the precision matrix stabilise the singular values of the SVD. Effectively, the diagonal α/h_i presents a lower limit to the eigenvalues, which otherwise drop to zero for increasing kernel width.

The consideration is different, when a Markov chain or other integration techniques are used. Especially for the adaptive diffusion model, see section 6.1.5, where the precision matrix is not easily obtained and integration is suggested anyway. Here the marginalisation of α is suggested.

7.2.3. Initial parameters

First choice

The initial parameters are typically scattered around a reference value on the first analysis. The scattering for the i'th parameter is done according to a simple modulo pattern with primes

$$\begin{aligned} \delta_i = \delta_0 \cdot (&i \bmod 19 - i \bmod 17 + i \bmod 13 - i \bmod 11 && (7.9) \\ &+ i \bmod 7 - i \bmod 5 + i \bmod 3 - i \bmod 2) && (7.10) \end{aligned}$$

for a scatter scale factor $\delta_0 \approx 10^{-6} \cdot p_0$. This is applied to the amplitudes as well as the spatial regularisation parameters. Compared to random numbers this allows the reproducibility using various computer systems.

The initial kernel width is to to $b_0 = 5\,\text{pixel}$. The initial amplitude to $h_0 = 1$.

Related, sequential problems

For the time trace analysis the consecutive heat flux patterns are typically similar to each other. For this reason the initial parameter set for the next profile is the same as the last. The change in the noise of the data is enough to lead to a quasi-random gradient for the amplitudes, even for a constant model as ground truth. For the kernel widths the parameters of the last profile at index k are adapted according to the rule

$$b_i^{k+1} = b_0^{1/3} \cdot b_i^{k\,2/3} \tag{7.11}$$

for the initial kernel width $b_0 = 5\,\text{pixel}$. This corresponds to the geometric mean, weighting the old parameter twice and the initial parameter once. This prevents the new profile to snap to the same configuration like the old, allowing the optimiser to find a more or less detailed solution compared to the old model.

On multi-modality

The initial set of parameters does not matter in the search for the minimum of a unimodal problem. However, it will influence the determined MAP result on multi-model problems. Unless a Markov Chain is used, which can transit between modes, the general hope is to find a *good enough* mode. For this reason, several runs with independent initial parameters – e.g. drawn from the prior – can be used and compared.

For the AK model slight differences in the kernel width distribution – arising from different starting points – affect the MAP widths distribution. However, experimental observations show that the resulting profile is not affected withing the error bars. For narrow features the kernel width is well defined and found reproducible.

In the case of larger kernel widths, the case of strong regularisation, the kernel widths are not as well defined in the posterior. However, a similar choice of widths will not affect the profile shape significantly, due to the regularisation.

Depending on the problem at hand this has to be studied for each case.

7.2.4. Posterior for a time trace of data

Here the behaviour of the posterior is considered, as more data get available over time, in our time-dependent problem. For simplicity a vector of scalar measurements d^k at consecutive points in time $t^k, k = 0, ..., K$ is considered. The heat flux density parameters q^k are denoted as scalars as well. This notation and structure works for one and two dimensional data and profiles as well.

The state of the system at index k is written as y^k and refers to the temperature distribution in the tile. The resulting measurement is denoted d^k and the parameter describing the heat load in between $k-1$ and k is denoted q^k. So that the initial state y^0 is progressing with heat load described by q^1 to state y^1 with the measurement d^1 being the first to depend on q^1.

We start with the posterior for the first parameter set q^1

$$p(q^1 \mid d^1) = \frac{p(d^1 \mid q^1) \cdot p(q^1)}{p(d^1)} \tag{7.12}$$

given one data set d^1. This is not the most informed distribution for q^1, as subsequent data provide additional information on q_1.

Now the second data set is included, influencing a new second parameter set via the new likelihood:

$$\begin{aligned} p(q^1, q^2 \mid d^1, d^2) &= \frac{p(d^1, d^2 \mid q^1, q^2) \cdot p(q^1, q^2)}{p(d^1, d^2)} && (7.13) \\ &= \frac{p(d^2 \mid q^1, q^2) \cdot p(q^2 \mid q^1)}{p(d^2)} \cdot \frac{p(d^1 \mid q^1) \cdot p(q^1)}{p(d^1)} \; . && (7.14) \end{aligned}$$

The right hand side on the first line shows a compact notation, where both data sets are included in one likelihood term. For independent measurements the likelihood can be split, leading, together with reorganisation of the prior and evidence, to the second line, which is discussed from here on. The first fraction contains the likelihood of the second data set, given both parameter sets and the prior for the second parameter set conditional on the first parameter set. Any interdependence of the parameter sets is captured in the prior $p(q^2 \mid q^1)$. The second fraction is the posterior obtained for q^1 considering only the first data set, which can be seen as prior for q^1. We use the fact that the likelihood $p(d^1 \mid q^1, q^2)$ is independent of q^2, due to causality, and reduces to $p(d^1 \mid q^1)$. This is a special case of the Bayesian update rule, where the former posterior – obtained including only d^1 – is the prior for the new posterior, where new data is included. In this special case the new data set affects two parameter sets, while the first parameter set only was affecting the first data set.

This procedure can be extended to a third data and parameter set, etc.:

$$\begin{aligned} p(q^1, q^2, q^3 \mid d^1, d^2, d^3) = & \frac{p(d^3 \mid q^1, q^2, q^3) \cdot p(q^3 \mid q^1, q^2)}{p(d^3)} \cdot \\ & \frac{p(d^2 \mid q^1, q^2) \cdot p(q^2 \mid q^1)}{p(d^2)} \cdot \\ & \frac{p(d^1 \mid q^1) \cdot p(q^1)}{p(d^1)} \ . \end{aligned} \tag{7.15}$$

Posterior truncation

The posterior is truncated similar to a Kalman Filter or Kalman smoother, see [Sär13].

To keep the posterior introduced above tractable, only a small set of the total parameter sets (heat flux profiles) is considered at a time. A truncation is introduced, though any heat influx in principle influences all following measurements. This is justified, as the diffusion equation obscures information exponentially in time. It should be noted that the justification for the truncation threshold depends on the length of the time steps between measurements, relative to the diffusivity D in the material, and on the correlation of sequential heat load profiles. A correlation can be introduced by the heat flux model on the plasma side.

As the influence of the heat load q at time t_k to the surface temperature T_s with respect to the equilibrium temperature T_eq decays with

$$T_\mathrm{s}(t) - T_\mathrm{eq} \propto q \cdot p(-(t - t_k)/D) \tag{7.16}$$

and causality prevents an influence before $t = t_k$, the data taken into account to infer parameters at index k are limited. The heat flux density at time k does not affect the system state before – e.g. y^{k-1} – and accordingly it does not affect the measurements d^{k-1} and before.

However, measurements at later times do provide information on earlier heat load profiles[1]. This is considered Bayesian smoothing [Sär13, ch. 8] and the reason to not treat the profiles entirely independent. However, any data with index $k + L$ and later are omitted, as later heat loads have a higher influence. This maximum range L has to be adapted to the time scales at hand. For the rest of this work $L = 2$ is used, so that two measurements following a heat profile are taken into account to infer its shape. This leaves us in principle – conditional on all other parameters – with

$$p(q^k \mid d^{k-1}, d^k, d^{k+1}, ..., d^N) \approx p(q^k \mid d^k, d^{k+1}) \, . \tag{7.17}$$

Returning to the example in (7.15), this corresponds to choosing q^1 to its MAP value from $p(q^1, q^2 \mid d^1, d^2)$. This can be written conditional to show the truncation as

$$p(q^2, q^3 \mid q^1, d^1, d^2, d^3) = p(q^1, q^2, q^3 \mid d^1, d^2, d^3) \cdot \frac{1}{p(q^1 \mid d1, d2, d3)} \tag{7.18}$$

where the probability distribution in the last term is approximated by

$$p(q^1 \mid d^1, d^2, d^3) \approx p(q^1 \mid d^1, d^2) \tag{7.19}$$

due to the decaying influence of q^1 on the system. Later data sets add less information than earlier ones. Accordingly the posterior for q^1 includes only the two most relevant data sets. Keeping the dependence in the prior leads to

$$\begin{aligned} p(q^2, q^3 \mid q^1, d^2, d^3) \propto & \frac{p(d^3 \mid q^2, q^3) \cdot p(q^3 \mid q^1, q^2)}{p(d^3)} \cdot && (7.20) \\ & \frac{p(d^2 \mid q^2) \cdot p(q^1 \mid q^2)}{p(d^2)} \, . && (7.21) \end{aligned}$$

When q^1 is replaced by its mean μ^1 the dependence of the posterior distribution for later parameters on d^1 is removed as well.
When there is no – or a negligible – dependence of p^3 on the prior of q^1, the posterior can be written – with implicit dependence on q^1 – as

$$\begin{aligned} p(q^2, q^3 \mid d^2, d^3) \approx & \frac{p(d^3 \mid q^2, q^3) \cdot p(q^3 \mid q^2)}{p(d^3)} \cdot && (7.22) \\ & \frac{p(d^2 \mid q^2) \cdot p(q^2)}{p(d^2)} && (7.23) \end{aligned}$$

It is noted that the truncation for the likelihood and prior can extend over a different range of time indices. The range of the likelihood affects how often the forward model has to be

---

[1] An important example is given in the discussion of ELM's in section 3.5.

solved, while a dependence on the prior can typically be included in a less costly way. Restricting the parameter set q^1 to the two next data sets, its MAP solution is obtained from a Gaussian approximation, as introduced in the end of this section. The mean and standard deviation are used to calculate the corresponding heat load profile and its credibility range. The posterior is collapsed to a delta at its mean, which is used as conditional starting point for the processing of further data. Otherwise, the forward model sequence to solve would grow with increasing size of the data vector.

For completeness, the form of the distribution for $p(q^2 \mid q^1)$ is presented in the following. With the first parameter set fixed, the conditional distribution is

$$\begin{aligned} p(q^2 \mid q^1, d^1, d^2) &= p(q^1, q^2 \mid d^1, d^2)/p(q^1 \mid d^1, d^2) && (7.24) \\ &= \frac{p(d^1 \mid {}^1)p(q^1)}{p(d^1)} \cdot \frac{p(d^2 \mid q^1, q^2)p(q^2 \mid q^1)}{p(d^2)} \cdot \frac{1}{p(q1 \mid d^1, d^2)} && (7.25) \end{aligned}$$

where the last part corresponds to q^1 for marginalised further parameter set(s) q^2

$$\begin{aligned} p(q^1 \mid d^1, d^2) &= \int p(q^1, q^2 \mid d^1, d^2)\mathrm{d}q^2 && (7.26) \\ &= \int p(q^1 \mid q^2, d^1, d^2)p(q^2)\mathrm{d}q^2 \,. && (7.27) \end{aligned}$$

Ignoring the constants – quantities not depending on q^2 – we find an expression, which is proportional to the posterior:

$$p(q^2 \mid q^1, d^1, d^2) \propto \frac{p(d^2 \mid q^1, q^2)p(q^2 \mid q^1)}{p(d^2)} \,. \tag{7.28}$$

Dropping the explicit conditional dependence on q^1 and d^1 we write

$$p(q^2 \mid d^2) = \frac{p(d^2 \mid q^2)p(q^2)}{p(d^2)} \tag{7.29}$$

which is Bayes' Theorem, for a fixed, implicit, conditional parameter. Still, the state of the system y^1 depends on the former parameter and technically inherits its uncertainty.

Note on truncation

Including more than two consecutive heat load and data profiles in the posterior is not beneficial to the analysis in scenarios like presented in section 8. Consecutive heat loads are anti-correlated. In addition the influence of heat input to the surface temperature is decaying in time, due to to diffusion process. The information gain is therefore limited and a trade of between computation time for the MAP and profile accuracy has to be made. Exceptions are fast events, see section 6.2.3, and radiation from sources which are not accounted for by the model, see section 3.5.

Gaussian approximation

Using a Gaussian approximation for $p(q^1 \mid d^1)$ we write

$$p(q^1 \mid d^1) \approx \mathcal{N}(\mu_1^1, \Sigma_1^1) \tag{7.30}$$

with the number of data sets considered denoted as subscript.
The first parameter set is updated when considering the second data set and the joint distribution from (7.14) together with the Gaussian approximation yields

$$p(q^1, q^2 \mid d^1, d^2) = \frac{p(d^2 \mid q^1, q^2) \cdot p(q^2 \mid q^1)}{p(d^2)} \cdot \mathcal{N}(\mu_1^1, \Sigma_1^1) \; . \tag{7.31}$$

This distribution itself is approximated as a joint normal distribution

$$p(\begin{pmatrix} q^1 \\ q^2 \end{pmatrix} \mid d^1, d^2) \approx \mathcal{N}(\begin{pmatrix} \mu_2^1 \\ \mu_2^2 \end{pmatrix}, \begin{pmatrix} \Sigma_2^{11}, \Sigma_2^{12} \\ \Sigma_2^{21}, \Sigma_2^{22} \end{pmatrix}) \tag{7.32}$$

with the symmetry $\Sigma_2^{21} = {\Sigma_2^{12}}^T$.
For the MAP solution of q^1 given d^1, d^2 it simply is

$$p(q^1 \mid d^1, d^2) = \mathcal{N}(\mu_2^1, \Sigma_2^{11}) \tag{7.33}$$

which is used for the MAP profile and credibility range.
The posterior for q^2 conditional on q^1 is found to be

$$p(q^2 \mid q^1, d^1, d^2) \approx \mathcal{N}(q^2 + \Sigma_2^{21} {\Sigma_2^{11}}^{-1} (q^1 - \mu_2^1), \Sigma_2^{22} - \Sigma_2^{21} {\Sigma_2^{11}}^{-1} \Sigma_2^{12}) \tag{7.34}$$

and fixing the MAP result μ_2^1 gives

$$p(q^2 \mid q^1 = \mu_2^1, d^1, d^2) \approx \mathcal{N}(\mu_2^2, \Sigma_2^{22} - \Sigma_2^{21} {\Sigma_2^{11}}^{-1} \Sigma_2^{12})) \; . \tag{7.35}$$

However, when sampling both parameters the forward model has to be evaluated for both heat load models, even when evaluation of the first likelihood is circumvented. A speed up requires the evaluation of the Gaussian approximation – a matrix-vector and a vector-vector product – to be faster then the calculation of the likelihood and prior of the model. For the MAP solution the evidence is typically ignored anyway. This can be made more beneficial, if the general forward calculation can be replaced – based on the limited range of parameters given the covariance matrix – by an available linearised or surrogate model, which evaluates faster.

Caveats

The approximation is based on two quantifiers: the mean and the covariance matrix. While the mean is taken from the MAP result, obtaining the Hessian is technically tricky, unless an analytic expression is available. Two problems can arise:
The model is compatible with the Gaussian approximation, but numerical errors lead to negative eigenvalues of the covariance matrix. This situation allows solutions, where the evaluation of the negative log-posterior

$$-\ln\left(\mathcal{N}(\mu, \mathbf{\Sigma})(x)\right) \propto \frac{-1}{2} \cdot \vec{x}\, \mathbf{\Sigma}^{-1}\, \vec{x} \tag{7.36}$$

leads to negative values. Depending on the magnitude of these eigenvalues and the new data this effect can drive parameters away from the old MAP solution. This can be expected when the data and the model leave two or more parameters with large (or small) covariance. An option to compensate numerical errors is to add a small contributions on the diagonal.
The second option is that – even if numerical problems do not occur – the underlying model is not sufficiently smooth. This can be expected, when a parameter is limited by a hard upper or lower limit.
The Gaussian approximation is available in the developed code, but due to the mentioned stability issues the exploration of the full system – (7.14) or (7.15) – is suggested.
Examples on handling probability distributions are presented in the appendix A.1.

8. Results for 1D profiles

This chapter presents results from the evaluation of one dimensional heat load profiles in ASDEX Upgrade. Section 8.1 presents a comparison between classic THEODOR and the statistical extension Bayesian THEODOR (BayTh) using the adaptive kernel (AK) model on synthetic data. Section 8.2 presents the analysis of measured data on the example of a discharge with external magnetic perturbation.

Further results are presented in separate chapters: fast events in chapter 9 and 2D deconvolution in 10.

8.1. Synthetic reference

As a reference an L-Mode like heat flux profile is used. The profile is characterised with the 1D equation (2.21) introduced in section 2.2. In the appendix – see B.1 – the time-dependent parameters are listed and the generating python-script are shown.

The typical procedure is to choose a setting for the IR-System, either with constant integration time or with the same adaptive method as used for the AUG System, where the histogram of the last frame is used to set the integration time for the next frame. This system allows a high signal-to-noise ratio throughout the entire duration of the experiment. For more details see [Sie+15]. For the presented case a constant integration time of 100 µs is used. This corresponds to the typical maximal integration time in experiments. This represents the behaviour in experiments with such low heat flux densities – or short duration – and the resulting small increase in peak surface temperature.

Figure 8.1 shows the heat flux density as colour map for the first 200 indices. The abscissa represents time and the ordinate the spatial location. The parameters of the profile are changed smoothly in time, with the dip in the peak at about 175 ms being mainly caused by a broadening of the profile. This mimics slow transitions of the plasma parameters in an experiment. The heat load is ramped up linearly for the first 100 ms starting with zero heat flux, to investigate the behaviour of the kernel for low heat load scenarios.

From this reference the surface temperature evolution is calculated using the solver underlying THEODOR. For this typical L-Mode camera setup and scenario the forward model and synthetic measurement is performed. One realisation of Gaussian noise is added and used for both inversions in order to compare the methods. Figure 8.2 shows the heat load obtained with THEODOR (left) and its absolute deviation to the reference (right). This is a typical noise patterns, uncorrelated in time and space. Figure 8.3 presents the same comparison for the BayTh result, where the range of the heat map for the deviation is identical to 8.2. By construction the profiles are smooth in space, while jumps time are possible. It can be noticed, that the deviation is strongly reduced in the flat areas, while the level is at most similar to the one of THEODOR around the peak of the signal.

