# Light curves and multidimensional reconstructions of photon observations

**Daniel Pumpe**

SIGILL·UNIVERSITAT·LUDOVIC·MAXIM·

München 2018

# Light curves and multidimensional reconstructions of photon observations

**Daniel Pumpe**

Dissertation
an der Fakultät für Physik
der Ludwig–Maximilians–Universität
München

vorgelegt von
Daniel Pumpe
aus Starnberg

München, den 23. Februar 2018

Erstgutachter: PD Dr. Torsten Enßlin
Zweitgutachter: Prof. Dr. Jochen Weller
Tag der mündlichen Prüfung: 11. Mai 2018

*"Don't be pushed by your problems. Be led by your dreams."*

*by Ralph Waldo Emerson*

## Zusammenfassung

Diese Dissertation konzentriert sich auf die Entwicklung und Anwendung von bayesianischen Inferenzmethoden, um physikalisch relevante Informationen aus verrauschten Photonenbeobachtungen zu extrahieren. Des Weiteren wird eine Methode entwickelt, um Beobachtungen von komplexen Systemen, welche sich stochastisch mit der Zeit entwickeln, anhand weniger Trainingsbeispiele in verschiedene Klassen einzuordnen.

Zu letztem Zweck entwickeln wir den Dynamic System Classifier (DSC). Dieser basiert auf der grundlegenden Annahme, dass viele komplexe Systeme in einem vereinfachten Rahmen durch stochastische Differentialgleichungen (SDE) mit zeitabhängigen Koeffizienten beschrieben werden können. Diese werden verwendet, um Informationen aus einer Klasse ähnlicher, aber nicht identischer simulierter Systeme zu abstrahieren. Der DSC ist in zwei Phasen unterteilt. In der ersten Lernphase werden die Koeffizienten der SDE aus einem kleinen Trainingsdatensatz gelernt. Sobald diese gelernt wurden, dienen sie für einen kostengünstigen Vergleich von Daten und abstrahierter Information. Wir entwickeln, implementieren und testen beide Schritte in dem Rahmen bayesianischer Logik für kontinuierliche Größen, nämlich der Informationsfeldtheorie.

Der zweite Teil dieser Arbeit beschäftigt sich mit astronomischer Bildgebung basierend auf Zählraten von Photonen. Die Notwendigkeit hierfür ergibt sich unter anderem aus der Verfügbarkeit von zahlreichen Satelliten, welche die Röntgen- und $\gamma-$Strahlen im Weltraum beobachten. In diesem Zusammenhang entwickeln wir den existierenden D$^3$PO-Algorithmus weiter, hin zu D$^4$PO, um multidimensionale Photonenbeobachtungen zu entrauschen, zu dekonvolvieren und in morphologisch unterschiedliche Komponenten aufzuteilen. Die Zerlegung wird durch ein hierarchisches bayesianisches Parametermodell gesteuert. Dieses erlaubt es, Felder zu rekonstruieren, die über den Produktraum von mehreren Mannigfaltigkeiten definiert sind. D$^4$PO zerlegt den beobachteten Fluss an Photonen in eine diffuse, eine punktförmige und eine Hintergrundkomponente, während er gleichzeitig die Korrelationsstruktur für jede einzelne Komponente in jeder ihrer Mannigfaltigkeiten lernt. Die Funktionsweise von D$^4$PO wird anhand eines simulierten Datensatzes hochenergetischer Photonen demonstriert.

Schließlich wenden wir D$^4$PO auf Daten der Magnetar-Flares von SGR 1806-20 und SGR 1900+14 an, um nach deren charakteristischen Eigenschwingungen zu suchen. Der Algorithmus rekonstruierte erfolgreich den logarithmischen Photonenfluss sowie dessen spektrale Leistungsdichte. Im Gegensatz zu früheren Arbeiten anderer Autoren können wir quasi- periodische Oszillationen (QPO) in den abklingenden Enden dieser Ereignisse bei Frequenzen $\nu > 17$ Hz nicht bestätigen. Deren Echtheit ist fraglich, da diese in das von Rauschen dominierende Regime fallen. Dennoch finden wir neue Kandidaten für Oszillationen bei $\nu \approx 9.2$ Hz (SGR 1806-20) und $\nu \approx 7.7$ Hz (SGR 1900+14). Für den Fall, dass diese Oszillationen real sind, bevorzugen moderne theoretische Modelle von Magnetaren relativ schwache Magnetfelder im Bereich von $\bar{B} \approx 6 \times 10^{13} - 3 \times 10^{14}$ G.

## Abstract

This thesis focuses on the development and application of Bayesian inference methods to extract physical relevant information from noise contaminated photon observations and to classify the observations of complex stochastically evolving systems into different classes based on a few training samples of each class.

To this latter end we develop the dynamic system classifier (DSC). This is based on the fundamental assumption that many complex systems may be described in a simplified framework by stochastic differential equations (SDE) with time dependent coefficients. These are used to abstract information from a class of similar but not identical simulated systems. The DSC is split into two phases. In the first learning phase the coefficients of the SDE are learned from a small training data set. Once these are obtained, they serve for an inexpensive data - class comparison. We develop, implement, and test both steps in a Bayesian inference framework for continuous quantities, namely information field theory.

Astronomical imaging based on photon count data is a challenging task but absolutely necessary due to todays availability of space based X-ray and $\gamma$-ray telescopes. In this context we advance the existing D$^3$PO algorithm into D$^4$PO to denoise, denconvolve, and decompose multidimensional photon observations into morphologically different components. The decomposition is driven by a probabilistic hierarchical Bayesian parameter model, allowing us to reconstruct fields, that are defined over the product space of multiple manifolds. Thereby D$^4$PO decomposes the photon count data into a diffuse, point-like, and background component, while it simultaneously learns the correlation structure over each of their manifolds individually. The capabilities of the algorithm are demonstrated by applying it to a simulated high energy photon count data set.

Finally we apply D$^4$PO to analyse the giant magnetar flare data of SGR 1806-20 and SGR 1900+14. The algorithm successfully reconstructes the logarithmic photon flux as well as its power spectrum. In contrast to previous findings we cannot confirm quasi periodic oscillations (QPO) in the decaying tails of these events at frequencies $\nu > 17$ Hz. They might not be real as these fall into the noise dominated regime of the spectrum. Nevertheless we find new candidates for oscillations at $\nu \approx 9.2$ Hz (SGR 1806-20) and $\nu \approx 7.7$ Hz (SGR 1900+14). In case these oscillations are real, state of the art theoretical models of magnetars favour relatively weak magnetic fields in the range of $\bar{B} \approx 6 \times 10^{13} - 3 \times 10^{14}$ G.

# Contents

# List of Figures

# List of Tables

# 1. Introduction

## 1.1. Life of a star in a tiny nutshell

After the birth of a protostar, the central regions of a star are being pulled together by gravity, causing an increasingly smaller radius and thus higher densities. The contraction is stopped when the star begins to fuse hydrogen nuclei into helium nuclei, generating heat and pressure to halt the collapse. A long time later, after the hydrogen in the core of the star has all been consumed, gravity takes over again and compresses the core, heating it up until helium burning sets in to form carbon. This produces enough heat to balance the gravitational force and the stars outer atmosphere expands.
The processes following, after the helium burning phase, depend on the mass of the star. Stars comparable to the mass of our sun begin their collapse right after the helium burning and form a white dwarf. More massive stars might explode. The remnant of such a supernova explosion can become a neutron star or be further compressed to form a black hole. As a rule of thumb, the more massive the star, the smaller its remnants eventually become due to the never ending force of gravity.

Due to the complex structure and large dynamic range of the physical processes to describe the life of a star we rely among others on numerical simulations to gain a deeper understanding. As these simulations are computationally very challenging and expensive it remains elusive to explore all possible initial conditions which might have lead to observations. Consequently we have to develop methods, which can abstract information from a few of these simulations and then confront the abstracted information to real observations. Thereby one may identify the initial conditions which explains a specific observation. This type of inference problem becomes especially apparent when analysing gravitational waves caused by a core collapse at the end of a stars life. Hence I will briefly discuss the nature and detection of gravitational waves in the next section.