In the initial phase the AK describes a more flat heat load pattern, while the signal is weak. In the second half of the time trace the deviation is basically homogeneous along single profiles. The model represents the peak shape clearly.

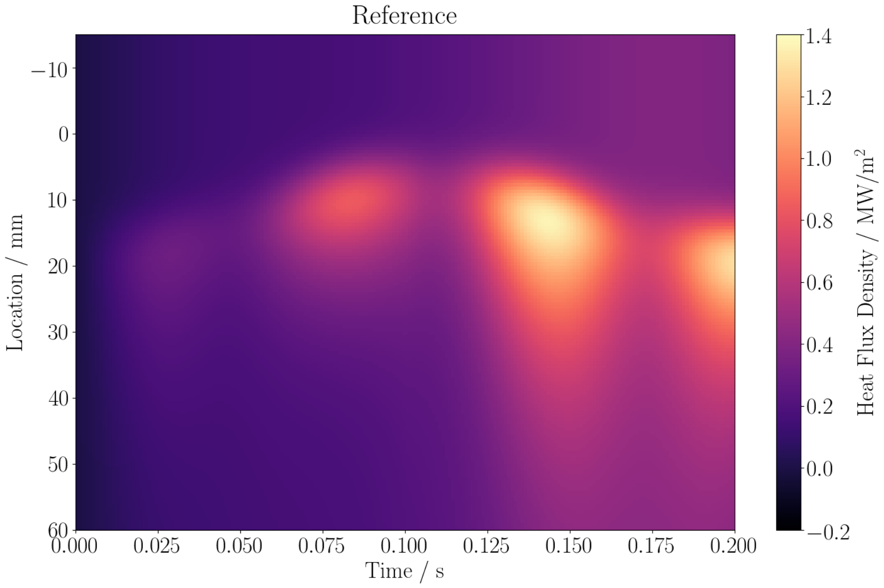


Figure 8.1.: Reference heat flux density used for the synthetic test.

Figure 8.4 shows a comparison between the 1D profiles towards the end of the ramp up at a time of 90 ms. The synthetic reference is shown as solid blue line, the result from THEODOR as cyan dots, where no error estimation is available. The red solid line represents BayTh maximum a posterior (MAP) result with the red dashed lines for the two sigma confidence interval. The THEODOR result is withing the two-sigma band of the BayTh result for most of the profile with 150 pixel, but outliers are apparent. The AK model is considered a good choice for the expected profile shape.

Figure 8.5 shows the histograms for the deviations in heat load in figures 8.2 and 8.3. The reconstruction with THEODOR has a standard deviation of about $24\,\mathrm{kW\,m^{-2}}$, which corresponds to the expected value for 1 kHz time resolution and 25 mK temperature resolution according to equation (4.3). The reconstruction with the AK model leads to spatial regularisation, depending on the local signal-to-noise ratio, forward model and signal complexity. This reduces the standard deviation of the reconstruction to $11\,\mathrm{kW\,m^{-2}}$. The mean deviation is

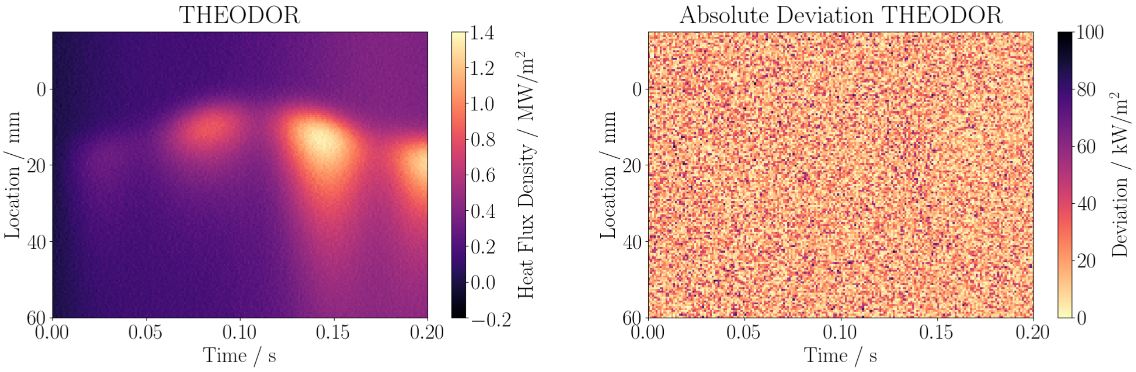


Figure 8.2.: Heat load obtained with THEODOR (left) and the absolute deviation to the reference (right) for a synthetic test case.

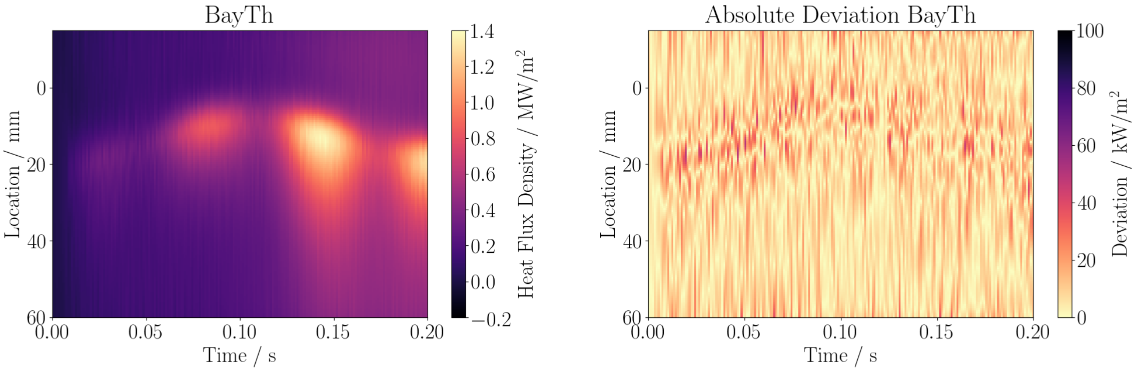


Figure 8.3.: Heat load obtained with BayTh (left) and the absolute deviation to the reference (right) for a synthetic test case.

both times well within the standard deviation. However, for THEODOR there is a systematic underestimation, although it is within the magnitude of the reconstruction uncertainty. This is due to the problems discussed in section 4.4. For example the insufficient spatial resolution to resolve the temperature distribution in the tile and the approach to obtain the surface heat load.

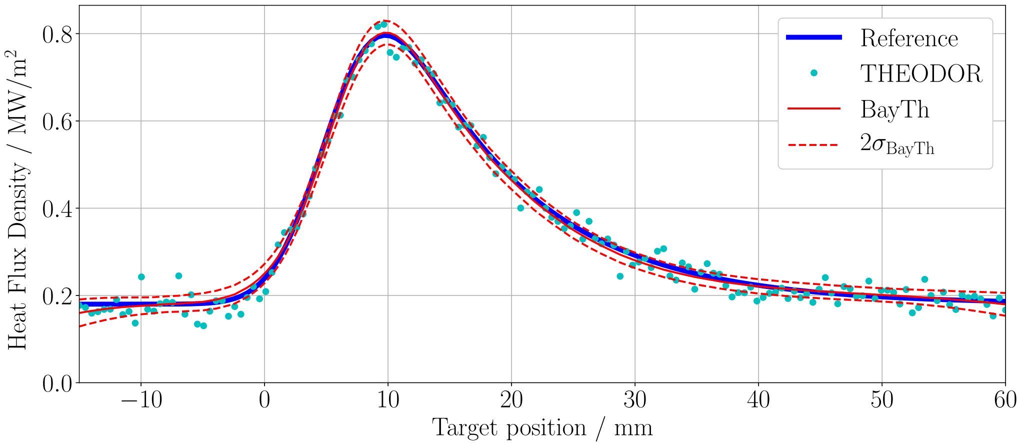
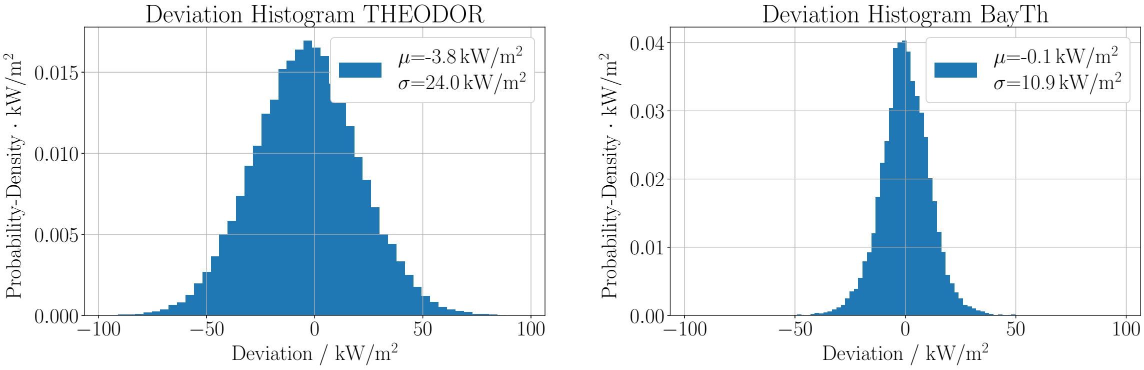


Figure 8.4.: Comparison of synthetic reference heat load (solid blue) to reconstructed profile from THEODOR (cyan dots) and BayTh (red).



(a) Histogram to the deviation of the THEODOR reconstruction in figure 8.2.

(b) Histogram to the deviation of the BayTh reconstruction in figure 8.3.

Figure 8.5.: Histograms for the deviation in heat flux density of the reconstructions with THEODOR (left) and BayTh (right) to the known synthetic profile.

A question arising is: how well does the kernel width represent the divertor spreading, which is often approximated by a Gaussian smoothing kernel? One could assume that the kernel width represents just this widening of the signal – though the regularisation is also needed for the ill-posed problem. Accordingly the hidden image would correspond to the upstream heat load profile, without smoothing or spreading. Figure 8.6 shows, that this view point does not hold. Initially, the information in the data is not sufficient to resolve features. Result is a very wide heat load profile for the first few profiles. At the peak position, the kernel width consistently is about 5 pixel. The underlying divertor spreading S is 5 mm in the presented range. Together with the resolution of 0.5 mm px^{-1}, this gives $S = 10$ px. The kernels seem to be a factor of two too narrow, however, the definition of the divertor spreading is a factor $\sqrt{2}$ smaller than the corresponding Gaussian point-spread function, see (2.2).

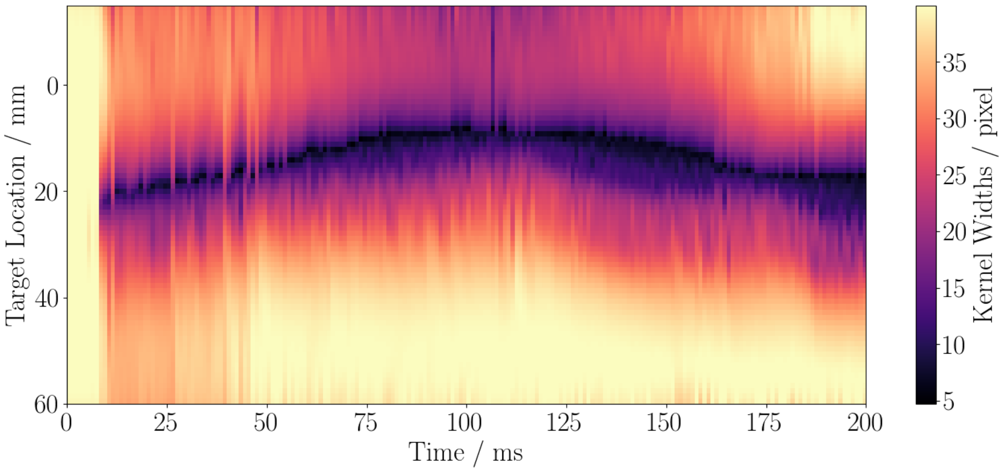
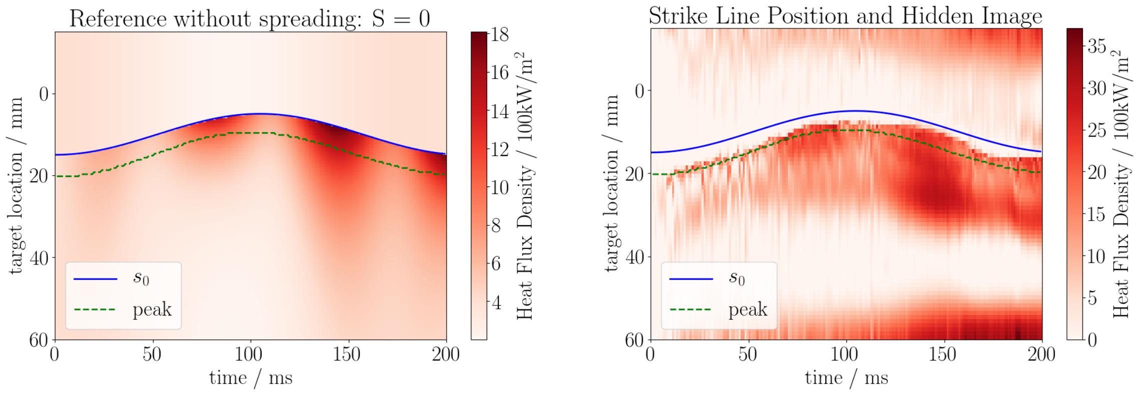


Figure 8.6.: Trace of kernel width, inferred together with hidden image in figure 8.7(b).

Figure 8.7 shows a comparison between the reference heat load before divertor spreading (left) and the hidden image of the AK MAP solution (right). The blue, solid line indicates the position of the strike line – the onset of the heat load before spreading, section 2.2. The dashed green line shows the peak heat load of the heat load profile, impinging on the target. One solution with the AK would be, to mimic the divertor spreading with the kernel width and recreate an exponential decay with the amplitudes. This naive view is not reflected in the inference results. Due to the smoothing property of the kernels, the peak in the hidden image is somewhere between the strike line and the peak of the profile. In addition, the spatially constant background heat load forces smaller kernel widths at the edges of the profile. This is necessary, in order to represent a flat profile at the boundary, given that the Gaussian kernels are being placed within the boundary.

Due to the variable widths of the kernel the amplitude density does not necessarily resemble the upstream profile, as seen at the time of about 150 ms. There the amplitude contribution in the



(a) Reference heat load pattern without spreading ($S = 0$).

(b) Hidden image.

Figure 8.7.: Reference heat load and inferred hidden image leading to figure 8.8(b) shown as time traces. The blue, solid line shows the reference value s_0, at which the exponentially decaying heat load arrives, without divertor spreading ($S = 0$). The green, dashed line indicates the peak of the reference profile with divertor spreading shown in figure 8.1.

centre is of the profile is larger than expected. However, together with the high contribution at the edge, the target profile is described well. In addition, the varying kernel widths has to be taken into account, when analysing the hidden image. A local reduction in kernel widths for a constant hidden image results in a peak in the model profile.

This implies, that the hidden image cannot be interpreted as resembling the upstream heat load profile. However, the edge of the profile is typically far enough from the structured part to not interfere, and is also weaker compared to it. Better understanding of the divertor spreading would allow to set the kernel width based on physical considerations, which would allow a more direct interpretation of the hidden image. So far, no knowledge about the profile shape, like the truncated exponential, is incorporated in the prior. The standard model of the entropic prior could be extended to express this.

8.2. Experimental data – determining lobe position

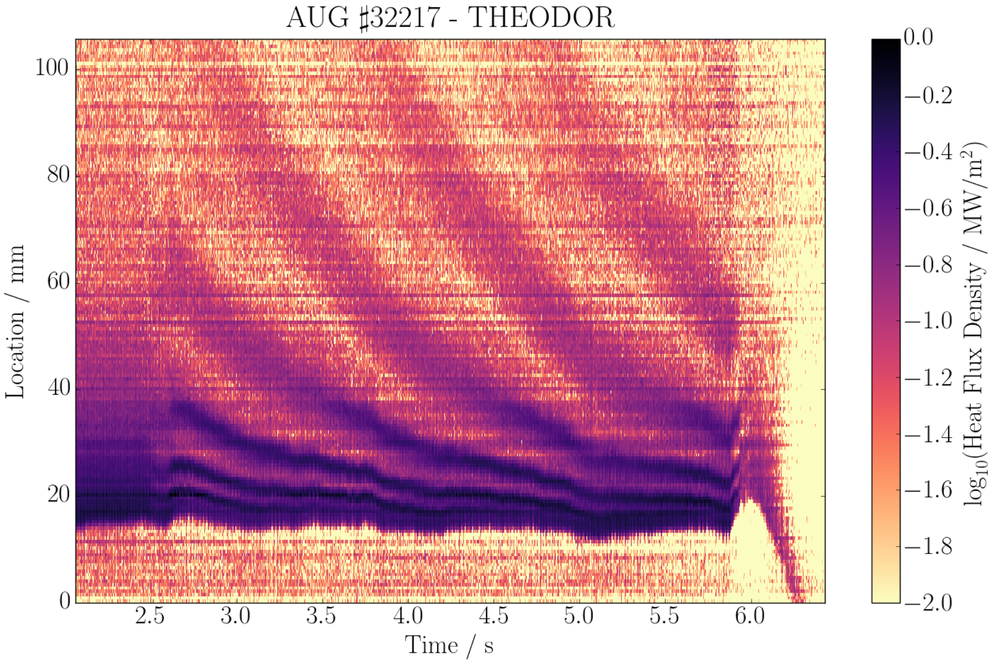
This section shows results of the inference with BayTh and compares them to results of the study in [Fai+17], which used the classic THEODOR to determine the heat load profiles. The mentioned study analysis heat flux profiles in the presence of external magnetic perturbations. One aspect is the position of features, which can be comparable to the heat flux resolution of the system. The perturbation of the toroidal symmetry of the magnetic field – which is a central property of a tokamak – leads to a feature known as strike line splitting. When turning the magnetic perturbation toroidally with respect to the position of the diagnostic, the 2D heat load pattern can be analysed. This cyclic change and the desire to learn more about the lobe position are both, a test bed for the AK and a valuable analysis. A brief introduction to magnetic perturbation and strike line splitting is given in section 1.1.5.

The aim of the cited study was to determine the position of the peak in the heat load profiles and compare them to the expected position obtained from a field-line tracing algorithm. To obtain the peak positions from the THEODOR heat load profiles a wiener filter along both, the space and time coordinates, is used. On the smoothed profiles local maxima are determined by comparing each cell with a heat load above a threshold to the two neighbouring cells.

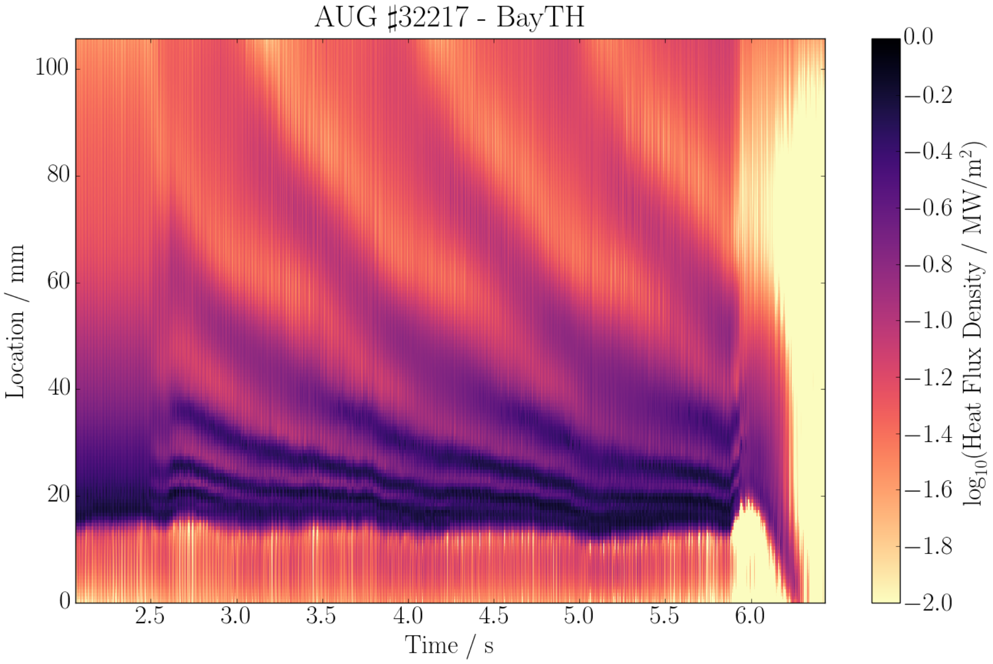
The heat flux density time trace of AUG discharge ♯32217 are shown in figure 8.8(a) – obtained by THEODOR – and 8.8(b) – obtained by BayTh – both on log-scale. The strike line appears similar, with the splitting clearly visible around a target position of 20 mm. However, as the heat load decreases into the scrape-off-layer (SOL) – upward in the plot – the SNR decreases and the lobe structure is much clearer in the result of BayTh. Also, a slight error in the calibration of the IR system is visible as brighter line at target location 100 mm for the THEODOR result. In the statistical analysis this deviation is within the uncertainty of the measurement and accordingly not attributed to a feature in the heat load. The heat flux density averaged over one turn of the pattern at the edge is on the order of some $10\,\mathrm{kW\,m^{-2}}$ but still the lobe position is well described by the AK model.

Figure 8.9 compares the deduced lobe positions to a model prediction, based on [Fai+17]. The ordinate shows the relative distance to the strike line in the range $(0 - 40)$ mm The abscissa shows the angle of the perturbation with respect to the diagnostic. As the mode number is two, the structure repeats after π. The black dots correspond to a model based on a field line tracing algorithm using a time resolved reconstruction of the magnetic field for the experimental pulse. The red dots indicate the position derived from the local maximum of the inferred heat load, by comparison to the nearest neighbour. For figure 8.9(a) a Wiener filter is applied to the THEODOR result, both in time and space. Figure 8.9(b) shows the same quantities, but with the peak-detection applied to the result of BayTh. No additional filter is applied, as the profiles are smooth by definition of the model.

For the THEODOR result a threshold of $250\,\mathrm{kW\,m^{-2}}$ is used to separate lobe signal from noise on the flat background signal. This limits the evaluation to effectively 30 mm on the tile. Even with this rather simplistic algorithm the strike line splitting is apparent at a toroidal angle of about $5/7\pi$ and the peaks are nicely scattered around the modelling result as the lobe evolves into the SOL. However, at the target location of about 20 mm an error in the calibration leads to false positives in the value between two lobes. For the probabilistic reconstruction the threshold is reduced to $70\,\mathrm{kW\,m^{-2}}$. The evaluation is thereby extended to the end of the tile, as shown in figure 8.10 – compared to figure 8.9 only the target location is extended. The kink at the toroidal angle of $3/4\pi$ and target location 75 mm is unexpected from the modelling point of view. Further analysis is needed to identify the reason for this behaviour. However,

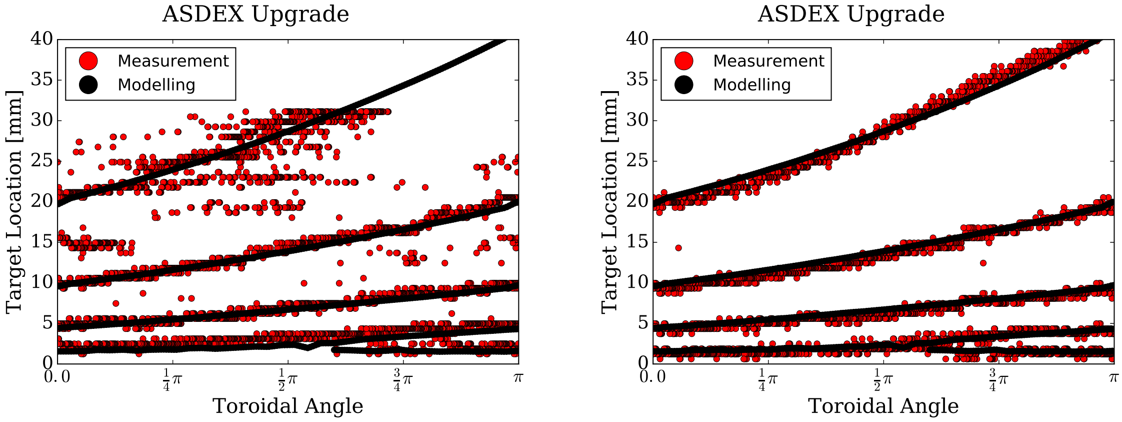


(a) Heat load determined by THEODOR, on a logarithmic scale.



(b) Heat load inferred by BayTh, on a logarithmic scale.

Figure 8.8.: Comparison of heat load profiles obtained from THEODOR and BayTh.



(a) Lobe position over angle of perturbation from THEODOR. Based on [Fai+17].

(b) Lobe position over angle of perturbation from BayTh.

Figure 8.9.: Comparison of the determined lobe position versus the angle of perturbation for THEODOR (left) and BayTh (right).

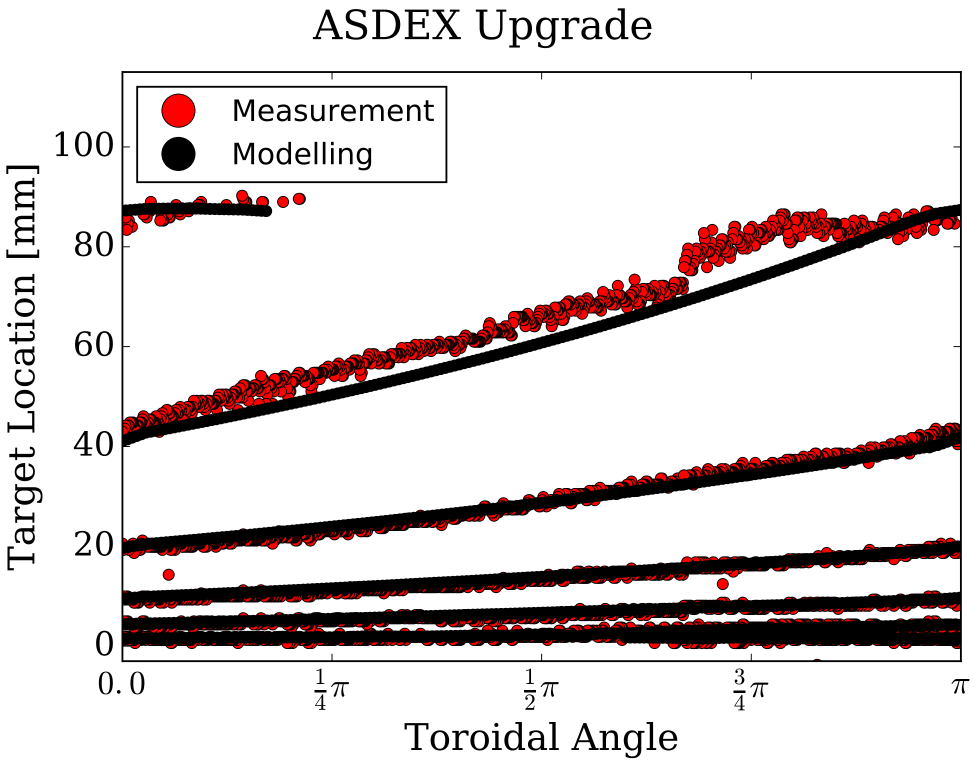


Figure 8.10.: Lobe position over angle of perturbation from BayTh for the full evaluation range in the SOL.

the kink is found in two full turns in discharge ♯32217 and is considered a real feature.

Discussion of lobe/peak position

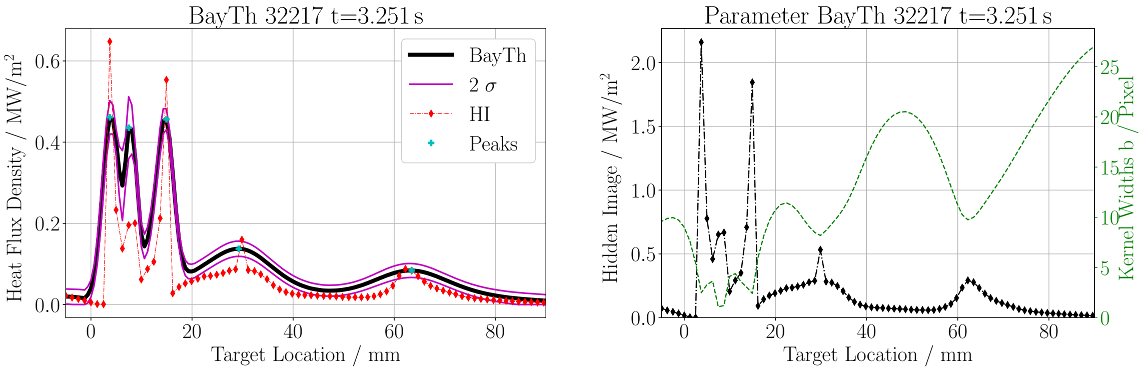
The field line tracer allows to determine the heat load pattern for pure parallel transport along the field lines. Additional perpendicular transport processes lead to a widening of the pattern. As neighbouring contributions can overlap, the peak of the signal is not necessarily the centre of a contribution.

Figure 8.11 shows a comparison between heat flux density profile and the hidden image for two different profiles. In general the local peak positions in the smooth profile – the black line with peaks emphasised by cyan dots in the left figures – does not necessarily coincide with a local peak in the hidden image (HI), represented as red dots, connected with a dashed line to guide the eye. The cyan marked peaks are used later on as the lobe position. The purple line shows the 2σ confidence interval of the MAP solution. However, as far as the position of the peak in the smooth profile can be determined, the peak in the hidden image is a good indication. Especially for areas with small kernel width – at target location (0 – 20) mm – the hidden image profile is often well peaked.

The right figures – b) and d) – show a comparison between the hidden image in black and the kernel width in green. The main difference between the two profiles is the peaking in the hidden image. In figure 8.11(a) the hidden image always has a clearly visible peak when the lobe position is unambiguous. This coincides with a local reduction in the kernel widths. In the far SOL in figure 8.11(b), at about 80 mm target position, the peak position is less clear and the amplitudes show a wider peak. At the same time, the kernel widths stays at a larger value, implying less structure.

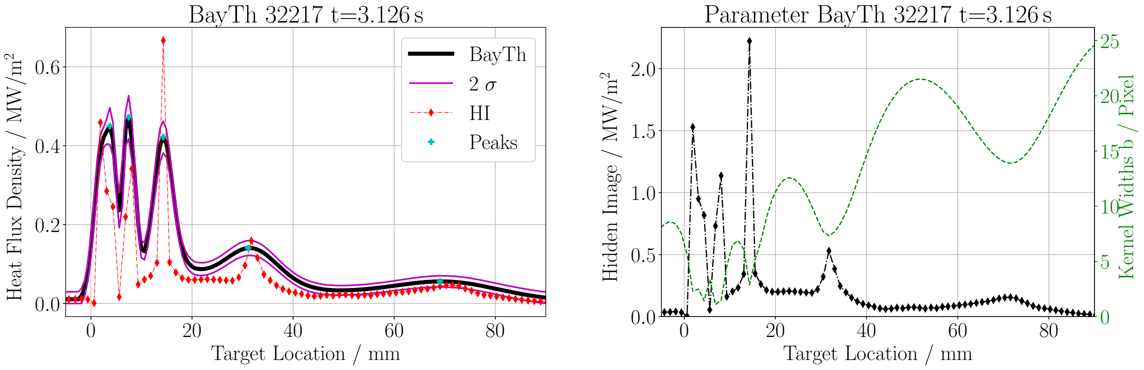
Figures 8.11(c) and 8.11(d) show the profiles at times $t = 3.251\,\mathrm{s}$ and $t = 3.3126\,\mathrm{s}$ from the time trace in figures 8.11(a) and 8.11(b). The last peak is less pronounced in the hidden image, as well as in the kernel widths. Here the data do not allow a better determination of the peak position, due to the lower SNR of the camera and intensity of the heat load.

Figure 8.12 shows a comparison between the results of the classic THEODOR and BayTh. The top frame shows a comparison of single profiles, also used in figures 8.11(c) and 8.11(d). The bottom part shows the difference between the two. Using the AK model raises the questions, if features might be flattened artificially. The cyan dots show the raw difference and appears unstructured. For comparison the difference signal has been smoothed using a Gaussian filter with a width of 3 pixel – magenta line – and 5 pixel – black line. For comparison: at the right-most lobe the kernel widths is about 20 pixel wide. The magenta line shows a minor oscillatory behaviour, but is mostly within two deviations. the black line is essentially zero. Especially at the strike line the deviations are essentially zero, due to the low kernel widths of few pixel and the resulting low smoothing. Even if there was more structure than is revealed by the AK model, it cannot be quantified within the uncertainty of the measurement. For



(a) Comparison between heat load pattern and hidden image. In addition, the local peaks are marked.

(b) Comparison between hidden image (black diamonds) and kernel widths (green, dashed line).



(c) Comparison between hidden image (black diamonds) and kernel widths (green, dashed line).

(d) Comparison between hidden image and kernel width.

Figure 8.11.: Comparisons between heat load profile and underlying hidden image from BayTh in a) and c) from two different times during experiment AUG ♯32217. b) and d) show a comparison between the hidden image (black) and the kernel widths (green).

this evaluation of BayTh the Kalman filter is used to correlate subsequent frames, which is not possible for THEODOR.

Further investigation is needed, to clarify the difference in figure 8.10 between the predicted and inferred lobe position. Possibly the difference is due to currents in the SOL, which are hard to quantify. These currents influence the magnetic geometry and could lead to such a behaviour. Examples can be found in [Rac+14a; Rac+14b]. Another uncertain parameter is the flux expansion, representing the mapping of the magnetic geometry to the divertor plates. This quantity is varying locally and could explain the discrepancy arising at above a position of 40 mm. However, the transition is expected to be smooth and cannot account for the jump.

For more details on the uncertainties of the inference of the magnetic equilibrium see [Ill18].