Another channel to probe the nature of stars is their emission of photons in different wavelengths. To handle the data of various space based telescopes powerful data analysis techniques are needed. To this end we develop D$^4$PO, a generic algorithm to denoise, deconvolve, and decompose multidimensional photon observations. This algorithm is applied to the analysis of the light curves of giant magnet flares. We will briefly review the physical fundaments of magnetar (flares) in section 1.1.2.

### 1.1.1. Gravitational waves of a core collapse supernova

The field of gravitational wave astronomy has become reality and today we can observe gravitational waves (GW) from different sources on an almost regular basis [1–6]. One yet undetected but important source of these waves are stellar gravitational collapses. This source class embraces the entire spectrum of stellar masses, from the collapse of a white dwarf induced by the accretion of low mass stars ($M_{\text{star}} < 10 M_{\text{sun}}$) to the collapses of more massive stars under their own weight ($M_{\text{star}} > 10 M_{\text{sun}}$). All of these phenomena have the potential of being detected by GW observations as they involve a rapid change of dense matter distributions that should lead to the emission of GW.

#### Gravitational wave emission

Most of the GW emission studies are based on a multipole expansion of the perturbation $h_{\mu\nu}$ to a background spacetime metric tensor $g^b_{\mu\nu}$. This yields [7]

$$h^{\text{TT}}_{jk} = \left[ \frac{2}{d} \frac{G}{c^4} \frac{d^2}{dt^2} \mathcal{I}_{jk} \left( t - r \right) + \frac{8}{3d} \frac{G}{c^5} \epsilon_{pq(j} \frac{d^2}{dt^2} \mathcal{S}_{k)p} \left( t - r \right) n_q \right]^{\text{TT}} . \tag{1.1}$$

$\mathcal{I}_{jk}$ and $\mathcal{S}_{jk}$ are the 3D-mass and quadruple moment of the source, $d$ is the distance between source and observer, $\epsilon_{ijk}$ is the antisymmetric tensor and $n_q$ is the unit vector pointing in propagation direction. Parentheses in the subscripts indicate symmetrisation over the enclosed indices. Further one has to sum over the indices $p$ and $q$ in the second term. The superscript TT refers that one has to take the traceless traverse gauge; $G$ is the gravitational constant and $c$ is the speed of light.

In case the bulk mass motion dominates the dynamics, the first term of eq. (1.1) describes the radiation. This term causes the well known 'chirp' signal of the binary in-spiral. The second term in eq. (1.1) describes radiation caused by mass currents in the collapsing core. In case the background spacetime is flat (or can at least be approximated as flat) the mass and current moments have a particular simple form. In Cartesian coordinates the mass quadrupole becomes

$$\mathcal{I}_{jk} = \int d^3x \rho \left( x_j x_k - \frac{1}{3} r^2 \delta_{jk} \right) , \tag{1.2}$$

with $\rho$ as the mass density and $\delta_{jk} = 1$ for $j = k$ and 0 otherwise. The second term ensures that the integrand is trace free.

However $\mathcal{I}_{jk}$ and $\mathcal{S}_{jk}$ are in general analytically not accessible due to the complex dynamics of a core collapse. Therefore these have been subject to multiple numerical studies in the past decades. A deeper understanding of GW will provide us with unique information, complementary to the information of electromagnetic and neutrino detections.

As shown in eq. (1.1) GW arise from the coherent movement of mass, whereas electromagnetic waves are mostly produced by incoherent superposition of radiation of electrons, atoms, and molecules. Thus GW encode a different kind of information. Furthermore electromagnetic waves interact strongly with their environment and can therefore provide us only with information from the surface of last scattering. GW in contrast propagate from their point of creation to the observer almost unperturbed due to their small coupling constant.
Neutrinos, which are also produced in a core collapse, are extremely sensitive to microphysics, while GW are only sensitive to large scale motions of mass. Hence combining these three channels, electromagnetic radiation, neutrinos and GW, can tell us much about the conditions of a collapsing core.

### Simulating gravitational wave emission

As GW encode unique and yet unexplored information about a stars interior, constant interest of scientists has been drawn to this field astronomy in the past decades [8–19]. During this period scientists have produced estimates of GW strength through numerical simulations. The amplitude of the results vary in order of magnitudes due to the complexity of the collapse, seen through the GW channel as encoded by eq. (1.1). Some of the current and past problems are [20]:

- accurate progenitor models, including realistic angular momentum distributions,
- proper treatment of microphysics, including the use of realistic equations of state and neutrino transport,
- non-axisymmetric effects in three-dimensions,
- general-relativistic effects,
- magnetic-field effects, and
- effects of the envelope on core behavior.

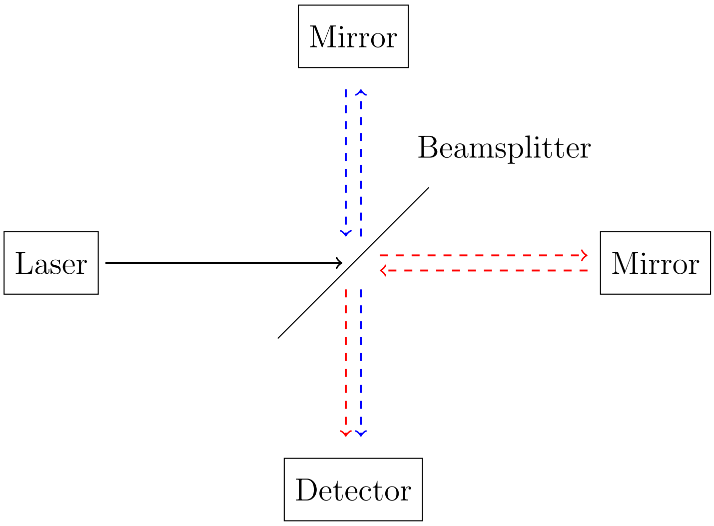


Figure 1.1.: Topology of a Michelson interferometer

Presently feasible simulations cannot capture all of the above stated physical issues. For example, most studies that include microphysics are often only considering Newtonian dynamics. Furthermore, most simulations have not reached convergence in spatial resolution. Additionally it is computationally too expensive to explore the whole space of possible initial conditions in order to match these simulations with observations. On top of the fact that we cannot be completely certain how a GW signal from a core collapse may look like, we face a very complex measurement process due to the small coupling contstant of gravitational waves.

### Observing gravitational waves

In the recent past the laser interferometric gravitational wave observatory LIGO [21] became the most successful observatory for gravitational waves.

In a simplified picture LIGO can be considered to be an advanced Michelson interferometer (fig. 1.1). This laser interferometer can measure small changes of the length of a optical path.
It produces a laser beam which is split into two orthogonal paths. After some distance they are reflected back by a mirror. Changes in the interference pattern of the two reunited beams allow to identify changes in the optical length. From these differential length changes it is possible to deduce the passage of a GW through the detector. The sensitivity of such a laser interferometer can easily be estimated. Suppose the optical paths have length $l_1$ and $l_2$. Consequently the change in the interference of the light is proportional to $\Delta l = \Delta l_1 - \Delta l_2$. GW would induce a strain $h := \Delta l / l$ on the detector, where $h$ is of the order of the GW perturbation. Hence