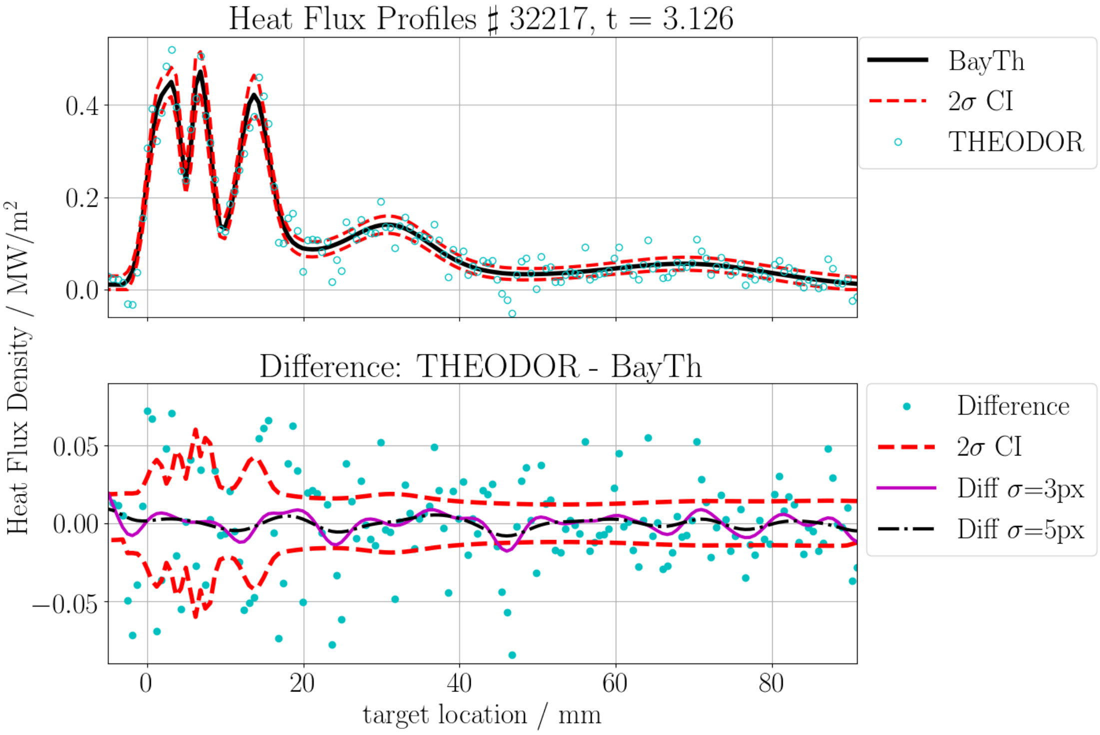
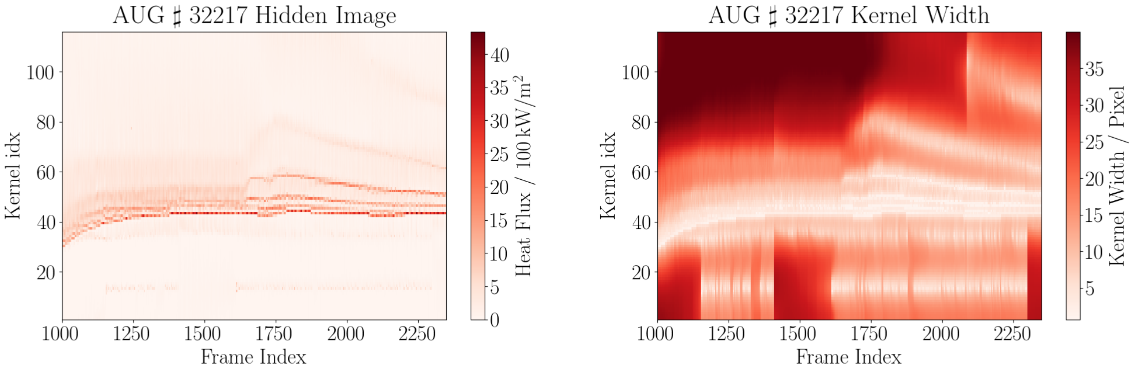


Figure 8.12.: Comparison between a single profile obtained with THEODOR and with BayTh. Top graph: comparison of the heat flux density, with the AK result as black solid line and the 2σ confidence interval (CI) as red dashed line. The cyan dots show the THEODOR result. Bottom graph: difference between the two profiles, with the 2σ CI as reference. Gaussian filtered difference in magenta and black – 3 and 5 pixel width.



(a) Hidden Image.

(b) Kernel Width Distribution.

Figure 8.13.: Parameters underlying the BayTh reconstruction in figure 8.8(b) shown as time traces.

Future work should include a forward model, mapping from the current distribution in the plasma to the heat load pattern on the divertor. This would potentially allow to improve the understanding of the plasma response to the external magnetic perturbation.

Figure 8.13 shows the hidden image (left) and kernel widths (right) composing the inferred heat flux profiles in figure 8.8(b). Before frame index 1600 the magnetic perturbation is not active, therefore no lobes are present. The tile surface is hottest at the strike line, when the perturbation is activated. The lobe structure leads to an increase of the surface temperature in time and the signal-to-noise ratio (SNR) locally improves. At about frame 2100 the kernel widths around index 100 decrease, resolving the lobe. The behaviour of the kernel widths and hidden image distributions behaves similar to the synthetic test case presented in section 8.1. Once the perturbation coils are active, the lobes are well pronounced in the hidden image distribution.

For further studies on the lobe position or other faint contributions, it appears more natural to look at the distribution of the hidden image, than to search for a local maxima in the heat profile itself. However, the kernel width in this analysis is not equivalent to the perpendicular spreading of heat in the plasma. This is shown in the synthetic data section above – see figure 8.7. In addition, the hidden image does not have to correspond to the heat flux distribution in the plasma distant to the target, but only represents the result of a deconvolution related to the plasma spreading. However, the position of a well localised contribution in the smooth profile – distant to the boundary – coincides within the local minimum of the kernel width. Figure 8.13 shows the time trace of the kernel widths. At the strike line the kernel widths is about 1 pixel wide, as the heat flux profile drops fast toward the private flux region. With magnetic perturbation the area with low kernel width increases in order to resolve the strike line splitting. At the lobes the kernel width is as low as is compatible with the SNR – which potentially leads to an overestimation of the feature widths and an underestimation of the peak heat flux. In the valleys between the lobes the kernel width increases.

The jumps of the kernel width in the around index 15 are related to a hot spot on the tile. This is an area with locally higher emissivity, possibly a dust particle [Roh+09], leading to an overestimation of the surface temperature. This hot spot is at the boundary of the resolution for the inference and hardly contributes to the heat flux profile. In the THEODOR result in figure 8.8(a) it is visible in log space. Finding it shows the potential of the AK model to resolving features contributing to a single pixel only, and that the evaluation needs experience with the system.

As already mentioned – 7.2 – the maximal kernel widths is limited to 40 Pixel, as larger kernel widths do not lead to an improvement in the smoothness of the profile and this value is only reached when there is virtually no shape contribution to the profile from this area. Higher values are detrimental for the numerical analysis, when computing the MAP.

The local kernel width cannot be seen as quantifier for the local spreading S as introduced in section 2.2. By construction the AK represent a deconvolution for a distribution blurred by a Gaussian kernel – which is an approximation to the transport in the plasma. However, the finite widths kernel width therefore represents an upper limit for S, while lower values can in general not be excluded. This is because of the limited accuracy of the data and the forward model, which limits the inference of such details. For magnetic perturbation this property of the adaptive kernel offers another source to infer the plasma temperature at the lobes in the divertor region. The local spreading is connected with the competing perpendicular and parallel transport in the plasma [Nil+19]. However, for plasma conditions like in this experiment at the far SOL the plasma temperature in front of the target is low. As indicated in the paper, for low plasma temperatures volume effects contribute to the target value of S. The introduces simple model cannot account for this.

This suggested approach to infer plasma parameters will probably only yield meaningful results in context of an integrated data analysis (IDA), combining measurements from various diagnostics. Such a system is available at ASDEX Upgrade, e.g. [FD04; Vie+12; Ill18; Ill18], but does not include IR based divertor heat loads yet.

9. Results for fast events

Fast events refers to events in the heat load in time, which are shorter than the sample interval of the diagnostic. They cannot be resolved with the classic THEODOR and similar tools. This section presents results of a synthetic study to resolve these fast events and first results on experimental data. The presented work is a proof of principle, as the model has to be refined further for more quantitative and reliable evaluations.
The experimental phenomenon in mind arises when analysing I-Mode data, which is an attractive confinement regime in between L- and H-Mode. Its characteristics tend to be a weakly coherent mode, leading to a decoupling of particle and energy confinement, and intermittent, bursty events at the mid plane [Hap+16a; Hap+16b; Man+17]. The time scale at the mid plane is about 10 µs. In the divertor volume of the plasma Bolometer data indicate times on the order of 100 µs.

9.1. Synthetic

For the synthetic test a heat flux density trace is defined and forward modelled on a 1 MHz time base. A burst arrives every 20.1 ms, which is enough to view them as independent, in the sense that the temperature distribution in the tile is close to a case with constant heat load in time. The additional 0.1 ms is used to scan different arrival times, relative to the sample interval of 1.0 ms.
Figure 9.1 shows the synthetic test case, on the left the heat flux in time and on the right the resulting surface temperature. The heat flux density is presented on log-scale, to see background and peak value. The background is a constant heat flux density of $1\,\mathrm{MW\,m^{-2}}$ and the burst adds $100\,\mathrm{MW\,m^{-2}}$ for 10 µs – so the effective heat load is $101\,\mathrm{MW\,m^{-2}}$. The peak temperature excursion is about 18 K and drops within 1 ms after the burst to about 1 K.

The rise and fall time of the surface temperature is connected to the time scale and transport process. The additional burst energy density of $Q_{\mathrm{burst}} = 1\,\mathrm{kJ/m^2}$ is initially localised on the surface of the tile. The heat is transported by diffusion into the depth of the tile quickly, leading to the rapid decrease of the surface temperature.
Figure 9.2 shows two examples for the reconstruction using THEODOR. Case A) (left) is for a burst arriving just after the last non-affected frame has finished recording. The reference surface temperature T_{Ref} and the temperature as deduced from the infrared (IR)-Signal T_{IR}

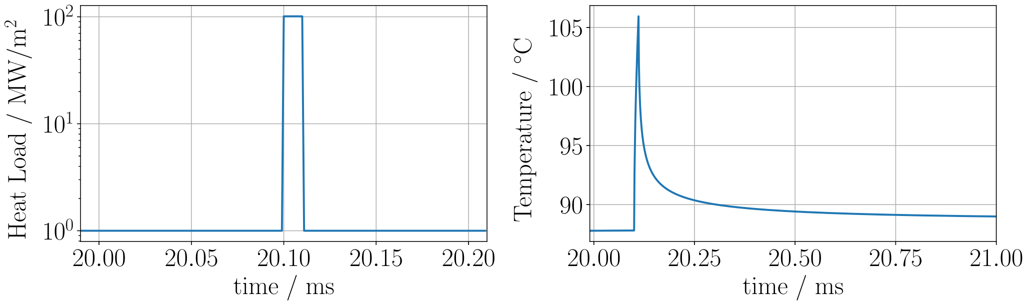


Figure 9.1.: Example of the synthetic heat load (left) and the resulting surface temperature excursion (right). The time intervals shown are not identical.

are shown on the top. Case B) (right) shows the same kind of burst with the same background heat load, but arriving during the integration of the camera. A comparison between A1) and A2) shows the difference in the inferred surface temperature of the first affected frame. The second line shows the heat load obtained by THEODOR on the synthetically measured signals. The dashed line represents the underlying ground truth. The discrete heat load density is presented as a constant contribution between the samples, which is the typical interpretation of the results. The most common case is an underestimation of the heat input during the first interval as shown in case A). This is compensated by an overestimation of the heat load density in the next intervals. When the measured surface temperature appears to rise quickly, the inferred heat load density in the interval leading to this measurement is overestimating the heat input of the burst – case B).

In the second case the additional energy is *removed* during the next intervals, thereby underestimating the background. When the burst has just the *right* timing, the burst heat input is captured correctly in the first THEODOR iteration, without significant over or underestimation of the following heat loads.

The burst in case B) arrives from (180.90 – 180.91) ms while the integration is performed during (180.90 – 181.00) ms. In this case, the camera sees the sharp increase and decrease of the temperature. The shorter the integration time of the camera, the larger the deduced peak temperature can be.

The over-/undershoot and the the following under-/overestimation in the heat load are independent of the background heat load. The integrated heat load of THEODOR from the start of the burst to some frames after it minus the known background gives a good estimate of the heat input of the burst. However, typically there is no known, constant background to subtract. The deterministic THEODOR has the property of estimating the integral heat influx reliably, over several frames. This is used in practice to estimate the integrated heat input of intermittent events like ELM's instead of using the peak heat load density [Eic+11].

This example shows, that fast events cannot be reliably analysed using the deterministic

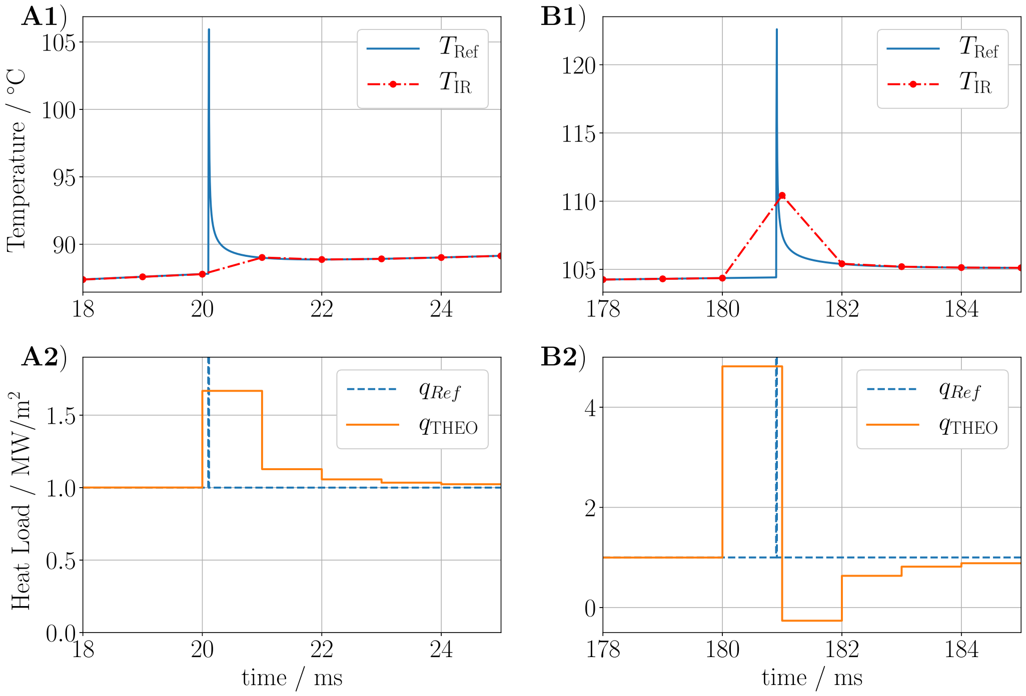


Figure 9.2.: Two examples for the reconstruction using THEODOR. Case A (left) is for a burst arriving just after the last non-affected frame has finished recording. Case B (right) is for a burst arriving right before the integration of the camera. The top row shows the surface temperatures from the reference and as inferred. The bottom row shows the inferred heat load from THEODOR.

approach behind THEODOR and similar tools. The only possibility to infer the heat input, is to separate the burst from the background and ask for the separate contributions. When the arrival time of the burst is not well known, it has to be inferred from the IR data as well.

Figure 9.3 shows the reconstruction using Bayesian THEODOR (BayTh), which distinguishes between a constant background and the short pulse. The time is relative to the start of the analysis and corresponds to case A) in figure 9.4, with an offset of 20.1 ms. The areas used for the IR signal are marked green. In the used parameter sample the additional burst-heat-load is distributed over about 180 µs, being distributed over two of the time discrete solver intervals.

Figure 9.4 shows on the left the histogram for the burst energy E_{B} obtained using a Markov Chain. The ground truth is $1\,\mathrm{kJ/m^2}$, with which the inference result is compatible. On the right the underlying two dimensional posterior distribution for the burst heat load versus the burst duration is shown. The burst energy is the product of heat load and duration. The two dimensional histogram includes a line indicating the obtained mean constant energy together with the 2σ confidence lines as dashed lines. The ground truth is a burst heat load of $100\,\mathrm{MW\,m^{-2}}$ with a duration of 10 µs. The analysis indicates a shorter burst duration,

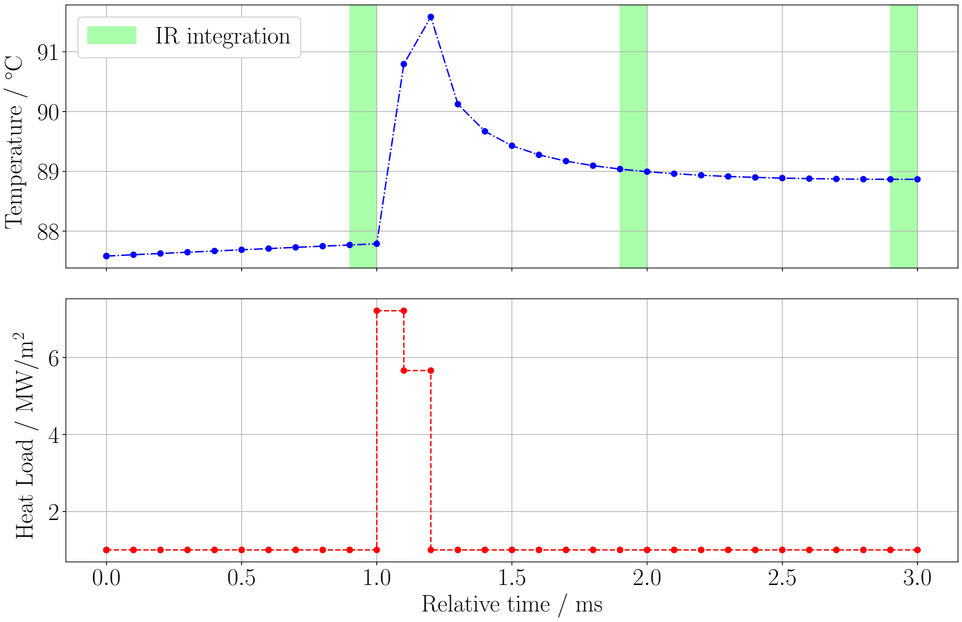
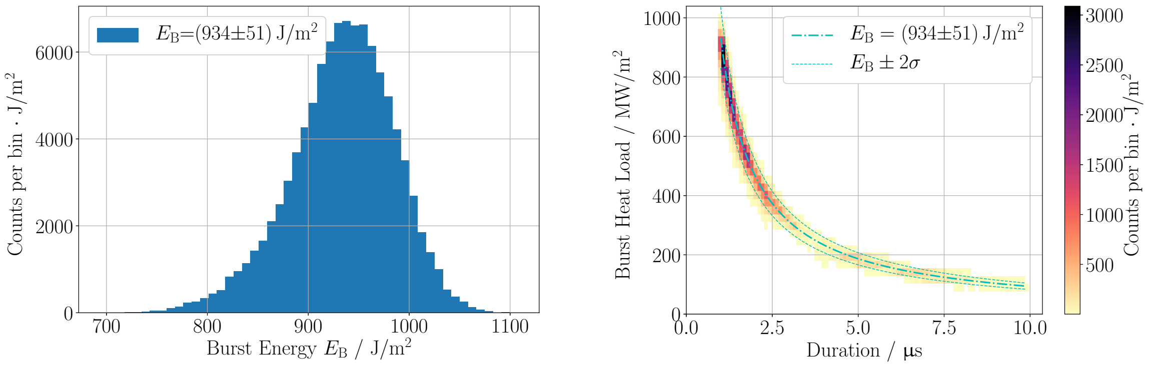


Figure 9.3.: Details for burst analysis. Top: The surface temperature, with dashed line showing the linear interpolation used for the measurement. Bottom: surface heat flux density applied during the discrete steps.

which is considered to be an artefact of the used numerical solving schemes. The burst duration is underestimated, but the energy input is well described and within two standard deviations to the ground truth. The time step for the numerical solver was a tenth of the measurement interval, using ten-times oversampling of the HDE solver. The integration of the diagnostic uses linear interpolation. The burst heat load is applied according to the same weighting scheme. This method was used to prevent a change in the precision of the solver with changing time steps.