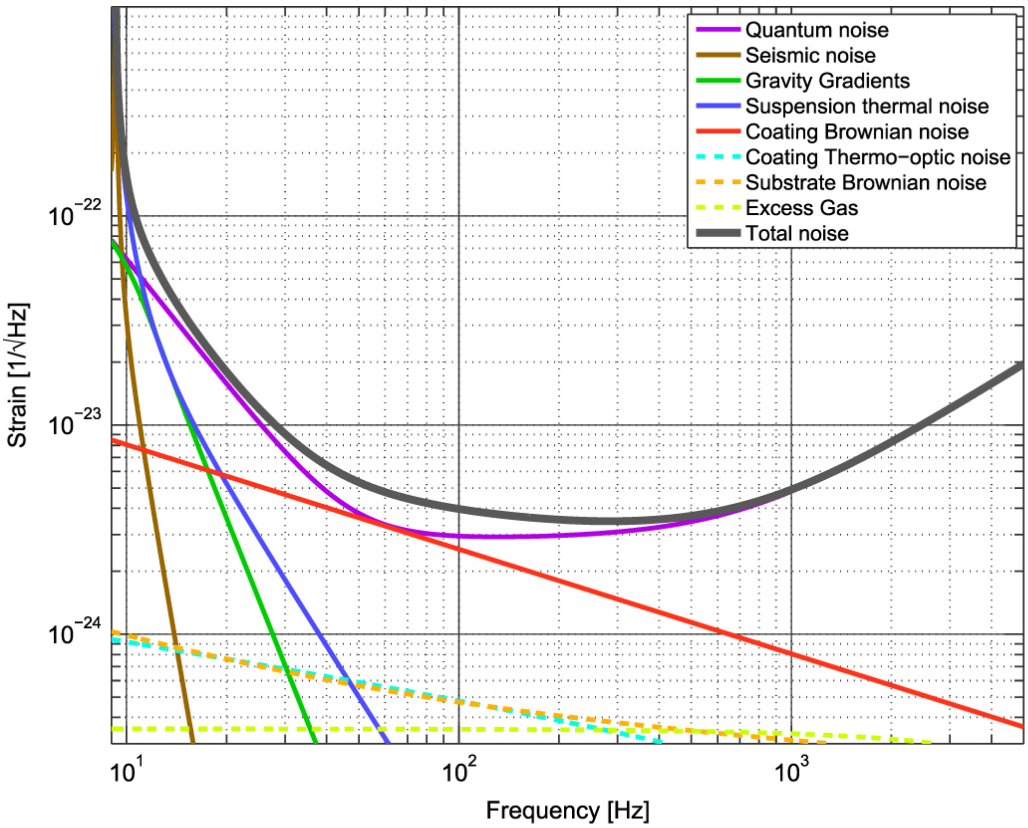


Figure 1.2.: Sensitivity of LIGO, showing the most dominant noise sources. Image credit LIGO scientific collaboration [21].

a measurable effect has to be at the order of the wavelength of the laser, $\Delta l \approx \lambda_{\text{Laser}}$. Assuming $\lambda_{\text{Laser}}$ to be a $\mu m$ then a metric perturbation of the order of

$$h = \frac{\Delta l}{l} \approx \frac{\lambda_{\text{Laser}}}{l} \approx \frac{10^{-6}\text{m}}{10^{3}\text{m}} = 10^{-9} \tag{1.3}$$

would be measurable by a kilometer sized interferometer. Unfortunately this is insufficient to detect GW of a core collapse, whose signal is expected to be $h \approx 10^{-22}$. Consequently a number of advanced techniques, such as beam recycling are used in order to lower the sensitivity significantly. For a detailed description see LIGO Scientific Collaboration *et al.* [21].

As outlined, the detection of a GW depends crucially on the length of the arms of the Michelson interferometer. However, many distortions move the mirrors and/or affect the laser light and therefore mimic or mask a gravitational wave signal. These noise artefacts must be treated in great detail. The sum of all these noise artefacts as a function of frequency are shown in fig. 1.2.

The most dominant sources of noise are [21],

- *Seismic noise*, due to the motion of the mirrors from seismic vibrations.
- *Gravity gradients*, which arise from ground vibrations, wind, ocean waves and human activities.
- *Thermal noise*, from microscopic fluctuations of the individual atoms in the mirrors and their suspensions.
- *Quantum noise*, due to the discrete nature of light and statistical uncertainty of the photodetectors (i.e. Poissonian shot noise).

As the full noise structure and the exact shape of GW from a core collapse are unknown, the detection renders a challenging problem. Hence we must rely on statistical methods which can abstract information from numerical simulations of GW caused by a core collapse, classify their waveforms in an abstract way and then indentify the abstract classes in noisy measurement data. In principle this allows to infer the most plausible set of initial conditions which have lead to the observation. However this is challenging as complex systems such as supernovae cannot be forward simulated for all possible initial conditions as it would be required for an identification of the exact initial conditions. This problem is discussed in greater detail in chapter 2.

### 1.1.2. Neutron stars

Neutron stars are after black holes the densest objects in our Universe. They have masses between one and two solar masses and are almost perfect spheres with radii of about 8 to 15 km. These stars are so compact that general relativistic effects have to be considered to describe their properties. The internal structure of neutron stars is widely unknown. Simple models assume mainly pure baryonic matter with few leptons, i.e. mainly neutrons with a small fraction of protons and electrons. Other models require significant fractions of muons, pions and/or kaons. As the density in the core of the neutron star is assumed to exceed nuclear densities by a factor of a few, it is not guaranteed that the matter inside the core is purely hadronic. Quantum chromodynamics (QCD) predicts the presence of a quark gluon plasma. The aim of all different composition models of neutron stars is to provide a relationship between density and pressure, the equation of state (EoS). The EoS defines the structure of a neutron star and thus can be consulted to compare theoretical models with observations. With a better understanding of the EoS scientists can draw conclusions on the interaction

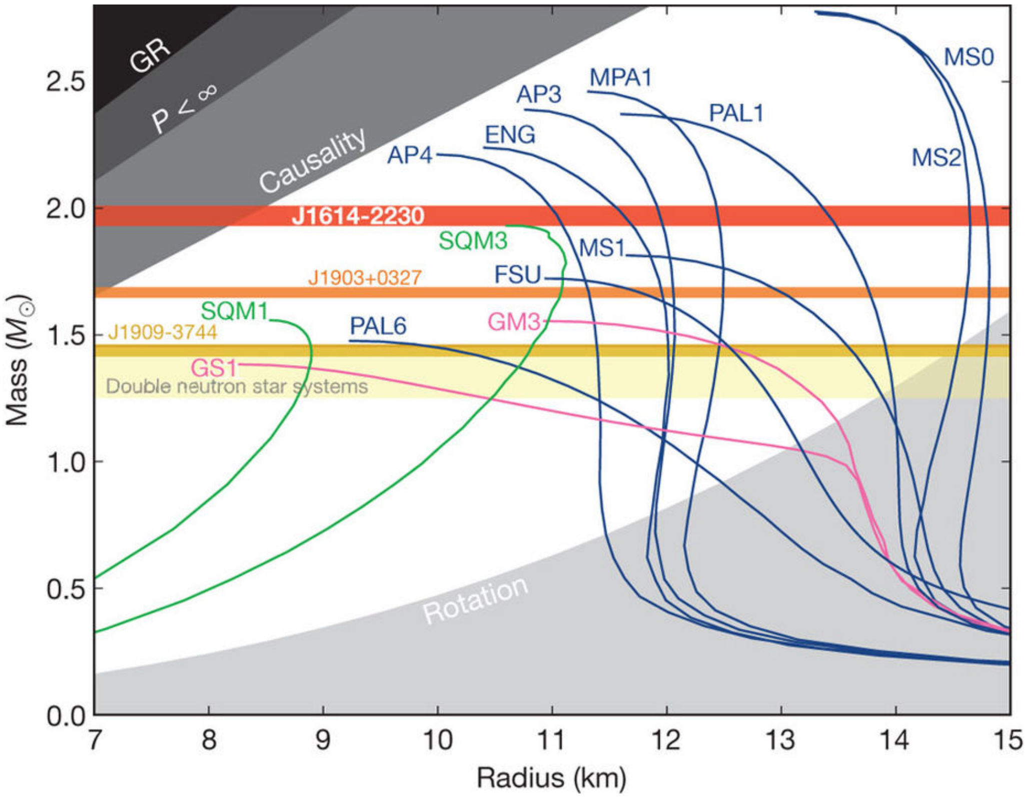


Figure 1.3.: Mass-radius relation for neutron stars with different EoS. Green lines indicate models with strange quark matter, magenta lines models with exotic hadronic matter like kaons or pions, and blue lines give models with ordinary hadronic matter. Horizontal bands give the range of observed masses with error bars. Illustration adapted from [22].

between fundamental particles at an energy scale which will never be accessible on Earth.

In fig. 1.3 the influence of the EoS on the structure of a neutron star is illustrated. The shown models are obtained by providing a central density and integrating the Tolman-Oppenheimer-Volkoff equations [23, 24],

$$\frac{dP}{dr} = -\frac{G\left(\frac{P}{c^2}+\rho\right)\left(m\left(r\right)+\frac{4\pi r^3}{c^2}P\right)}{r^2\left(1-2G\frac{m(r)}{c^2 r}\right)}\,. \tag{1.4}$$

This equation describes the pressure gradient that is needed to compensate the change in gravity as one moves radially by $dr$ in a spherically symmetric static matter distribution described by $m(r)$, its pressure $P$, and energy density $\rho$. Its solution is completely defined by the central density, but different EoS lead to different mass-radius relations. Hence a simultaneous measurement of the mass and the radius of a neutron star could distinguish between different EoS and thus constrain the number of potential theoretically possible models significantly.