(a) Histogram for the burst energy E_B as product of time and intensity.

(b) Histogram for the intensity in MW m^{-2} versus the duration in µs.

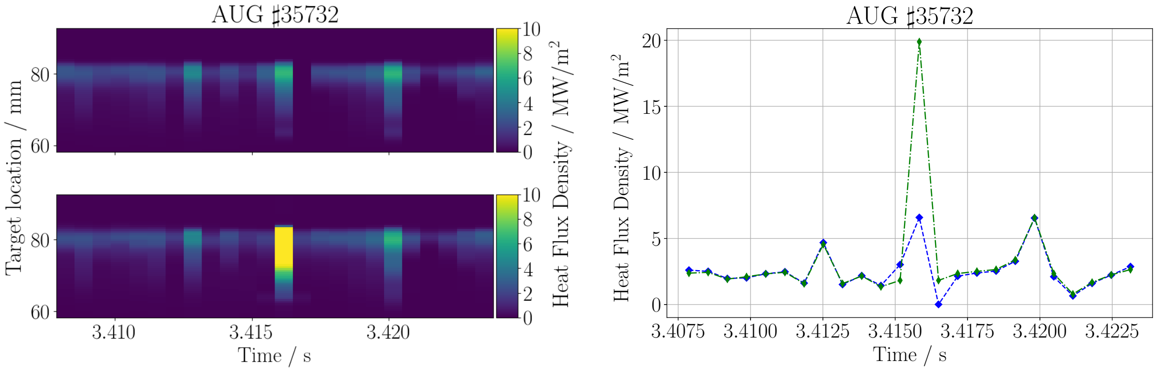
Figure 9.4.: Comparison of the histograms showing the distribution of the burst energy E_B(left) and its factors burst heat load versus burst duration (right).

9.2. Measurements

This section provides the first application of the method, tested in the section before, to experimental data.

A small set of burst events were evaluated from AUG discharge ♯35732, and one of them is presented here. The sample rate of the IR camera was 1.5 kHz, corresponding to a sample distance of 0.66 ms. Figure 9.5(a) shows a comparison of the surface heat flux density evolving in time. The abscissa represents the time, thy ordinate the target location. The colour scale is limited to an upper value of 10 MW m^{-2}, which caps the burst contribution, but resolves the behaviour of the background. The top image shows the result using the standard BayTh analysis, setting the heat load constant between data samples. The bottom image shows the result when using the burst analysis. The burst-heat-load is presented as wide as the other heat loads, but is actually shorter, with the background being shown before and after the burst. Figure 9.5(b) shows the same event as figure 9.5(a), but as 1D profile in time evaluated at the strike line. In the THEODOR evaluation the heat flux after the burst drops below zero, which is prevented in BayTh via the prior enforcing non-negativity. The heat load being smaller before the burst indicates that the constant background value before and after the burst might be not a good model. Still, the difference indicates the expected underestimation of the following heat loads, without using another model for the burst. The heat flux density for the sample with the burst analysis presents the sum of the background plus the burst. The energy density of the burst is about 18.2 MW m^{-2} $\cdot$ 100 µs $\approx$ 1.8 kJ/m^2. This is small compared to ELM's at ASDEX Upgrade with about 30 kJ/m^2 [Eic+17].

While the initial expectation of the burst duration was about 100 µs, more recent analysis indicates a burst duration of about 500 µs. This however is still less than the time between the two samples. The analysis of the presented burst indicates a duration of about 100 µs. Given the results of the synthetic test cases, this is expected to be a lower boundary, though it is resolved by the numerical scheme. However, the heat input – in terms of energy density – is expected to be reliable.



(a) Comparison between heat load trace without (top) and with (bottom) burst analysis.

(b) Comparison of the heat load density without (blue) and with (green) burst analysis from BayTh at strike line.

Figure 9.5.: Comparison of the standard BayTh evaluation and BayTh with the burst analysis method. The heat flux density of the burst refers to the effective heat flux with burst.

Figure 9.6 shows the posterior for a scan of the start time of the burst. For events with durations on the order of 100 µs the arrival time analysis seems reliable within 50 µs. For start time beyond 0.3 ms the burst extends until after the measurement and the influence of the burst heat load on the posterior decreases. This typically results in an exponential increase of the negative log-posterior, in this case reaching values beyond 10^6 for the arrival tenth of µs before the first affected measurement. This is in parts due to the constant background model, which describes the three measurements together with the burst.

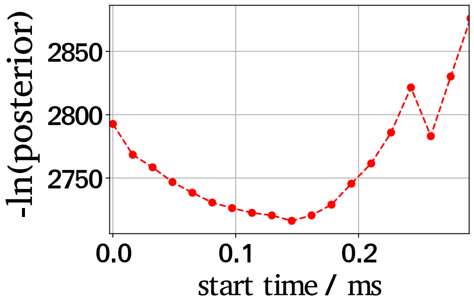


Figure 9.6.: Example for the negative log posterior from a scan of the start time of the burst, relative to the end time of the last unaffected measurement.

9.3. Summary and outlook

The presented analysis allows to infer the event duration – given a reasonable model of the time behaviour – peak heat load and the resulting energy transferred to the wall. Even the arrival time can be estimated solely from IR data. However, additional data from other diagnostics would improve the inferred arrival time, which in turn also benefits the uncertainty of the other parameters.

Also for other events like ELM's, which show an exponential decay in time, this method can provide a better evaluation of the transferred energy. Typical studies determine the decay time based on a few profiles obtained from the classic THEODOR. The decay time in JET is reported to be in the range (262 – 528) µs [Eic+11]. The time-behaviour of ELM's is reasonably well known, but the energy is typically determined via the time integral during one decay time, for which no uncertainty estimation is available. An approach like the one presented here would allow a better quantification of the uncertainty of the event parameters.

A higher time resolution of the solver can improve the results. However, a trade-off between precision and calculation time has to be found. The choice depends mainly on the time scale of the fast events and the sample rate of the diagnostic. A potential improvement would be to include the uncertainty of the solving scheme in the likelihood.

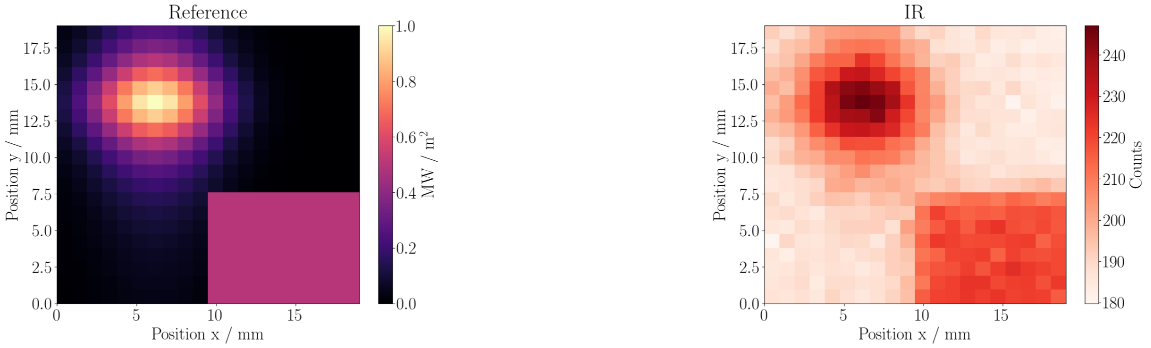
10. Results from the 2D extension

10.1. 2D adaptive kernel and heat load profiles

This section is based on the publication [NT19].

The results of three approaches for inference with the 3D heat diffusion equation (HDE) solver as forward model are presented. The approaches are to compute the maximum a posterior (MAP) using singular value decomposition (SVD) and stochastic trace estimation (STE), and Markov chain Monte Carlo (MCMC). The figures in 10.2 show the results for the same data for comparison. The properties of how to the adaptive kernel (AK) model extension to 2D is discussed.

For this comparison, a static reference 2D test heat load is applied to obtain synthetic IR data. Figure 10.1(a) shows the reference profile with 20 × 20 pixels, to which a dense set of kernels – one for every pixel – will be applied. Starting with a tile at equilibrium at 80 °C, the resulting synthetic data including noise after 50 ms are shown in figure 10.1(b).



(a) Reference 2D heat load pattern, applied in time.

(b) Synthetic measurement after 5 steps with 10 ms length.

Figure 10.1.: a) Reference heat flux pattern. Peak: along x-axis a Cauchy distribution with width of 5 pixel, along y-axis a Gaussian with width of 6 pixel. Peak set to $1\,\mathrm{MW\,m^{-2}}$. The bottom right corner is overwritten by a plateau of $0.5\,\mathrm{MW\,m^{-2}}$. b) Resulting IR data– in counts – after 50 ms exposure to the heat flux pattern shown in a).

10.1.1. MCMC

Classic Metropolis Hastings Monte Carlo provides the reference results for the distributions of the amplitudes and kernel widths. For 1200 parameters the integration is still feasible, see section 10.1.4. Figure 10.2(a) shows the inferred heat flux pattern in the top left. In the top right is the spatial amplitude distribution. In the bottom left and right are the distributions of the kernel width along the horizontal axis – σ_x – and the vertical axis – σ_y. At the edges of the step function the kernel widths are lowest, reflecting the sharp transition.

10.1.2. Singular value decomposition – SVD

To evaluate the Gaussian approximation for the kernel widths – see section 6.1.2 – the SVD is used, which is available as a robust algorithm. Run time scales typically with $\mathrm{O}(n^3)$, which is feasible for *small* systems of around 100 kernels – resulting in a 100 × 100 matrix – with calculation times on the millisecond scale. For the shown test system with 20×20 kernels, the SVD for the 400×400 matrix takes about 100 ms. The kernel widths are lower at the edges of the step in the bottom right than for the MC result. As the kernel width acts as regularisation and the resulting heat flux pattern is virtually identical, the Gaussian approximation for the kernel widths is justified. This provides the additional benefit of a speedup from about 10 hours to about 10 min per frame.

The total memory demand including automatic differentiation is 5.0 GB and the time per function call including gradient evaluation is (1600 – 2000) ms. Without the gradient information, the values are 70 MB and 100 ms.

By increasing the image resolution by a factor of 2 – leading to 40 × 40 pixels and kernels – the computation time without gradients increases to 13 s and the overall memory demand is 730 MB. With gradients, the memory demand exceeds 120 GB, which is the upper limit on the computer system used.

10.1.3. Stochastic trace estimation – STE

For larger matrices the SVD becomes too expensive in terms of time and memory consumption. An alternative is to sample the matrix with test-vectors, in order to estimate logarithm of the determinant (log-det). This is known as STE and explained in section 5.4.2.

The resulting distributions are shown in figure 10.2(c), which come close to the results of the SVD. The downside is, that the number of test-vectors and the order of the expansion has to be set a-priori for the optimisation procedure. For the shown example the expansion order is 10 and 50 test-vectors have been used.

Using automatic differentiation, the total memory demand is 4.7 GB and the time per function call is about 2000 ms. Without the gradient information the values are 70 MB and 100 ms. The computation time and memory demand scale linear with both, the number of test-vectors

and expansion order.
For larger systems the memory demand and computation time can be controlled by the order of the expansion and number of test vectors used. In a case with 40 × 40 pixels and kernels – overall 4801 parameters – the time raises to 10 s and the memory to 730 MB without gradients. Including gradients, the computation time increases to 20 s for 10 vectors and 10 expansion orders, the memory demand is about 200 MB per test vector and expansion order. Intermittent evaluation of the gradients is possible to free the memory, as the results are independent. For 20 vectors and expansion order 10, the memory demand is just above 100 GB.

Remark 1

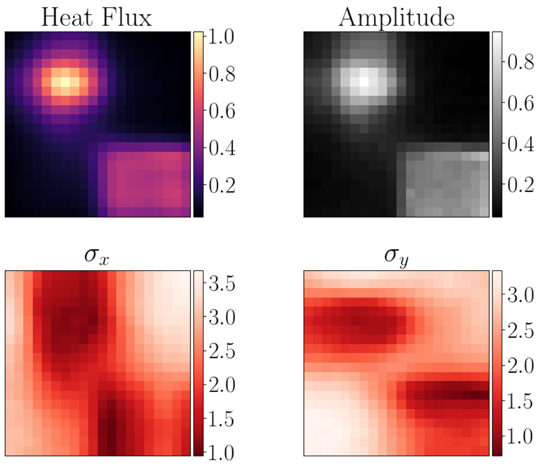
For large matrices, the matrix-matrix product $\mathbf{B}^T\mathbf{B}$ turns out to be memory consuming in the adjoint formulation. More memory-efficient implementations become necessary for increasing matrix sizes beyond 1000×1000 – here about 20 GB are reserved for the matrix multiplication. An improvement could be, to run the optimisation on a subset of the parameters and obtain only the corresponding gradients. When memory consumption prevents the use of automatic differentiation, finite difference can used instead. The performance difference was not investigated so far.
In addition to the SVD and STE, a conjugate gradient method was tested as well. However, due to the iterative nature the determination of the gradients mostly fails. In addition, although the precision of the result can be controlled, jumps in the log-det for slight changes of the kernel matrix prohibit the use in an optimisation routine. This fact limits the alternatives to the SVD for large matrices – to the authors knowledge – to the STE approach.

Remark 2

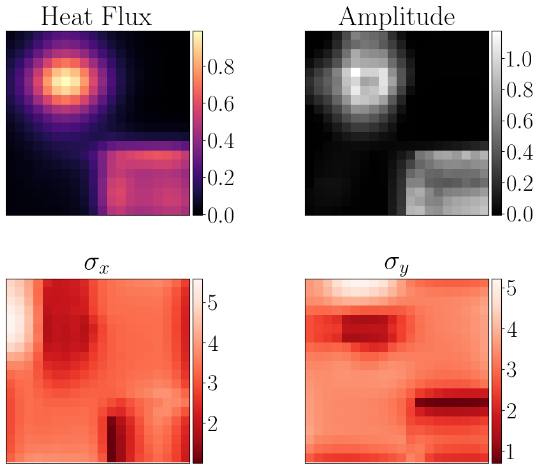
For the IR system it seems natural to place a kernel for every pixel. This however is probably not needed for most of the cases, especially when the profile has some known decay lengths. An example is the heat flux which may vary fast – on a few pixels basis – along the poloidal orientation, but slow along the toroidal orientation. For application to large data sets the number of kernels needed should be investigated beforehand to reduce the problem size.

10.1.4. Computation time

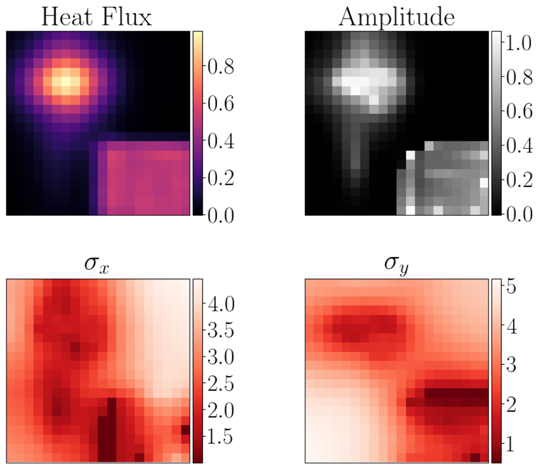
For the shown example with 1200 parameters the number of function calls from a *standard* parameter distribution – constant values for hidden image and kernel widths, initial likelihood about $3.4 \cdot 10^4$ for 400 data points – is on the order of some thousands. On the example shown, the bottle neck is the evaluation of the log determinant, independent of the method used – with some seconds for the full evaluation. The time needed for the optimisation is therefore on the order of (5 – 60) min per frame. When the starting point in parameter space is close



(a) Result using MCMC



(b) Result using SVD



(c) Result using STE

Figure 10.2.: Comparison between the three approaches. Structure of subplots: top left – Heat Flux Pattern. Top right – amplitudes of each kernel. Bottom left – kernel width along x-axis. Bottom right – kernel width along y-axis.

to the MAP solution, this time can drop to about 1 min. This is typically the case for similar consecutive frames, where the last MAP solution is used as new starting point.

An alternative is to use MCMC to explore the parameter space. This circumvents the calculations of the log-det all together. The run-time of the forward model for a reasonable system size is on the order of 1 s – without the need to compute the kernel matrix explicitly and taking the system response into account. However, for 20×20 kernels we already have to deal with 1200 parameter. Using 1000 sweeps – each representing a sequential scan through the parameters – and 10 bins leads to about $10^3 \cdot 10^3 \cdot 10 = 10^7$ function evaluations. Assuming a run-time of 1 s per call, this sums up to 116 days of computation. For the shown example, the forward model evaluation took about 2.5 ms, which corresponds to about 7 h of total sampling time.

On this scale minimisation seems to be the only feasible way, although the log-det calculation becomes cumbersome. Alternatively Hamiltonian Monte Carlo can make use of the gradient information, which speeds up the process significantly. However, the comparison presented here is based on results from classic MCMC.

10.2. Comparing adaptive kernel and adaptive diffusion

In two dimensions the evaluation time for the AK grows significantly as outlined in the modelling section 6.1.5. As alternative it is proposed to use the diffusion operator for regularisation of the hidden image. Given a small test case – where the AK including the determinant can be evaluated with gradient information – a comparison is made and presented.

The forward model in this section is a simple linear operator: a Gaussian smoothing kernel with $\sigma = 5$ pixel. As before, a 20×20 grid is used.

Figure 10.3 shows the ground truth on the left, which is defined via a Gaussian of width $\sigma = 3$ pixel along the y-axis and a Cauchy of width $s = 3$ pixel along the x-axis. The peak is normalised to 1000 counts. In the bottom right corner this distribution is overwritten by a flat plane of height 500. The middle image in the figure is the result from applying the smoothing operator, which is treating the surrounding to be equal to zero, hence the losses at the end of the step. In the right image is the final image used as data for the inference, with Gaussian noise according to the counts.

Figure 10.4 shows a comparison between the results of the adaptive kernel (top) and the adaptive diffusion (bottom). The left part shows the model for the quantity of interest, again overlaid the lines for the profiles. The middle and right graphs show the profiles along the x- and the y-axis, colours as in figure 10.3. The light colour is for the ground truth. The profiles are similarly resolved, the ringing at the step being the largest difference. The overshoot at the step is smaller for the AD than for the AK, as the model is better suited for such features. By construction the AD model allows two separate constant levels next to each other, by setting only the conductivities along the interface to low values. Though the specific level

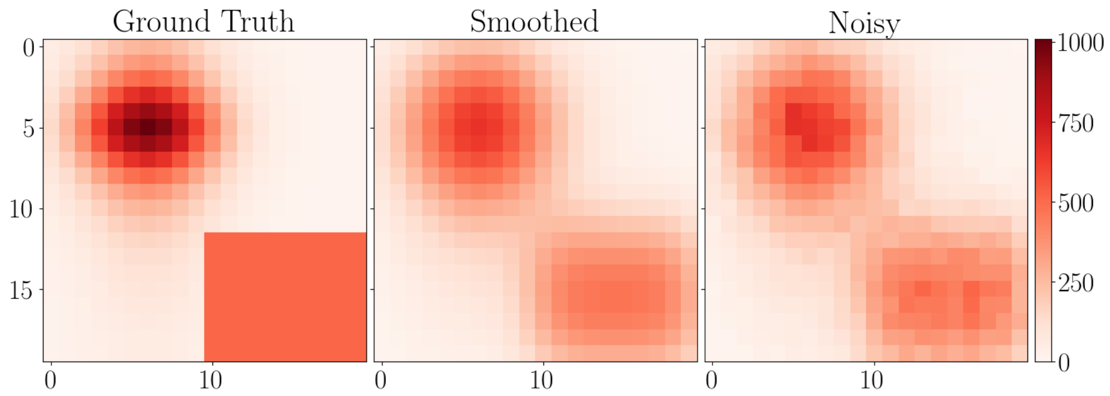


Figure 10.3.: Ground truth (left), which is to be restored after applying a smoothing operator (middle) and adding normal distributed noise (right).

shown for the AD model is above the ground truth, the neighbouring profiles are at a lower level. Due to the forward operator they are anti-correlated, so that this model still agrees well with the measurement. The adaptive diffusion shows a flattening at the boundary of the domain, which is due to the reflecting boundary condition. A steeper behaviour would require a lower conductivity in the boundary area, to prevent a sort of piling up. Depending on the problem at hand, the boundary conditions can be chosen adequately, or be inferred.

Figure 10.5 shows another comparison for the adaptive kernel (top) and adaptive diffusion (bottom) results. In the left column the deviation from the reference is shown, which is clearly strongest around the step. The corresponding regularisation parameters (kernel width σ and conductivity κ) are shown in the middle for the x-axis and in the right for the y-axis. Interestingly, the kernel widths are getting sharper in the top left, where the peak of the Gaussian is located along the x-axis. In contrast, the diffusive kernels are getting smoother in this area. The DK deviate slightly stronger from the ground truth – c.f. figure 10.3 bottom right – than the AK for this peak – in the top right.

Note: While the AK are defined via the widths of the kernel, the conductivity sets the regularisation for the AD, which is here associated with single cells. For the solver this quantity is of interest at the interfaces between cells. It is suggested to use the conductivity of the interface as parameters, but for comparison to the AK model this was changed. The parameter refers to the cells, with the interface values being defined as the harmonic mean of the two adjacent cell conductivities for the relevant direction, see equation (6.51). Just like the kernel widths reach their lowest values at the step in the profile, the conductivities approach zero there as well. This allows for a steep step, as the areas are basically decoupled for zero conductivity. This case corresponds to a pixel-wise model, locally.

The performance is significantly better for the adaptive diffusion (AD) model, as a few steps of the implicit solver use less operations than evaluating the 2D kernel function for all kernels

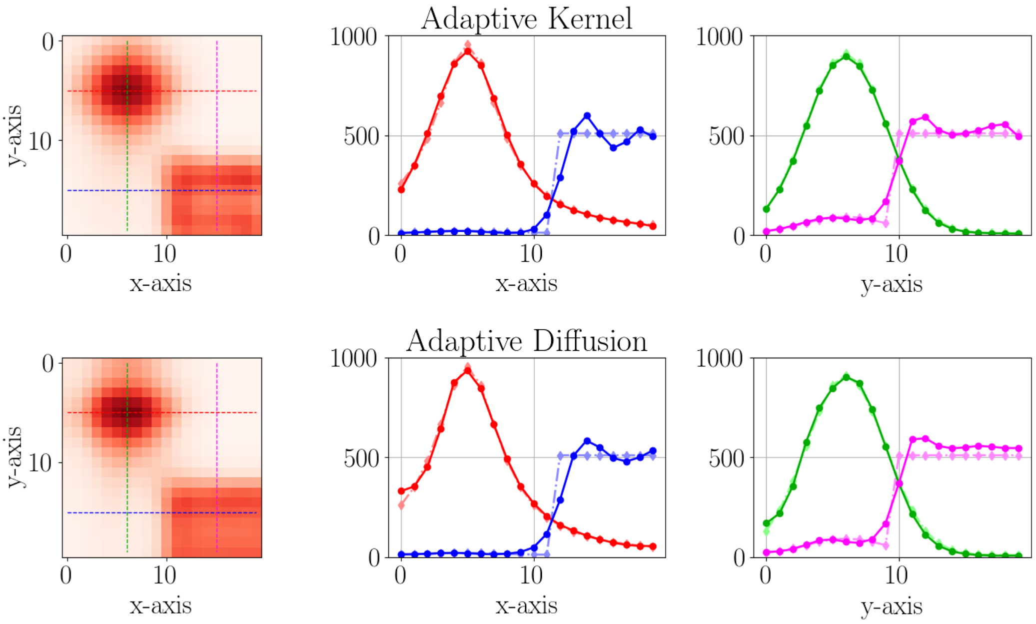


Figure 10.4.: Top: adaptive kernel results. Bottom adaptive diffusion results. Left: MAP solution from diffusive kernel model. Middle and right: cross-sections as shown in 10.3

on potentially every cell. The evaluation time for the 20 × 20 cells dense fields is 0.01 ms for the AD model compared to 1 ms for the AK model, this is two orders of magnitude difference. Optimisation potentially can improve the performance of these models, but by construction the finite difference diffusion solver is much more efficient in higher dimensions.

Downside of the adaptive diffusion model is, that a Markov Chain – or similar methods – have to be used, as the Hessian is not obtained easily for the Gaussian approximation. On the one hand, for more complex forward models the evaluation of the adaptive diffusion or adaptive kernel model can become negligible. On the other hand, for strongly non-linear models the Gaussian approximation might become insufficient, so that an MC routine has to be used anyway.

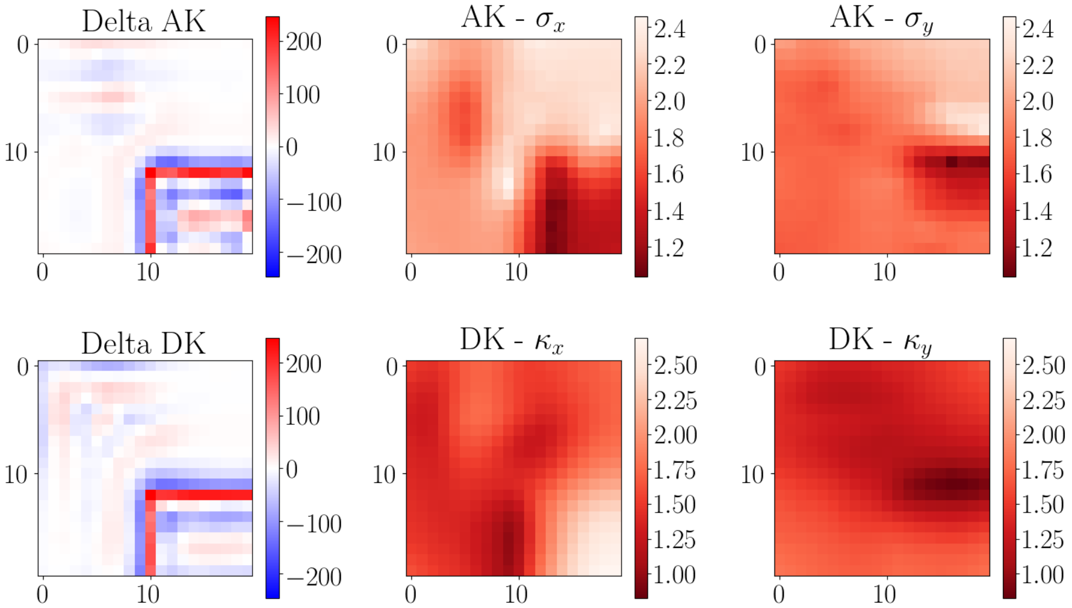


Figure 10.5.: Top: adaptive kernel results, Bottom adaptive diffusion results. Left: Difference to ground truth. Middle: Regularisation parameter along the x-axis. Right: Regularisation parameter along the y-axis.

10.3. Conclusion and outlook

Inferring two-dimensional scalar fields in the scope of ill-posed problems can become numerically prohibitive expensive. When a spatially resolved regularisation is desired, the adaptive kernel model provides model selection via Gaussian approximation, which is efficient for small fields. For larger fields and non-linear forward models, adaptive diffusion can become an interesting alternative.

Both of these models can facilitate an adaptive resolution, in order to reduce the parameter count. Linear interpolation of the hidden image and regularisation parameters could be used to start with an under-resolved model and later refined where needed.

11. Summary and Outlook

This chapter presents a short summary of the work this thesis is based on and the conclusions drawn. An outlook with some ideas for further studies is given at the end.

11.1. Summary

An analysis of the inverse problem of inferring heat loads to surfaces from surface temperature information is presented. Commonly used finite difference solving schemes for the forward model are compared and the final choice is justified with the available data quality in mind. For the statistical analysis of long time traces of data for this time-dependent problem a truncation method is presented, allowing to analyse the data sequentially.

The adaptive kernel (AK) model is used to represent the heat load profiles – the quantity of interest. It allows spatially varying regularisation, being especially suited for a scenario with dynamic signal-to-noise ratios of the data. In one dimension – validated by using synthetic data – it is shown that the adaptive kernel model provides as much regularisation as is appropriate for inferring heat loads profiles. The inferred credibility range is reliable for narrow features as well as for a flat background.

The inference is performed using either optimisation routines to find the maximum a posterior (MAP) solution or Monte Carlo routines. To select the best kernel widths for the MAP, the logarithmic matrix determinant is needed. As an alternative to the singular value decomposition (SVD) to determine the ideal kernel widths distribution a stochastic trace estimation (STE) is presented and tested. STE allows to determine the MAP solution for larger model profiles. However, beyond 400 kernels with the SVD or 1500 kernels with the STE the memory demand using automatic differentiation is at about 100 GB, limiting the applicability.

An extension of the AK model to two dimensions is introduced and applied to synthetic data. Although the implementation of the expansion was successful, it is considered too time demanding for the routine analysis of thousands of profiles.
As alternative model with adaptive regularisation the adaptive diffusion (AD) model is introduced and a comparison to the AK model is made. In two or more dimensions the cost of evaluating the AD model is significantly lower. However, as the precision matrix cannot be obtained efficiently and not processed for large problems, time consuming Monte Carlo methods have to be used for the inference. The estimated regularisation is comparable to the

AK model.

A new approach to resolve events shorter than the sample interval of the diagnostic is presented and tested. By using a set of three consecutive data profiles the short event can be distinguished from the background heat flux density profile.

11.2. Conclusion

For analysis tools like THEODOR, the uncertainty of the measurement typically dominates the analysis of experimental data. Instead of spending resources for optimising the numerical integration of a PDE, a statistical analysis with sufficiently precise tools – like the underlying finite difference solver – pushes the analysis to a new quantitative level. Using adaptive kernel in conjunction with automatic differentiation for model selection gives good computational performance and access to uncertainty estimation.
This can be used to determine the position of contributions in heat load profiles, which are at the level of the signal-to-noise ratio (SNR). An example is strike line splitting in presence to external magnetic perturbation.

The AK model is a potent multi-resolution approach for 1D and 2D profiles, revealing structures supported by the data while preventing over-fitting. For 1D profiles the determination of the MAP solution to the non-linear, time-dependent problem is fast enough to treat thousands of profiles in a reasonable time. About one minute on a single core machine per index, including the estimation of the confidence interval, allows to treat a thousand profiles sequentially withing a day. Longer traces can be split, analysed in parallel, and finally merged, to keep the wall time at about one day. The obtained confidence intervals allow better interpretation of weak contributions, like from strike line splitting. This can be the basis for further investigation into the effects of external magnetic perturbation onto the scrape-off-layer (SOL) and e.g. the shielding properties of the plasma. This is also important for the understanding of transport processes in the plasma.

For fast events – being shorter than the sample interval of the diagnostic – integrated quantities can be estimated with the statistical approach. The inference of the energy density of short events was solely based on infrared (IR)-data in this thesis. As the intensity and duration of the events are anti-correlated, they cannot be well resolved with this approach. Also the arrival time is not estimated easily. However, including further diagnostics can provide additional information, thereby provide narrower boundaries for the parameter. Deterministic tools like THEODOR are not able to resolve these fast events at all. After such a fast event, several profiles are affected and biased towards lower or higher heat flux densities. However, the integrated energy density is reasonably estimated, if all affected profiles are taken into account. It is considered problematic to use THEODOR in scenarios, where the heat load cannot be reasonably approximated as constant in between samples.

For a higher number of dimensions, the optimisation with gradient information is limited by the computation time and memory consumption of the gradients, as large matrices have to be processed. The SVD as bottleneck can be replaced by stochastic trace estimation, but larger images beyond 100 × 100 cells are presently still out of scope. Here other schemes would be needed to treat the linearised problem. The alternative is to use numerical integration, which does not require computations on large matrices. In contrast to the classic Metropolis Hastings Monte Carlo approach – based on a random walk – Hamiltonian MC reduces the auto-correlation time – in terms of samples and computation time – significantly. Hamiltonian MC is an efficient exploration tool, especially in combination with automatic differentiation.

11.3. Outlook

Methods for heat flux determination using Bayesian inference are available as a result of this thesis and first applications have been presented. However, not every question regarding divertor heat loads was answered. This section is supposed to illustrate some ideas of further options.

The combination of BayTh with a plasma models for the SOL is expected to improve the understanding of external magnetic perturbation. The plasma response can be determined from the heat load profile, for which a quantitative evaluation is now available.

Next steps include the use of an IR-system, which measures a spectrum instead of a single wave-length range. This allows to disentangle contributions like Bremsstrahlung from Planck-radiation. This method does not only increase the reliability of the system, but it also yields more information for other inference tasks.

In the case of fast events, additional data obtained from other diagnostics will allow to determine more parameters, like pulse duration and intensity, with greater precision. The model however should be refined and tailored for the specific case, as at the moment a constant background from the sample interval before to after the burst is assumed.

For a faster camera system the single heat flux profiles will be correlated stronger with each other, unless there are fast events happening. The effective noise in heat load scales with the square root of the sample frequency. Model selection would allow to select between a model for the heat load in time, which is smooth and allows for a high heat load resolution, and a model resolving fast events, at the expense of certainty in the heat load amplitude. As with increasing frame rate the interaction of neighbouring pixels decreases, simpler models for a single profile can be used. A possible formulation would be a form free description of a single profile, with the heat load in time being coupled. Here AK or adaptive diffusion can be used for model selection.