However the simultaneous determination of both, mass and radius of a neutron star is a challenging task, and so far the observational uncertainties are large. In consequence there are no strong constraints on the EoS up to date. Via Keplers' law one may obtain estimates about the mass of the neutron star [25]. The Shapiro delay provides masses for neutron star binaries [22]. This method is based on the relativistic time delay of an electromagnetic wave passing through a gravitational potential.
Methods to obtain the radius of a neutron star are based on estimates of the surface of the photosphere of thermally emitting neutron stars or the exterior of thermonuclear explosions of accreted matter on the surface [26–28]. In fig. 1.3 some fundamental constraints on the mass-radius relation are shown. The shaded regions in the upper left corner are forbidden due to theoretical limits, such as causality or the appearance of infinite pressure. The lower shaded areas are forbidden due to the mass shedding limit of the fastest spinning pulsars. Horizontal bands are mass measurements including their error bars. The highest measured mass of about two solar masses excludes already a number of models for the EoS, in particular SQM1 and SQM3, which include strange quark stars, and stars with exotic hadronic matter, which include pions or kaons (GS1, GS3). For details see [22].

In chapter 4 we aim to follow another path to obtain information about the structure of a neutron star and hence its EoS. In contrast to the recently observed gravitational wave signal from a neutron star merger event [1], which carries information about the interior of the star, we use information which is encoded in its electromagnetic radiation. A giant flare of a so called soft gamma-ray repeater (SGR) [29] is considered to carry information about the magneto-elastic oscillations of a magnetar and thus may carry information about the EoS.

### Soft gamma-ray repeater

In todays state of knowledge SGRs are believed to be a subclass of magnetars. Magnetars emit an average energy of a few $\times 10^{35}$ erg/s. The hard X-ray radiation carries a significant part of the total energy, whereas the optical and infrared emission is faint, if detectable at all [30]. The characteristic property of a SGR is the presence of recurring periods of active bursting. Depending on various sources a few up to hundreds of bursts with luminosities of around $10^{42}$ erg/s have been observed. The typical duration of a burst is between 0.1 and 1 s. The quiescent period between these bursts

may have approximatively a log-normal distribution. However, there also exists SGR events that show much more energetic explosions with luminosities of $\approx 10^{45}$ erg/s. Up to date only three of these giant flares have been observed. Giant flares can last for several hundreds of seconds and are so luminous that most detectors have been over saturated by the burst signal. The spectra of the decaying tails are in general softer, and show a complex pulse profile, from which one might constrain the EoS.
We show in chapter 4 a detailed analysis of the observed data of two of such events.

## 1.2. Information Theory

Information theory describes the transmission, processing, and extraction of information. In a more abstract sense it can be thought of as resolution of uncertainty. Information does not necessarily need to be consistent or reliable. For this reason it is natural to quantify the plausibility of information and to this end information theory is mathematically based on probability theory and statistics. I will shortly review a consistent probability theory to quantify the plausibility of information, namely Bayes theorem.

### 1.2.1. Bayesian logic

In Bayesian logic the probability $P(A)$ describes a degree of rational belief in an argument $A$ in terms of a positive real number, from the interval $[0,1]$. By $P(B|A)$ we denote the conditional probability, such that it describes the degree of rational belief in statement $B$ given (or assuming) the truth of statement $A$. Further one can show that the consistency of a probability measure is implemented by the sum rule of probabilities

$$P(B|A) = P\left(B,\bar{C}|A\right) + P(B,C|A) \ , \tag{1.5}$$

and the product rule of probabilities

$$P(A,B|C) = P(A|B,C)\,P(B|C) = P(B|A,C)\,P(A|C) \ , \tag{1.6}$$

with a third statement $C$ and its negotiation $\bar{C}$. With these two rules we can answer the question how likely is $B$ given the statement $A$. Rewriting the product rule and dropping the conditional $C$ for compactness of the notation yields Bayes theorem:

$$P(B|A) = \frac{P(A|B)\,P(B)}{P(A)} \tag{1.7}$$

This theorem describes a natural way of learning by updating rational states of belief and is closely connected to the confrontation of physical theories to physical experiments. We update our physical theories, encoded in $P\left(B\right)$ as we obtain new information. This update involves multiplication with the probability of the data $P\left(A|B\right)$ and division by a new normalisation constant $P\left(A\right)$.

In this context I briefly want to point out an analogy to classical dynamics. In Newtonian dynamics the state of motion of a system is described by the momentum, while a change from one state to another is explained by an applied force. In Bayesian inference this is similar. A state of belief is described in terms of probabilities and the change of rational beliefs is due to information. Hence it follows naturally from the definition of a force, which induces a change from one state to another state, that information is the force which induces a change from one state of belief to another.
This correspondence suggests immediately that one has only obtained enough information if additional information does not change the state of belief any more. Unfortunately such a state is never achievable in any physical experiment. Consequently this gives rise to the obvious question how one can construct methods that allow to infer information from a noisy finite dimensional data set in a information theoretically consistent way.

### 1.2.2. Statistical inference

Up to now we have only treated states of belief. If we intend to infer physical quantities, such as a photon flux and other statistical properties of a phenomena, we must extend the classical Bayesian logic towards a field theory. Lemm [31] showed that this can be done and Enßlin *et al.* [32] named the resulting theory information field theory (IFT). Within IFT the aim is to reconstruct a continuous signal $\boldsymbol{s} = s\left(x\right)$, which lives over some manifold $\Omega$ with $x \in \Omega$, given a finite dimensional noise contaminated data set $d$. In this language Bayes theorem reads,

$$\mathcal{P}\left(\boldsymbol{s}|d\right) = \frac{\mathcal{P}\left(d|\boldsymbol{s}\right)\mathcal{P}\left(\boldsymbol{s}\right)}{\mathcal{P}\left(d\right)} \tag{1.8}$$

where $\mathcal{P}\left(d|\boldsymbol{s}\right)$ describes the likelihood probability density function (PDF), which states how likely a given signal $\boldsymbol{s}$ imprints itself in the observed data $d$. Thereby the likelihood incorporates descriptions of all measurement processes, such as a convolution of $\boldsymbol{s}$ with a point spread function, integration over time and so on. Any a priori knowledge about $\boldsymbol{s}$ is phrased in the prior $\mathcal{P}\left(\boldsymbol{s}\right)$ that is not dependent on $d$. The evidence of $d$ is described by $\mathcal{P}\left(d\right)$

and involves the marginalisation of $\boldsymbol{s}$,

$$\mathcal{P}\left(d\right) = \int \mathcal{D}\boldsymbol{s}\, \mathcal{P}\left(d|\boldsymbol{s}\right) \mathcal{P}\left(\boldsymbol{s}\right) \, . \tag{1.9}$$

In general this infinite dimensional phase space integral is analytically not accessible and makes the calculation of the posterior inaccessible in general. To circumvent these limitations we rely on approximations of the posterior PDF. Typical approaches try to evaluate some statistic of the posterior, for instance its mean or its most likely configuration in the maximum a posteriori approach.
In the following chapters we will show some of these approaches to obtain estimates of $\boldsymbol{s}$ given observed data $d$ and how these may be used to gather reliable information about the physical phenomena outlined in section 1.1.2 and section 1.1.1.

## 1.3. Outline

In chapter 2 we introduce the DSC to classify complex dynamical systems. Thereby we assume that many physical, biological and sociological systems can be described by stochastic differential equations. We propose the usage of simple stochastic differential equations to characterize and classify complex dynamical systems within a Bayesian framework. The DSC first abstracts training data of a system in terms of time dependent coefficients of the descriptive stochastic differential equation. Thereby the DSC identifies unique correlation structures within the training data. For definiteness we restrict the presentation of DSC to oscillation processes with a time dependent frequency $\omega(t)$ and damping factor $\gamma(t)$. Although real systems might be more complex, this oscillator captures many characteristic features. The $\omega$ and $\gamma$ timelines represent the abstract system characterization and permit the construction of efficient signal classifiers. A classifier is trained with only a few samples, in comparison, the training of a neural network classifier requires typically hundreds of samples. Numerical experiments show that the DSC classifier performs well even in the low signal-to-noise regime.