The use of more prior information can be investigated. For example the upstream profile of the heat load can be used as standard model for the hidden image of the AK model.

Bibliography

[Age17] International Energy Agency. *World Energy Outlook 2017.* 2017, p. 763. DOI: `https://doi.org/https://doi.org/10.1787/weo-2017-en`.

[aut+05] U. Schumacher (auth.) et al. *Plasma Physics: Confinement, Transport and Collective Effects.* 1st ed. Lecture Notes in Physics 670. Springer-Verlag Berlin Heidelberg, 2005. ISBN: 9783540252740.

[Bar+18] Alessandro Barp et al. 'Geometry and Dynamics for Markov Chain Monte Carlo'. In: *Annual Review of Statistics and Its Application* 5.1 (2018), pp. 451–471. DOI: `10.1146/annurev-statistics-031017-100141`.

[Béd08] Mylène Bédard. 'Optimal acceptance rates for Metropolis algorithms: Moving beyond 0.234'. In: *Stochastic Processes and their Applications* 118.12 (2008), pp. 2198–2222. ISSN: 0304-4149. DOI: `https://doi.org/10.1016/j.spa.2007.12.005`.

[Bet+17] Michael Betancourt et al. 'The geometric foundations of Hamiltonian Monte Carlo'. In: *Bernoulli* 23.4A (Nov. 2017), pp. 2257–2298. DOI: `10.3150/16-BEJ810`. URL: `https://doi.org/10.3150/16-BEJ810`.

[Bit04] J. A. Bittencourt. *Fundamentals of Plasma Physics.* Third. Fundamentals of Plasma Physics, Third Edition by J.A. Bittencourt. Published by Springer-Verlag, New York, Inc.; 2004. ISBN 0-387-20975-1., 2004.

[Bol+02] H Bolt et al. 'Plasma facing and high heat flux materials – needs for ITER and beyond'. In: *Journal of Nuclear Materials* 307-311 (2002), pp. 43–52. ISSN: 0022-3115. DOI: `https://doi.org/10.1016/S0022-3115(02)01175-3`.

[BS11] Michael Betancourt and Leo C. Stein. 'The Geometry of Hamiltonian Monte Carlo'. In: *arXiv e-prints*, arXiv:1112.4118 (Dec. 2011).

[BS98] Hans Dieter Baehr and Karl Stephan. 'Heat conduction and mass diffusion'. In: *Heat and Mass Transfer.* Berlin, Heidelberg: Springer Berlin Heidelberg, 1998, pp. 105–250. ISBN: 978-3-662-03659-4. DOI: `10.1007/978-3-662-03659-4_2`.

[CFL28] R. Courant, K. Friedrichs and H. Lewy. 'Über die partiellen Differenzengleichungen der mathematischen Physik'. In: *Mathematische Annalen* 100.1 (Dec. 1928), pp. 32–74. ISSN: 1432-1807. DOI: `10.1007/BF01448839`.

[Che16] Francis F. Chen. 'Plasma Applications'. In: *Introduction to Plasma Physics and Controlled Fusion*. Cham: Springer International Publishing, 2016, pp. 355–411. ISBN: 978-3-319-22309-4. DOI: `10.1007/978-3-319-22309-4_10`.

[CN96] J. Crank and P. Nicolson. 'A practical method for numerical evaluation of solutions of partial differential equations of the heat-conduction type'. In: *Advances in Computational Mathematics* 6.1 (Dec. 1996), pp. 207–226. ISSN: 1572-9044. DOI: `10.1007/BF02127704`.

[CP04] Ariel Caticha and Roland Preuss. 'Maximum entropy and Bayesian data analysis: Entropic prior distributions'. In: *Phys. Rev. E* 70 (4 Aug. 2004), p. 046127. DOI: `10.1103/PhysRevE.70.046127`.

[Eic+11] T. Eich et al. 'Type-I ELM power deposition profile width and temporal shape in JET'. In: *Journal of Nuclear Materials* 415.1, Supplement (2011). Proceedings of the 19th International Conference on Plasma-Surface Interactions in Controlled Fusion, S856–S859. ISSN: 0022-3115. DOI: `https://doi.org/10.1016/j.jnucmat.2010.11.079`.

[Eic+13] T. Eich et al. 'Scaling of the tokamak near the scrape-off layer H-mode power width and implications for ITER'. In: *Nuclear Fusion* 53.9 (2013), p. 093031. DOI: `10.1088/0029-5515/53/9/093031`.

[Eic+17] T. Eich et al. 'ELM divertor peak energy fluence scaling to ITER with data from JET, MAST and ASDEX upgrade'. In: *Nuclear Materials and Energy* 12 (2017). Proceedings of the 22nd International Conference on Plasma Surface Interactions 2016, 22nd PSI, pp. 84–90. ISSN: 2352-1791. DOI: `https://doi.org/10.1016/j.nme.2017.04.014`.

[Fai+17] M. Faitsch et al. '2D heat flux in ASDEX Upgrade L-Mode with magnetic perturbation'. In: *Nuclear Materials and Energy* 12 (2017). Proceedings of the 22nd International Conference on Plasma Surface Interactions 2016, 22nd PSI, pp. 1020–1024. ISSN: 2352-1791. DOI: `https://doi.org/10.1016/j.nme.2017.03.008`.

[FD04] R. Fischer and A. Dinklage. 'Integrated data analysis of fusion diagnostics by means of the Bayesian probability theory'. In: *Review of Scientific Instruments* 75.10 (2004), pp. 4237–4239. DOI: `10.1063/1.1787607`.

[Fed+14] G. Federici et al. 'Overview of EU DEMO design and R&D activities'. In: *Fusion Engineering and Design* 89.7 (2014). Proceedings of the 11th International Symposium on Fusion Nuclear Technology-11 (ISFNT-11) Barcelona, Spain, 15-20 September, 2013, pp. 882–889. ISSN: 0920-3796. DOI: `https://doi.org/10.1016/j.fusengdes.2014.01.070`.

[Fis02] R. Fischer. ‘Depth profiles and resolution limits in accelerator-based solid state analysis’. In: *Analytical and Bioanalytical Chemistry* 374.4 (Oct. 2002), pp. 619–625. ISSN: 1618-2650. DOI: `10.1007/s00216-002-1505-6`.

[Fit+17] Jack K. Fitzsimons et al. ‘Entropic Trace Estimates for Log Determinants’. In: *CoRR* abs/1704.07223 (2017). URL: `http://arxiv.org/abs/1704.07223`.

[Fit+18] JK Fitzsimons et al. ‘Improved stochastic trace estimation using mutually unbiased bases’. In: AUAI Press, 2018, pp. 310–318. URL: `http://arxiv.org/abs/1608.00117`.

[Fou09] Jean Baptiste Joseph Fourier. *The Analytical Theory of Heat - Cambridge Library Collection. Mathematics.* Cambridge University Press, New York, 2009. ISBN: 978-1-108-00178-6.

[FS13] Farhan Feroz and John Skilling. ‘Exploring multi-modal distributions with nested sampling’. In: *AIP Conference Proceedings* 1553.1 (2013), pp. 106–113. DOI: `10.1063/1.4819989`.

[FVD96] R. Fischer, W. Von Der Linden and V. Dose. ‘On the Importance of α Marginalization in Maximum Entropy’. In: *Maximum Entropy and Bayesian Methods.* Ed. by Kenneth M. Hanson and Richard N. Silver. Dordrecht: Springer Netherlands, 1996, pp. 229–236. ISBN: 978-94-011-5430-7.

[FvD96] R. Fischer, W. von der Linden and V. Dose. ‘Adaptive kernels and occam’s razor in inversion problems’. English. In: *Proceedings of the Maximum Entropy Conference (MAXENT 96).* Ed. by M. Sears et al. 1996, pp. 21–30.

[Gar+13] M Garcia-Munoz et al. ‘Fast-ion losses induced by ELMs and externally applied magnetic perturbations in the ASDEX Upgrade tokamak’. In: *Plasma Physics and Controlled Fusion* 55.12 (Nov. 2013), p. 124014. DOI: `10.1088/0741-3335/55/12/124014`.

[Gla11] Graham M. L. Gladwell. ‘Matrix Inverse Eigenvalue Problems’. In: *Dynamical Inverse Problems: Theory and Application.* Ed. by Graham M. L. Gladwell and Antonino Morassi. Vienna: Springer Vienna, 2011, pp. 1–28. ISBN: 978-3-7091-0696-9. DOI: `10.1007/978-3-7091-0696-9_1`.

[Gri97] Hans R. Griem. *Principles of Plasma Spectroscopy.* Cambridge Monographs on Plasma Physics. Cambridge University Press, 1997. DOI: `10.1017/CBO9780511524578`.

[Han+16] Insu Han et al. ‘Approximating the Spectral Sums of Large-scale Matrices using Chebyshev Approximations’. In: abs/1606.00942 (2016). URL: `http://arxiv.org/abs/1606.00942`.

[Hap+16a] T Happel et al. 'The I-mode confinement regime at ASDEX Upgrade: global properties and characterization of strongly intermittent density fluctuations'. In: *Plasma Physics and Controlled Fusion* 59.1 (Oct. 2016), p. 014004. DOI: `10.1088/0741-3335/59/1/014004`.

[Hap+16b] T. Happel et al. 'Turbulence intermittency linked to the weakly coherent mode in ASDEX Upgrade I-mode plasmas'. In: *Nuclear Fusion* 56.6 (May 2016), p. 064004. DOI: `10.1088/0029-5515/56/6/064004`.

[Hed97] Allan Hedin. *Spent nuclear fuel - how dangerous is it? A report from the project 'Description of risk' (SKB-TR–97-13). Sweden.* Mar. 1997. URL: `https://inis.iaea.org/collection/NCLCollectionStore/_Public/29/015/29015601.pdf?r=1&r=11`.

[Her+11] A. Herrmann et al. 'Real-time protection of in-vessel components in ASDEX Upgrade'. In: *Fusion Engineering and Design* 86.6 (2011). Proceedings of the 26th Symposium of Fusion Technology (SOFT-26), pp. 530–534. ISSN: 0920-3796. DOI: `https://doi.org/10.1016/j.fusengdes.2011.02.037`.

[Her+95] A Herrmann et al. 'Energy flux to the ASDEX-Upgrade diverter plates determined by thermography and calorimetry'. In: *Plasma Physics and Controlled Fusion* 37.1 (1995), p. 17. DOI: `10.1088/0741-3335/37/1/002`.

[Hog14] R.J. Hogan. 'Fast Reverse-Mode Automatic Differentiation Using Expression Templates in C++'. In: *ACM Transactions on Mathematical Software* 40.4 (June 2014), 26:1–26:24. DOI: `10.1145/2560359`.

[Hut02] I. H. Hutchinson. *Principles of Plasma Diagnostics.* 2nd ed. Cambridge University Press, 2002. DOI: `10.1017/CBO9780511613630`.

[Hut90] M.F. Hutchinson. 'A stochastic estimator of the trace of the influence matrix for laplacian smoothing splines'. In: *Communications in Statistics - Simulation and Computation* 19.2 (1990), pp. 433–450. DOI: `10.1080/03610919008812866`.

[Igl+17] D. Iglesias et al. 'Digital twin applications for the JET divertor'. In: *Fusion Engineering and Design* 125 (2017), pp. 71–76. ISSN: 0920-3796. DOI: `https://doi.org/10.1016/j.fusengdes.2017.10.012`.

[Igl+18] D. Iglesias et al. 'An improved model for the accurate calculation of parallel heat fluxes at the JET bulk tungsten outer divertor'. In: *Nuclear Fusion* 58.10 (Aug. 2018), p. 106034. DOI: `10.1088/1741-4326/aad83e`.

[Ill18] Johannes Illerhaus. 'Estimation, Validation and Uncertainty of the Position of the Separatrix Contour at ASDEX Upgrade'. Technische Universität München, 2018. URL: `http://hdl.handle.net/21.11116/0000-0003-71DD-5n`.

[IPP18] IPP. IPP. www.aug.ipp.mpg.de. Accessed at 30.09.18. 2018.

[J G11] Charles J. Geyer. 'Introduction to Markov chain Monte Carlo'. In: *Handbook of Markov Chain Monte Carlo* (Jan. 2011).

[Jay03] E. T. Jaynes. *Probability Theory: The Logic of Science.* Ed. by G. LarryEditor Bretthorst. Cambridge University Press, 2003. DOI: `10.1017/CBO9780511790423`.

[Jen14] Prof. Dr. Frank Jenko. 'Plasmaphysik 2'. In: (2014).

[JHG14] N. Jaksic, A. Herrmann and H. Greuner. 'Transient Thermal and Structural Mechanics Investigation of the New Solid Tungsten Divertor Tile for Special Purposes at ASDEX Upgrade'. In: *IEEE Transactions on Plasma Science* 42.6 (June 2014), pp. 1790–1795. ISSN: 0093-3813. DOI: `10.1109/TPS.2014.2311158`.

[Jij09] L.M. Jiji. *Heat Convection.* Springer Berlin Heidelberg, 2009. ISBN: 9783642029714. URL: `https://books.google.de/books?id=ZdyOoyA8F%5C_QC`.

[JJ96] Lijenzin Jan-Olov and Rydberg Jan. *Risks from nuclear waste.* Nov. 1996. URL: `https://inis.iaea.org/collection/NCLCollectionStore/_Public/28/028/28028626.pdf?r=1&r=1`.

[Kan+16] C. S. Kang et al. 'Study on the heat flux reconstruction with the infrared thermography for the divertor target plates in the KSTAR tokamak'. In: *Review of Scientific Instruments* 87.8 (2016), p. 083508. DOI: `10.1063/1.4961030`.

[Kes09] H. K. Kesavan. 'Jaynes' maximum entropy principleJaynes' Maximum Entropy Principle'. In: *Encyclopedia of Optimization.* Ed. by Christodoulos A. Floudas and Panos M. Pardalos. Boston, MA: Springer US, 2009, pp. 1779–1782. ISBN: 978-0-387-74759-0. DOI: `10.1007/978-0-387-74759-0_312`.

[Kik11] Mitsuru Kikuchi. 'Sun on Earth: Endless Energy from Hydrogen'. In: *Frontiers in Fusion Research: Physics and Fusion.* London: Springer London, 2011, pp. 1–14. ISBN: 978-1-84996-411-1. DOI: `10.1007/978-1-84996-411-1_1`. URL: `https://doi.org/10.1007/978-1-84996-411-1_1`.

[KW96] Robert E. Kass and Larry Wasserman. 'The Selection of Prior Distributions by Formal Rules'. In: *Journal of the American Statistical Association* 91.435 (1996), pp. 1343–1370. DOI: `10.1080/01621459.1996.10477003`.

[LDT14] Wolfgang von der Linden, Volker Dose and Udo von Toussaint. *Bayesian Probability Theory: Applications in the Physical Sciences.* Cambridge University Press, 2014. DOI: `10.1017/CBO9781139565608`.

[LY15] M. Li and J.-H. You. 'Interpretation of the deep cracking phenomenon of tungsten monoblock targets observed in high-heat-flux fatigue tests at 20 MW/m2'. In: *Fusion Engineering and Design* 101 (2015), pp. 1–8. DOI: `10.1016/j.fusengdes.2015.09.008`.

[Mak+12] M. A. Makowski et al. ‘Analysis of a multi-machine database on divertor heat fluxesa)’. In: *Physics of Plasmas* 19.5, 056122 (2012). DOI: `http://dx.doi.org/10.1063/1.4710517`.

[Man+17] P. Manz et al. ‘Turbulence characteristics of the I-mode confinement regime in ASDEX Upgrade’. In: *Nuclear Fusion* 57.8 (July 2017), p. 086022. DOI: `10.1088/1741-4326/aa7476`.

[Mao+18] X. Mao et al. ‘Exploring the engineering limit of heat flux of a W/RAFM divertor target for fusion reactors’. In: *Nuclear Fusion* 58.6 (Apr. 2018), p. 066014. DOI: `10.1088/1741-4326/aabb64`.

[MD15] Hadi Mohasel Afshar and Justin Domke. ‘Reflection, Refraction, and Hamiltonian Monte Carlo’. In: *Advances in Neural Information Processing Systems 28.* Ed. by C. Cortes et al. Curran Associates, Inc., 2015, pp. 3007–3015. URL: `http://papers.nips.cc/paper/5801-reflection-refraction-and-hamiltonian-monte-carlo.pdf`.

[Mey+19] H. Meyer et al. ‘Overview of physics studies on ASDEX Upgrade’. In: *Nuclear Fusion* 59.11 (July 2019), p. 112014. DOI: `10.1088/1741-4326/ab18b8`.

[NAG16] NAG. The NAG C Library, The Numerical Algorithms Group (NAG), Oxford, United Kingdom `www.nag.com`. Accessed 15 Juli 2016. 2016.

[Nea93] Radford M Neal. *Probabilistic inference using Markov chain Monte Carlo methods.* 1993.

[Nem17] R. D. Nem. ‘’Surface Temperature Measurement of In-Vessel Components on ASDEX Upgrade using Infrared Spectroscopy’’. 2017. DOI: `doi:10.17617/2.2506539`.

[Neu74] A. Neumaier. ‘Rundungsfehleranalyse einiger Verfahren zur Summation endlicher Summen’. In: *ZAMM - Journal of Applied Mathematics and Mechanics / Zeitschrift für Angewandte Mathematik und Mechanik* 54.1 (1974), pp. 39–51. DOI: `10.1002/zamm.19740540106`. eprint: `https://onlinelibrary.wiley.com/doi/pdf/10.1002/zamm.19740540106`.