In chapter 3 we develop the D$^4$PO algorithm to denoise, deconvolve, and decompose multidimensional photon observations. The primary objective is to incorporate accurate and well motivated likelihood and prior models in order to give reliable estimates about morphologically different but superimposed photon flux components present in the data set. We denoise and deconvolve photon count data, while simultaneously decomposing them into diffuse, point-like and uninteresting background radiation fluxes. The decomposition is based on a probabilistic hierarchical Bayesian parameter model within the framework of IFT. D$^4$PO can reconstruct several emission components of which each is defined over its own manifold. The manifold over which an entire component is defined may be a direct product of multiple manifolds, such as location, time and/or energys. Beyond that, D$^4$PO may reconstruct correlation structures over each of the components' manifolds separately. A diffuse emission line is for example expected to show rather smooth changes between neighbouring locations and rather monochromatic emission in its energy manifold. In contrast to that are the correlations of a point-flux component which does not show any spatial correlations but rather smooth spectral distribubtions. The inferred correlations can implicitly define the morphologically different source components. The capabilities of the algorithm, to denoise, deconvolve, and decompose multidimensional photon observations are demonstrated with a mock photon count data set, that captures many of the properties and difficulties of real world data sets. D$^4$PO successfully denoises, deconvolves, and decomposes a photon count image into diffuse, point-like and background flux. The algorithm provides a posteriori uncertainty estimates of the reconstructions and the correlation structure of the fields with respect to the manifolds they reside on.

The D$^4$PO algorithm and its predecessor D$^3$PO are applied to study the light curves of the giant flares SGR 1806-20 and SGR 1900+14 in chapter 4. Using this fully noise-aware algorithm, we tried to confirm previously reported frequency lines at $\nu \gtrsim 17\,\mathrm{Hz}$ but cannot. Although they are in the noise-dominated regime of the data, tests with injected signals suggest that D$^4$PO should detect them if as strong as reported in the literature. However, we find two new potential candidates for oscillations at $9.2\,\mathrm{Hz}$ (SGR 1806-20) and $7.7\,\mathrm{Hz}$ (SGR 1900+14). If these are real and the fundamental magneto-elastic oscillations of the magnetars, current theoretical models would favour relatively weak magnetic fields $\bar{B} \sim 6 \times 10^{13} - 3 \times 10^{14}\,\mathrm{G}$ (SGR 1806-20) and a relatively low shear velocity inside the crust compared to previous findings $\bar{B} \sim 1.8 \times 10^{14} - 9 \times 10^{14}\,\mathrm{G}$ (SGR 1806-20).

Finally, we conclude in chapter 6 and give a brief outlook about future prospectives of this thesis.

# 2. Dynamic System Classifier

*This chapter is used additionally as a journal publication in Physical Review E (Pumpe* et al.*, [33]).*
*I am the principal investigator of the research described in the following chapter. My contributions include development of the novel idea, working out the algorithm, implementing and testing it. Further I wrote the chapter. Maksim Greiner helped to find errors in the algorithm and its implementation. Ewald Müller provided me with necessary background information about gravitational wave signatures of core collapse super nova. Torsten Enßlin also fulfilled the role of a principal investigator as he is my PhD supervisor. All authors read, commented, and approved the final manuscript.*

## 2.1. Introduction

### 2.1.1. Basic idea

A classification problem for complex dynamic systems might look as follows: For a number of different system classes a few training samples of the evolution of some of the variables for each class are obtained, either by observation or by numerical simulation. Now, different observed systems should be classified with respect to the reference classes.

As many physical systems tend to be very complex, it is practically impossible to obtain a training data set which precisely resembles the observed one. Therefore, a way to abstract the training samples appropriately, permitting to recognize general characteristics and features of a system class, is desirable. To this end we propose the usage of simplified stochastic dynamical systems with time dependent coefficients to abstract the time evolution of system classes. Once the time dependent coefficients are obtained, they serve as a reference for an inexpensive data-system class comparison. We develop both steps, the coefficient determination and the data classification in a Bayesian inference setting for continuous quantities within information field theory (IFT) [32].

As periodically evolving systems play a remarkable role in the description of dynamical systems we use the stochastic oscillator equation

$$\frac{d^2x(t)}{dt^2} + \gamma\,(t)\,\frac{dx(t)}{dt} + \omega^2(t)x(t) = \xi(t)\,. \tag{2.1}$$

Where the time dependent frequency $\omega(t)$ and damping factor $\gamma(t)$ characterize the system classes and the external force $\xi(t)$ describes the variance within the class. The non-stationarity of eq. (2.1), induced by the time dependent parameters $\gamma(t)$ and $\omega(t)$ allows us to also model anharmonic system classes which do not show a clear oscillating behavior.

The strategy presented in this work is first to analyze system classes in terms of time-wise varying frequency $\omega$ and damping factor $\gamma$. To do so, we apply a hierarchical Bayesian ansatz within the framework of IFT as this also allows us to simultaneously reconstruct the two point correlation function of $\omega$ and $\gamma$. For each analyzed system class, $\omega$ and $\gamma$ serve as abstract classification signatures to which observational data can be confronted. Consequently, we introduce a Bayesian model comparison approach using these signatures to state the probability that a given observation is explained by one of the learned system classes.

### 2.1.2. Previous work

In the last decades, extensive research has been made within the field of estimating parameters of deterministic and stochastic systems, respectively [34–41]. Most attention has been drawn to local and global nonlinear optimization methods. The use of stochastic processes within the global optimization approaches is called *Bayesian global optimization* or the *random function approach*. In a more generic setting general parameter estimation has been performed extensively by maximum likelihood estimations [42]. In financial mathematics the parameter estimation for stochastic models has been extensively studied in a frequentist maximum likelihood as well as in a Bayesian framework [43].

In the past, model selection has primarily been performed by likelihood ratio tests [44]. Due to the enormous increase of computational power, Bayesian methods have been coming into use more and more [45].

The PCA, emerging from Pearson [46] has successfully been used to reduce data to be represented by a linear superposition of a few uncorrelated principle components [47, 48]. As a PCA might perform poorly in case the superposition principle is violated neural networks (NN) for pattern recognition and model selection were used more extensively [49]. Despite the success of NN in the domain of pattern recognition there is still an ongoing discussion how to properly adjust NN to their desired task [50–53].

### 2.1.3. Structure of this work

This paper is structured as follows: In section 2.2 we introduce some basic notation for signal inference in IFT. In section 2.3 we establish the training algorithm to infer the time dependent fields $\omega$ and $\gamma$ from training data. In section 2.4 the inferred fields are used to construct our model selection algorithm. After deriving the whole DSC algorithm we apply it to a realistic test scenario in section 2.5. We conclude in section 2.6.

## 2.2. Inference of fields

### 2.2.1. Basic notation

First of all, we have to establish some notations and basic assumptions. To do so, we primarily follow the notation used in IFT. In this paper we will suppose that we are analyzing a discrete set of data $d = (d_1, \dots, d_\mathrm{r})^\mathsf{T} \in \mathbb{R}^\mathrm{r}$, which may depend on the underlying signal $s : \mathcal{S} \to \mathbb{R}$. By $\mathcal{S}$ we denote the continuous space on which the signal is defined.

From the *principal of minimal updating* [54] and the *principal of maximum entropy* [55] it follows that a Gaussian is the proper probability distribution for a quantity which is only characterized by its first and second momentum, as it often occurs in physical experiments. By

$$\mathscr{G}\left(\phi, \boldsymbol{\Phi}\right) = \frac{1}{|2\pi\boldsymbol{\Phi}|^{1/2}} \exp\left(-\frac{1}{2}\phi^\dagger \boldsymbol{\Phi}^{-1} \phi\right) . \tag{2.2}$$

we denote a multivariate Gaussian probability distribution function of a continuous field $\phi$. $|\boldsymbol{\Phi}|$ denotes the determinant and $\phi^\dagger$ the transposed and complex conjugated $\phi$. The covariance $\boldsymbol{\Phi} = \langle \phi\phi^\dagger \rangle_{\mathscr{G}(\phi,\boldsymbol{\Phi})}$ can be regarded as a function of two arguments $\boldsymbol{\Phi}(x, y) = \langle \phi(x)\phi(y)^\dagger \rangle$ or as a matrix-like operator $\boldsymbol{\Phi}_{xy} = \langle \phi_x \phi_y^\dagger \rangle$, where we introduced the index notations for functions $\psi_x = \psi(x)$. Under the assumptions of a statistical stationary or homogeneous process over a $u$-dimensional space

$$\boldsymbol{\Phi}_{xy} = C(x - y) \tag{2.3}$$

one can show that the covariance $\boldsymbol{\Phi}$ becomes diagonal in Fourier space,

$$\boldsymbol{\Phi}_{kq} = (2\pi)^u \delta(k - q) P_\phi(k) . \tag{2.4}$$

We use $k$ to denote frequencies in the Fourier convention $f_k = \int dt\, e^{ikt} f_t$ and $f_t = \int \frac{dk}{2\pi} f_k e^{-ikt}$. In eq. (2.4) we have also introduced the power spectrum $P(k)$ which is identical to the Fourier transform of $C(x - y)$. In order to apply any operator to a field we have to specify the scalar product, which we take as

$$\phi^\dagger \psi = \int_{\mathcal{S}} dx\, \overline{\phi\left(x\right)} \psi\left(x\right) \quad \forall\, \phi, \psi : \mathcal{S} \to \mathbb{R} . \tag{2.5}$$

Stationary Gaussian processes with zero mean are fully determined by their power spectrum. Here we aim at characterizing non-stationary processes, for which eq. (2.3) and eq. (2.4) do not hold. We construct a non-stationary process $x(t)$ by eq. (2.1), in which a fixed time evolution of $\omega(t)$ and $\gamma(t)$ imprints non-stationary correlation structures onto $x(t)$ for a set of noise $\xi(t)$ realizations. The noise realizations $\xi(t)$ as well as the characteristics of $\omega(t)$ and $\gamma(t)$ of a model class are themselves assumed to be realizations of stationary processes. Nevertheless, $x(t)$ is non-stationary for fixed $\omega(t)$ and $\gamma(t)$.

### 2.2.2. Signal inference

Informative physical experiments provide data from which an unknown signal of interest can be inferred. Since there might be infinitely many possible signal field configurations leading to the same data set, there is no exact solution to this inference problem. Consequently, we have to use probabilistic data analysis methods to obtain the most plausible signal field including its uncertainty.

Given a set of data $d$, the *posterior* probability distribution $\mathcal{P}(s|d)$ describes how probable the signal $s$ is given the observed data $d$. This posterior can be calculated according to Bayes' theorem,

$$\mathcal{P}(s|d) = \frac{\mathcal{P}(d|s)\mathcal{P}\left(s\right)}{\mathcal{P}\left(d\right)}\,, \tag{2.6}$$

which is the quotient of the product of the *likelihood* $\mathcal{P}\left(d|s\right)$ and the signal *prior* $\mathcal{P}(s)$ divided by the *evidence* $\mathcal{P}(d)$. The likelihood describes how likely it is to measure the observed data set $d$ given a signal field $s$. It covers all processes that are relevant for the measurement of $d$. The prior characterizes all a-priori knowledge and therefore does not depend on the data itself. As we are interested to find the most plausible signal field configuration given the observed $d$, we use the *maximum a posteriori* ansatz (MAP). The outcome of the MAP- Ansatz states the most probable field configuration

$$m = \underset{s}{\operatorname{argmax}}\{\mathcal{P}(s \mid d)\}\,. \tag{2.7}$$

In the following sections we will now discuss the likelihood and the prior of the evolution of a stochastic system, which is described by eq. (2.1).

## 2.3. Model Training

### 2.3.1. The likelihood of a stochastic differential equation

As outlined, we use an oscillator model with evolving frequency $\omega$ and damping factor $\gamma$ to characterize a class of systems. To ensure strict positive definiteness of the time dependent frequency we parametrize it as $\omega^2(t) = \omega_0^2 e^{\beta(t)}$ in eq. (2.1). Here we choose time units such that $\omega_0 = 1$.

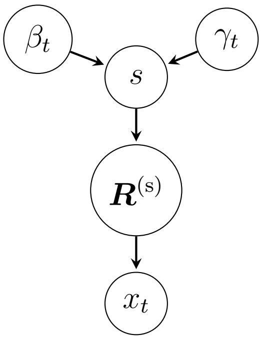


Figure 2.1.: Hierarchical Bayes model for the key quantities. The two fields $\beta_t$ and $\gamma_t$ form together a signal $s$, defining the response $\boldsymbol{R}^{(\mathrm{s})}$, with which the system reacts to the driving white noise $\xi$. The application of $\boldsymbol{R}^{(\mathrm{s})}$ on white Gaussian noise solves eq. (2.1) and consequently yields to its solution $x = \boldsymbol{R}^{(\mathrm{s})}\,\xi$. From sufficient training data vectors $x$ a plausible signal should be inferred in the learning phase.

Consequently, the actual fluctuations of the frequency are characterized by $\beta\,(t)$.

Figure 2.1 shows a Bayesian network to infer the time dependent parameters $\beta_t$ and $\gamma_t$ from the solution $x_t$ of a stochastic differential equation as in eq. (2.1). The inferred vector $s = (\beta_t, \gamma_t)$ will serve as the characteristic signature of a system class in section 2.4. To construct this Bayesian inference network we note that the linearity of eq. (2.1) implies

$$x = \boldsymbol{R}\xi\,, \tag{2.8}$$

and in scalar product notation implying

$$x_t = \int dt' \boldsymbol{R}_{tt'} \xi_t'\,, \tag{2.9}$$

where we have introduced the response operator $\boldsymbol{R}_{tt'} = \boldsymbol{R}(t,t')$. The force $\xi_t$ to the oscillator is assumed to be white Gaussian noise, i.e. $\mathcal{P}\,(\xi|\Xi) = \mathscr{G}\,(\xi, \Xi)$, with a given diagonal and constant covariance $\Xi_{tt'} = \delta_{tt'}\tilde{\Xi}$ and the amplitude $\tilde{\Xi}$. Given that the response operator $\boldsymbol{R}$ and its inverse $\boldsymbol{R}^{-1}$ exist, eq. (2.9) can be rewritten as

$$\int dt'\, \boldsymbol{R}^{-1}_{tt'} x_t = \xi_t\,, \tag{2.10}$$

which allows us to identify the functional form of the response operator $\boldsymbol{R}_{tt'}$ by comparing eq. (2.10) with eq. (2.1). Hence $\boldsymbol{R}_{tt'}$ is a reformulation of the differentials[1] in eq. (2.1) and is defined as

$$\left(\boldsymbol{R}^{(\mathrm{s})}_{\mathrm{tt'}}\right)^{-1} := \delta^{(2)}_{tt'} - \gamma(t)\delta^{(1)}_{tt'} + e^{\beta(t)}\delta_{tt'}\,. \tag{2.11}$$

As already used in eq. (2.11) we will from now on refer to the response operator as $\boldsymbol{R}^{(\mathrm{s})}$ to indicate that it depends on the signal $s$, which is characteristic for a system class. With eq. (2.9) the likelihood $\mathcal{P}(x|s)$ becomes

$$\begin{aligned}\mathcal{P}(x|s) &= \int \mathcal{D}\xi\, \delta(x - \boldsymbol{R}^{(\mathrm{s})}\,\xi)\,\mathscr{G}(\xi, \Xi)\\ &= \mathscr{G}\left(x, \boldsymbol{R}^{(\mathrm{s})^\dagger}\,\Xi\,\boldsymbol{R}^{(\mathrm{s})}\right)\,,\end{aligned} \tag{2.12}$$

where we have marginalized over all possible realizations of the driving force $\xi_t$. We see that $x_t$ is modeled as a Gaussian random field with a temporarily structured covariance $\boldsymbol{X} = \boldsymbol{R}^{(\mathrm{s})^\dagger}\,\Xi\,\boldsymbol{R}^{(\mathrm{s})}$. Thus, $\boldsymbol{X}$ is not of the from given by eq. (2.3) and eq. (2.4), respectively and specified a non-stationary process, which is characterized by $s$.