[Nil+19] D Nille et al. ‘Analytic 1D approximation of the divertor broadening S in the divertor region for conductive heat transport’. In: *Plasma Physics and Controlled Fusion* 61.8 (June 2019), p. 085016. DOI: `10.1088/1361-6587/ab240f`.

[Nil16] D Nille. ‘Comparison between 1D and 2D approximations describing heat diffusion in the ASDEX Upgrade Divertor’. Feb. 2016. DOI: `10.17617/2.3030046`.

[NT19] D. Nille and U. von Toussaint. ‘2D Deconvolution using Adaptive Kernel’. In: *Bayesian Inference and Maximum Entropy Methods in Science and Engineering.* Multidisciplinary Digital Publishing Institute, 2019. DOI: `10.3390/proceedings2019033006`.

[PA15] *Paris Agreement.* URL: `https://treaties.un.org/pages/ViewDetails.aspx?src=TREATY&mtdsg_no=XXVII-7-d&chapter=27&clang=_en` (visited on 28/03/2019).

[Pla00a] M. Planck. 'Über eine Verbesserung der Wienschen Spektralgleichung.' In: *Deutsche Physikalische Gesellschaft* (1900).

[Pla00b] M. Planck. 'Zur Theorie des Gesetzes der Energieverteilung im Normalspektrum.' In: *Deutsche Physikalische Gesellschaft* (1900).

[Pre+07] William H. Press et al. *Numerical Recipes.* third edition. Cambridge University Press, 2007.

[Pue96] R. C. Puetter. 'Pixon-Based Multiresolution Image Reconstruction and Quantification of Image Information Content'. In: *Maximum Entropy and Bayesian Methods.* Ed. by Kenneth M. Hanson and Richard N. Silver. Dordrecht: Springer Netherlands, 1996, pp. 145–152. ISBN: 978-94-011-5430-7.

[Rac+14a] M. Rack et al. 'Findings of pre-ELM structures through the observation of divertor heat load patterns at JET with applied $n = 2$ perturbation fields'. In: *Nuclear Fusion* 54.7 (Apr. 2014), p. 072004. DOI: `10.1088/0029-5515/54/7/072004`.

[Rac+14b] M. Rack et al. 'Modified heat load deposition of the ELM crash due to $n = 2$ perturbation fields at JET'. In: *Nuclear Fusion* 54.6 (Mar. 2014), p. 064012. DOI: `10.1088/0029-5515/54/6/064012`.

[Rap+10] J. Rapp et al. 'Geometry and expected performance of the solid tungsten outer divertor row in JET'. In: *Fusion Engineering and Design* 85.2 (2010), pp. 153–160. ISSN: 0920-3796. DOI: `https://doi.org/10.1016/j.fusengdes.2009.08.009`.

[RC09] D. Reiser and D. Chandra. 'Plasma currents induced by resonant magnetic field perturbations in tokamaks'. In: *Physics of Plasmas* 16.4 (2009), p. 042317. DOI: `10.1063/1.3126548`.

[Rei14] Felix Reimold. 'Experimental Studies and Modeling of Divertor Plasma Detachment in H-Mode Discharges in the ASDEX Upgrade Tokamak'. PhD Thesis. Technische Universität München, 2014.

[Rie01] Max J. Riedl. *Optical Design Fundamentals for Infrared Systems, Second Edition.* 2nd ed. SPIE Tutorial Texts in Optical Engineering Vol. TT48. SPIE Publications, 2001. ISBN: 9780819440518. DOI: `10.1117/3.412729`.

[Roh+09] V Rohde et al. 'Dust investigations at ASDEX Upgrade'. In: *Physica Scripta* T138 (Dec. 2009), p. 014024. DOI: `10.1088/0031-8949/2009/t138/014024`.

[RS61] H. Raiffa and R. Schlaifer. *Applied statistical decision theory*. Studies in managerial economics. Division of Research, Graduate School of Business Administration, Harvard University, 1961. ISBN: 9780875840178. URL: `https://books.google.de/books?id=wPBLAAAAMAAJ`.

[Rub19] Marek Rubel. 'Fusion Neutrons: Tritium Breeding and Impact on Wall Materials and Components of Diagnostic Systems'. In: *Journal of Fusion Energy* 38.3 (Aug. 2019), pp. 315–329. ISSN: 1572-9591. DOI: `10.1007/s10894-018-0182-1`.

[RW06] CE. Rasmussen and CKI. Williams. *Gaussian Processes for Machine Learning*. Adaptive Computation and Machine Learning. Cambridge, MA, USA: MIT Press, Jan. 2006, p. 248.

[Sär13] Simo Särkkä. *Bayesian Filtering and Smoothing*. Institute of Mathematical Statistics Textbooks. Cambridge University Press, 2013. DOI: `10.1017/CBO9781139344203`.

[Sie+13] B Sieglin et al. 'Power load studies in JET and ASDEX-Upgrade with full-W divertors'. In: *Plasma Physics and Controlled Fusion* 55.12 (2013), p. 124039. URL: `http://stacks.iop.org/0741-3335/55/i=12/a=124039`.

[Sie+15] B. Sieglin et al. 'Real time capable infrared thermography for ASDEX Upgrade'. In: *Review of Scientific Instruments* 86.11, 113502 (2015). DOI: `http://dx.doi.org/10.1063/1.4935580`.

[Sie+16] B Sieglin et al. 'Investigation of scrape-off layer and divertor heat transport in ASDEX Upgrade L-mode'. In: *Plasma Physics and Controlled Fusion* 58.5 (2016), p. 055015. URL: `http://stacks.iop.org/0741-3335/58/i=5/a=055015`.

[Sie+17] B. Sieglin et al. 'Assessment of divertor heat load with and without external magnetic perturbation'. In: *Nuclear Fusion* 57.6 (May 2017), p. 066045. DOI: `10.1088/1741-4326/aa6c20`.

[Sie14] Bernhard Sieglin. 'Divertor Power Load Studies in All Metal ASDEX Upgrade and JET'. PhD Thesis. Technische Universität München, 2014.

[SK19] John Skilling and Kevin H. Knuth. 'The Symmetrical Foundation of Measure, Probability, and Quantum Theories'. In: *Annalen der Physik* 531.3 (2019), p. 1800057. DOI: `10.1002/andp.201800057`.

[Ski06] John Skilling. 'Nested sampling for general Bayesian computation'. In: *Bayesian Anal.* 1.4 (Dec. 2006), pp. 833–859. DOI: `10.1214/06-BA127`.

[Ski09] John Skilling. 'Conjugate Gradient for Bayesian Computation'. In: *AIP Conference Proceedings* 1193.1 (2009), pp. 269–276. DOI: `10.1063/1.3275624`. eprint: `https://aip.scitation.org/doi/pdf/10.1063/1.3275624`.

[Ski89a] John Skilling. ‘Classic Maximum Entropy’. In: *Maximum Entropy and Bayesian Methods: Cambridge, England, 1988*. Dordrecht: Springer Netherlands, 1989, pp. 45–52. DOI: `10.1007/978-94-015-7860-8_3`.

[Ski89b] John Skilling. ‘The Eigenvalues of Mega-dimensional Matrices’. In: *Maximum Entropy and Bayesian Methods: Cambridge, England, 1988*. Dordrecht: Springer Netherlands, 1989, pp. 455–466. ISBN: 978-94-015-7860-8. DOI: `10.1007/978-94-015-7860-8_48`.

[Ski98] J. Skilling. ‘Probabilistic data analysis: an introductory guide’. In: *Journal of Microscopy* 190.1-2 (1998), pp. 28–36. DOI: `10.1046/j.1365-2818.1998.2780835.x`.

[SM06] Arnaud Soirat and Lou Massa. ‘The number of independent parameters defining a projector: proof in matrix representation and resolution of previously conflicting arguments’. In: vol. 21. Nov. 2006, pp. 127–145. DOI: `10.1007/0-306-46943-X_8`.

[Sta00] P.C. Stangeby. *The Plasma Boundary of Magnetic Fusion Devices*. Series in Plasma Physics and Fluid Dynamics. Taylor & Francis, 2000. ISBN: 9780750305594.

[SWW93] C. E. M. Strauss, D. H. Wolpert and D. R. Wolf. ‘Alpha, Evidence, and the Entropic Prior’. In: *Maximum Entropy and Bayesian Methods: Paris, France, 1992*. Ed. by Ali Mohammad-Djafari and Guy Demoment. Dordrecht: Springer Netherlands, 1993, pp. 113–120. ISBN: 978-94-017-2217-9. DOI: `10.1007/978-94-017-2217-9_15`.

[Tho18] Nick Thomopoulos. *Probability Distributions*. Apr. 2018. ISBN: 978-3-319-76041-4. DOI: `10.1007/978-3-319-76042-1`.

[Tol17] P. Tolias. ‘Analytical expressions for thermophysical properties of solid and liquid tungsten relevant for fusion applications’. In: *Nuclear Materials and Energy* 13 (2017), pp. 42–57. ISSN: 2352-1791. DOI: `https://doi.org/10.1016/j.nme.2017.08.002`.

[Vie+12] E. Viezzer et al. ‘High-resolution charge exchange measurements at ASDEX Upgrade’. In: *Review of Scientific Instruments* 83.10 (2012), p. 103501. DOI: `10.1063/1.4755810`.

[Wil+16] M Willensdorfer et al. ‘Plasma response measurements of external magnetic perturbations using electron cyclotron emission and comparisons to 3D ideal MHD equilibrium’. In: *Plasma Physics and Controlled Fusion* 58.11 (Sept. 2016), p. 114004. DOI: `10.1088/0741-3335/58/11/114004`.

[Wil+18] M Willensdorfer et al. ‘Dynamics of ideal modes and subsequent ELM crashes in 3D tokamak geometry from external magnetic perturbations’. In: *Plasma Physics and Controlled Fusion* 61.1 (Nov. 2018), p. 014019. DOI: `10.1088/1361-6587/aadc39`. URL: `https://doi.org/10.1088%2F1361-6587%2Faadc39`.

[WK00] Scott W White and Gerald L Kulcinski. ‘Birth to death analysis of the energy payback ratio and CO2 gas emission rates from coal, fission, wind, and DT-fusion electrical power plants’. In: *Fusion Engineering and Design* 48.3 (2000), pp. 473–481. ISSN: 0920-3796. DOI: `https://doi.org/10.1016/S0920-3796(00)00158-7`.

[You+16] J.H. You et al. ‘European DEMO divertor target: Operational requirements and material-design interface’. In: *Nuclear materials and energy* 9 (2016). 31.40.03; LK 01, pp. 171–176. ISSN: 2352-1791. DOI: `10.1016/j.nme.2016.02.005`.

[Zoh17] Bahman Zohuri. ‘Inertial Confinement Fusion (ICF)’. In: *Inertial Confinement Fusion Driven Thermonuclear Energy*. Cham: Springer International Publishing, 2017, pp. 193–238. ISBN: 978-3-319-50907-5. DOI: `10.1007/978-3-319-50907-5_4`.

Acknowledgement

I want to thank my academic supervisor Udo von Toussaint for his support, advice and suggestions during my work, especially regarding probabilistic methods and inference. Furthermore I want to thank my colleagues Bernhard Sieglin and Michael Faitsch for their support and cooperation regarding the measurements and the valuable insights into plasma physics. Together with Thomas Eich they have built the infrared diagnostics used for this work and keep improving them, to expand our understanding of divertor and SOL physics.

The data could only be obtained thanks to the ASDEX Upgrade team, making the machine perform at its best, from technicians over engineers to physicists. A special thanks goes to everyone supporting me in pursuit of new data, from the theoretical side to the scientific operators. The ASDEX Upgrade Team is a fabulous community when it comes to push the knowledge about material, plasma and fusion related physics forward.

This work was conducted at the department of numerical methods in plasma physics (NMPP) within IPP, which is host to many brilliant people. Here was formed a knowledge base from elementary mathematical concepts to numerical engineering as well as inference. Valuable discussions improved my understanding from theory to numerical implementation.

Appendices

A. Probabilities

A.1. Example for working with probabilities

For readers not familiar with the handling of probability distributions, here a few expressions for the two data and two parameter expression are shown:

$$\begin{aligned}
p(q^1, q^2, d^1, d^2) &= p(q^1, q^2 \mid d^1, d^2) \cdot p(d^1) \cdot p(d^2) \\
&= p(d^1, d^2, q^2 \mid q^1) \cdot p(q^1) \\
&= p(d^1, d^2 \mid q^1, q^2) \cdot p(q^2 \mid q^1) \cdot p(q^1) \\
&= p(d^1 \mid d^2, q^1, q^2) \cdot p(d^2 \mid q^1, q^2) \cdot p(q^2 \mid q^1) \cdot p(q^1) \\
&= p(d^2 \mid d^1, q^1, q^2) \cdot p(d^1 \mid q^1, q^2) \cdot p(q^2 \mid q^1) \cdot p(q^1) \\
&= p(d^2 \mid d^1, q^1, q^2) \cdot p(d^1 \mid q^1) \cdot p(q^2 \mid q^1) \cdot p(q^1)
\end{aligned}$$

In the first line the independence of the data is used: $p(d^1 \mid d^2) = p(d^1)$. A joint distribution for independent quantities can be split, because the conditional expression can be dropped:

$$\begin{aligned}
p(d^1, d^2 \mid q^1, q^2) =& p(d^1 \mid d^2, q^1, q^2) \cdot p(d^2 \mid q^1, q^2) \\
=& p(d^2 \mid d^1, q^1, q^2) \cdot p(d^1 \mid q^1, q^2) \\
=& p(d^1 \mid q^1, q^2) \cdot p(d^2 \mid q^1, q^2)
\end{aligned}$$

The interpretation of the other expressions is left to the curious reader, who wants to understand the manipulation of probabilities.

B. Python Code Reference

B.1. Reference heat load

The profile has 150 pixel on a equidistant rid with 0.5 mm resolution, in the range (−15 − 60) mm. The sample rate is 1 kHz with a total of 5000 profiles. The time dependent parameters for profile k – corresponding to time $t = k \cdot 1\,\mathrm{ms}$ – are

$$q_0 = (1 + 0.5 \cdot \sin(k \cdot 1e - 1))\,\mathrm{MW\,m^{-2}} \tag{B.1}$$

$$\lambda_q = 5 \cdot (3 + 1.0 \cdot \cos(k \cdot 0.04))\,\mathrm{mm} \tag{B.2}$$

$$S = 5 \cdot (1 + 0.5 \cdot \sin(k/5000))\,\mathrm{mm} \tag{B.3}$$

$$s_0 = 5 \cdot (1 + 2.0 \cdot \cos(k \cdot 0.015))\,\mathrm{mm} \tag{B.4}$$

$$q_{BG} = (0.3 + 0.1 \cdot \cos(k \cdot 0.035))\,\mathrm{MW\,m^{-2}} \tag{B.5}$$

The first 100 profiles are scaled linearly with their index $k/100$ – starting with $k = 0$.
ELM like structures are position at indices 300, 800, 1000, 1500, 2000 and 2500. They are additive for the above L-Mode like heat load. The parameters are a length of 5 ms with 1 ms of onset. The Amplitude is $2\,\mathrm{MW\,m^{-2}}$.
Note: a factor of 5 is included for the spatial components λ_q and S, because the values are specified as upstream values. The flux expansion is typically a factor of 5 in AUG. For example $\lambda_q^{\mathrm{omp}} = 2\,\mathrm{mm}$ at the outer mid-plane corresponds to a mapped value $\lambda_q^{\mathrm{div}} = 5 \cdot 2\,\mathrm{mm} = 10\,\mathrm{mm}$.
The following code is used to create the reference heat load profile described in section 8.1, including the Eich-model from equation 2.21:

Listing B.1: Code to create the reference heat load.

```python
import numpy as np
from numpy import sin, cos, exp
from scipy.special import erfc

def Eichmodel(p, pos):
    # parameters p: [q0, lambda_q, S, s0, q_BG]
    z = p[2]/2.0/p[1]
    return (0.5*p[0]*exp(z*z-(pos-p[3])/p[1]) *
             erfc(z-(pos-p[3])/p[2]) + p[4])

def get_testsetup():
    Nt = 5000 # number of profiles in time
    Nl = 150  # pixel in the profile
    time = np.linspace(0, 1e-3*Nt, Nt) # 1ms resolution
    location = np.linspace(-15,60, Nl)*1e-3
    data = np.zeros([Nl,Nt])
    for i in xrange(0, Nt):
        data[:,i] = Eichmodel([1e6+5e5*sin(i*1e-1), # q0
                               5*(3e-3+1e-3*cos(i*4e-2)), # lamq
                               5*(1e-3+0.5e-3*sin(i*1./Nt)), # S
                               5*(1e-3+2e-3*cos(i*15e-3)**2), # s0
                               2.0e5 + 1e5*(1.0 + cos(i*35e-3))], #q_BG
                              location)
    # linear ramp-up at the beginning
    data[:, i] *= np.linspace(0, 1, 100+1)
    # indices for ELM's
    tELM = [300, 800, 1000, 1500, 2000, 2500]
    def gauss(location, amp, sigma):
        x = np.arange(location.size) - location.size * .5
        return amp * np.exp( -.5* (x/sigma)**2 )
    ELMlength = 5
    ELMonset = 1
    ELMamp = 2e6
    for i in tELM:
        for j in xrange(ELMonset):
            data[:,i+j] += gauss(location,
                                 ELMamp*np.exp((1.*j-1.*ELMonset)/.5),
                                 location.size*.2)
        for j in xrange(ELMlength):
            data[:,i+ELMonset+j] += gauss(location,
                                          ELMamp*np.exp(-1.*j/1.),
                                          location.size*.2)
    # return dictionary
    return {'data': data, 'time': time, 'location': location}
```