## 2.3.2. Prior assumptions

As the likelihood in eq. (2.12) only describes how $\beta_t$ and $\gamma_t$ are transformed into $x_t$ we need to model our prior knowledge about the time evolution of these parameters. To do so we briefly outline a hierarchical Bayesian prior ansatz, which leads to the so called *critical filter* [56, 57]. Figure 2.2 shows this hierarchical parameter model, which will be introduced in the following. We assume a priori that $\beta_t$ as well as the $\gamma_t$ parameters obey multivariate Gaussian distributions,

$$\mathcal{P}(\beta_t|\boldsymbol{\Omega}) = \mathscr{G}(\beta_t, \boldsymbol{\Omega})\,, \tag{2.13}$$

$$\mathcal{P}(\gamma_t|\boldsymbol{\Gamma}) = \mathscr{G}(\gamma_t, \boldsymbol{\Gamma})\,. \tag{2.14}$$

The covariances $\boldsymbol{\Omega}$ and $\boldsymbol{\Gamma}$ describe the strength of the temporal correlations of $\beta$ and $\gamma$, respectively and thus the smoothness of their fluctuations. A convenient parametrization of the covariances can be found, if our knowledge of the parameters does a priori not single out any instance; i.e. cor-

---

[1] The defining property of a Dirac delta distribution $\delta(x)$ is

$$\int_{-\infty}^{\infty} f(t)\delta(t-a)dt \equiv \int_{-\infty}^{\infty} f(t)\delta_{ta}dt = f(a)\,.$$

Its derivatives are given by

$$\int f(t)\delta^{(n)}(t)dt \equiv -\int \frac{\partial f(t)}{\partial t}\delta^{(n-1)}(t)dt\,.$$

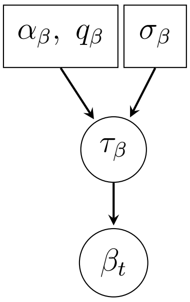


Figure 2.2.: The spectral parameters $\alpha_\beta$ and $q_\beta$ define together with the smoothness enforcing parameter $\sigma_\beta$ the prior for the shape of the spectral parameters $\tau_\beta(k)$. Hence the correlation structure of $\beta_t$ is described by $\tau_k$.

relations only depend on time intervals. This is equivalent to assume $\beta_t$ and $\gamma_t$ to be statistically stationary. Under this assumption, $\mathbf{\Gamma}$ and $\mathbf{\Omega}$ are diagonal in the Fourier space representation, see eq. (2.4).

To keep notations simple, we will only show the calculations for $\mathbf{\Omega}$ in full detail since the ones for $\mathbf{\Gamma}$ can be performed analogously. Later we will see that the priors for $\beta_t$ and $\gamma_t$ only differ by hyperprior-parameters.

Under the assumption of statistical stationarity, we choose the following ansatz for the covariance:

$$\mathbf{\Omega} = \sum_k e^{\tau_\beta(k)} \mathbf{\Omega}_k \tag{2.15}$$

Here $\tau(k)$ are spectral parameters, determining the power spectrum $P_\beta(k)$ and $\mathbf{\Omega}_k$ are projection operators onto spectral bands, with approximately identical power spectrum values. Since the covariances $\mathbf{\Gamma}$ and $\mathbf{\Omega}$ are unknown, one has to introduce another prior for them, i.e. a prior for properly describing the spectral parameters $\tau(k)$, for each $\beta$ and $\gamma$. As the power spectrum's shape might span over several orders of magnitude, this requires a logarithmically uniform prior for each element of the power spectrum and a uniform prior $\mathcal{P}_{\text{un}}$ for each spectral parameter $\tau_k$, respectively. In accordance with [56, 58] we therefore initially assume inverse-Gamma distributions for the individual elements

$$\mathcal{P}_{\text{un}}\left(e^{\tau_k} | \alpha_k,\, q_k\right) = \frac{q_k^{\alpha_k - 1}}{\Gamma\left(\alpha_k - 1\right)} e^{-\left(\alpha_k \tau_k + q_k e^{-\tau_k}\right)} \tag{2.16}$$

and hence

$$\mathcal{P}_{\text{un}}\left(\tau_k|\alpha_k,\, q_k\right) = \frac{q_k^{\alpha_k-1}}{\Gamma\left(\alpha_k-1\right)} \times e^{-\left(\alpha_k\tau_k+q_k e^{-\tau_k}\right)}\left|\frac{de^{\tau_k}}{d\tau_k}\right| \tag{2.17}$$

where $\alpha_k$ and $q_k$ denote shape and scale parameters for the spectral hyperpriors, and $\Gamma$ the Gamma function. If $\alpha_k \to 1$ and $q_k \to 0\,\forall k$, the inverse Gamma distribution becomes asymptotically flat on a logarithmic scale. In practice, $q_k \geq 0$ provide lower limits for the power spectra, which lead to a more stable inference algorithm. Note that the variations of $\alpha_k$ and $q_k$ with $k$ can be used to model specific spectral prior knowledge. However, in the absence of such knowledge, these will get the same values $\alpha_k = \alpha$, $q_k = q = \text{const.}$.

Until now we have only addressed each individual element of the power spectrum separately, but empirically we know that many power spectra do not exhibit strong fluctuations on neighboring Fourier scales. It is therefore natural for the spectrum to request some sort of smoothness. To enforce this behavior, we further incorporate a spectral smoothness prior $\mathcal{P}_{\text{sm}}$ [56, 57]. This spectral smoothness prior is based on the second logarithmic derivative of the spectral parameters $\tau$. Up to a normalization constant $\mathcal{P}_{\text{sm}}$ can be written as

$$\mathcal{P}_{\text{sm}}\left(\tau|\sigma\right) \propto \exp\left(-\frac{1}{2}\tau^{\dagger}\boldsymbol{T}\tau\right) \tag{2.18}$$

with

$$\tau^{\dagger}\boldsymbol{T}\tau = \int d\left(\log k\right)\frac{1}{\sigma^2}\left(\frac{\partial^2\tau_k}{\partial\left(\log k\right)^2}\right)^2 , \tag{2.19}$$

which is punishing deviations from any power-law behavior of the power spectrum. The strength of the punishment is encoded by $\sigma$. In total, the resulting prior for the spectral parameters $\tau$ is given by the product of the priors discussed above

$$\mathcal{P}\left(\tau|\alpha,\, q, \sigma\right) \propto \mathcal{P}_{\text{sm}}\left(\tau|\sigma\right)\prod_k \mathcal{P}_{\text{un}}\left(\tau_k|\alpha_k,\, q_k\right) , \tag{2.20}$$

with its three given quantities $\alpha_k, q_k$ and $\sigma$.

With this hierarchical Bayesian model we are able to state the posterior:

$$\mathcal{P}(s, \tau_\beta, \tau_\gamma|x) = \frac{\mathcal{P}(x|s)}{\mathcal{P}(x)} \times \prod_{i\in\{\beta,\gamma\}} \mathcal{P}(i|\tau_i)\mathcal{P}(\tau_i|\alpha_i,\, q_i,\, \sigma_i) \tag{2.21}$$

Now we are seeking for the most probable parameter configurations of $\beta_t$ and $\gamma_t$ given the training data $x_t$. Due to the complexity of the posterior, given by eq. (2.21), it is virtually impossible to solve this problem analytically. Consequently we use the numerical feasible MAP-ansatz. Rather than maximizing the posterior it is convenient to define the negative logarithm of the posterior $\mathcal{P}\left(s,\tau_\beta,\tau_\gamma|x\right)$ as the information Hamiltonian

$$\begin{aligned}\mathscr{H}\left(s,\tau_\beta,\tau_\gamma|x\right) &= -\log\mathcal{P}\left(s,\tau_\beta,\tau_\gamma|x\right)\\ &= \log\left(\det\left[\boldsymbol{R}^{(\mathrm{s})}\right]\right)\\ &\quad+\frac{1}{2}x_t^\dagger\left({\boldsymbol{R}^{(\mathrm{s})}}^{-1}\right)^\dagger\,\Xi^{-1}\,{\boldsymbol{R}^{(\mathrm{s})}}^{-1}\,x_t\\ &\quad+\frac{1}{2}\log\left(\det\left[\boldsymbol{\Omega}\right]\right)+\frac{1}{2}\beta^\dagger\boldsymbol{\Omega}^{-1}\beta\\ &\quad+\frac{1}{2}\log\left(\det\left[\boldsymbol{\Gamma}\right]\right)+\frac{1}{2}\gamma_t^\dagger\boldsymbol{\Gamma}^{-1}\gamma_t\\ &\quad+\sum_{i\in\{\beta,\gamma_t\}}\left(\alpha_i-1\right)^\dagger\tau+q_i^\dagger e^{-\tau_i}+\frac{1}{2}\tau_i^\dagger\boldsymbol{T}\tau_i\,,\\ &\quad+H_0\,,\end{aligned}\tag{2.22}$$

where we have absorbed all terms constant in $\beta_t$, $\gamma_t$, $\tau_\beta$, and $\tau_\gamma$ into $H_0$. By this reformulation the MAP solution is now seeking for the minimum of eq. (2.22). This minimum may be found by taking the derivative of the information Hamiltonian with respect to $\beta_t$, $\gamma_t$, $\tau_\gamma$, and $\tau_\beta$, respectively and equating them with zero. This yields four implicit equations. The minimum we are seeking for may be found by an iterative downhill algorithm such as the steepest descent. To better understand the MAP solution we focus on the resulting filtering formulas of this ansatz. The ones for the frequency and damping factor evolution read

$$\begin{aligned}\left.\frac{\partial\mathscr{H}}{\partial\beta}\right|_{\beta=\beta_{\min}} &= \mathrm{Tr}\left[{\boldsymbol{R}^{(\mathrm{s})}}^{-1}\,\frac{\partial\,\boldsymbol{R}^{(\mathrm{s})}}{\partial\beta}\right]\\ &\quad+\frac{1}{2}x_t^\dagger\left(\frac{\partial\,{\boldsymbol{R}^{(\mathrm{s})}}^{-1}}{\partial\beta}\right)^\dagger\,\boldsymbol{\Xi}^{-1}\,{\boldsymbol{R}^{(\mathrm{s})}}^{-1}\,x_t\\ &\quad+\frac{1}{2}x^\dagger\left({\boldsymbol{R}^{(\mathrm{s})}}^{-1}\right)^\dagger\boldsymbol{\Xi}^{-1}\frac{\partial\,{\boldsymbol{R}^{(\mathrm{s})}}^{-1}}{\partial\beta}x_t\\ &\quad+\boldsymbol{\Omega}^{-1}\beta_t\\ &\overset{!}{=}0\end{aligned}\tag{2.23}$$

and

$$
\begin{aligned}
\left.\frac{\partial \mathscr{H}}{\partial \gamma_t}\right|_{\gamma_t=\gamma_{\min}} &= \mathrm{Tr}\left[\boldsymbol{R}^{(\mathrm{s})^{-1}}\,\frac{\partial\,\boldsymbol{R}^{(\mathrm{s})}}{\partial \gamma_t}\right] \\
&+ \frac{1}{2} x_t^{\dagger} \left(\frac{\partial\,\boldsymbol{R}^{(\mathrm{s})^{-1}}}{\partial \gamma_t}\right)^{\dagger} \boldsymbol{\Xi}^{-1}\,\boldsymbol{R}^{(\mathrm{s})^{-1}}\,x_t \\
&+ \frac{1}{2} x^{\dagger} \left(\boldsymbol{R}^{(\mathrm{s})^{-1}}\right)^{\dagger} \boldsymbol{\Xi}^{-1} \frac{\partial\,\boldsymbol{R}^{(\mathrm{s})^{-1}}}{\partial \gamma_t} x_t \\
&+ \boldsymbol{\Gamma}^{-1}\gamma_t \\
&\stackrel{!}{=} 0\,.
\end{aligned}
\tag{2.24}
$$

While the filter formula for the power spectra of $\beta_t$ is

$$
\frac{\partial \mathscr{H}}{\partial \tau_\beta} \stackrel{!}{=} 0 \to e^{\tau} = \frac{q_\beta + \frac{1}{2}\left(\mathrm{Tr}\left[\beta_t \beta_t^{\dagger} \boldsymbol{\Omega}_k^{-1}\right]\right)_k}{(\alpha_\beta - 1) + \frac{1}{2}\left(\mathrm{Tr}\left[\boldsymbol{\Omega}_k \boldsymbol{\Omega}_k^{-1}\right]\right)_k + \boldsymbol{T}\tau_\beta}
\tag{2.25}
$$

and the one for $\gamma_t$ is

$$
\frac{\partial \mathscr{H}}{\partial \tau_\gamma} \stackrel{!}{=} 0 \to e^{\tau} = \frac{q_\gamma + \frac{1}{2}\left(\mathrm{Tr}\left[\gamma_t \gamma_t^{\dagger} \boldsymbol{\Gamma}_k^{-1}\right]\right)_k}{(\alpha_\gamma - 1) + \frac{1}{2}\left(\mathrm{Tr}\left[\boldsymbol{\Gamma}_k \boldsymbol{\Gamma}_k^{-1}\right]\right)_k + \boldsymbol{T}\tau_\gamma}\,.
\tag{2.26}
$$

The filtering formulas in eq. (2.25) and eq. (2.26) have previously been derived [57, 59]. Due to the construction of the hierarchical Bayesian parameter model the covariance structures of $\beta_t$ and $\gamma_t$ get also inferred from $x_t$. The spectral shapes of $\boldsymbol{\Omega}$ and $\boldsymbol{\Gamma}$ are only constrained by eq. (2.20).

In total, the model learning phase leads to a restriction of possible signals to the set

$$
\mathscr{S} = \{s_1, s_2, \ldots s_n\}\,,
\tag{2.27}
$$

which is finite in case of a finite number of considered system classes.

## 2.4. Model selection

So far we only faced the inverse problem to reconstruct time dependent parameters, such as the frequency $\beta_t$ and the damping factor $\gamma_t$ including their their power spectra from an oscillator driven by a stochastic force. From now on we will extend our model to a measurement scenario, involving a measurement response $\boldsymbol{R}_{\mathrm{obs}}$ and additive Gaussian measurement noise $n \hookleftarrow \mathscr{G}(n, \boldsymbol{N})$, with related covariance $\boldsymbol{N}$. Consequently the data model is now given by

$$
\boldsymbol{d} = \boldsymbol{R}_{\mathrm{obs}} x + n = \boldsymbol{R}_{\mathrm{obs}}\,\boldsymbol{R}^{(\mathrm{s})}\,\xi + n\,.
\tag{2.28}
$$

where $\boldsymbol{R}^{(\mathrm{s})}$, including $s = (\beta_t, \gamma_t)$, serves as an abstract operator to classify, identify and distinguish between different physical systems $s_1, s_2, \dots$. Hence each $s$ acquires its $\boldsymbol{R}^{(\mathrm{s})}$ from training data of $s$, according to the previously described algorithm. To state the probability of a model $s_i$ in the set of possible signals $\mathscr{S}$ given the observed data we again use Bayes' theorem

$$\mathcal{P}\left(s_i|\mathscr{S}, \boldsymbol{d}\right) = \frac{\mathcal{P}\left(\boldsymbol{d}|\mathscr{S}, s_i\right)\mathcal{P}\left(s_i|\mathscr{S}\right)}{\mathcal{P}\left(\boldsymbol{d}\right)}\,. \tag{2.29}$$

The involved likelihood turns out to be

$$\begin{aligned}\mathcal{P}(d|s_i) &= \int \mathcal{D}x\,\mathcal{P}(d|x)\mathcal{P}(x|s_i)\\ &= \int \mathcal{D}x\;\mathscr{G}(d - \boldsymbol{R}_{\mathrm{obs}}x, N)\\ &\qquad \times \mathscr{G}(x, \boldsymbol{R}^{(\mathrm{s})^\dagger}\,\Xi\,\boldsymbol{R}^{(\mathrm{s})})\\ &\propto \frac{1}{\sqrt{|D|}}\exp\left(\frac{1}{2}j^\dagger D j\right)\end{aligned} \tag{2.30}$$

with

$$j = \boldsymbol{R}^{(\mathrm{s})^\dagger}\,\boldsymbol{R}^\dagger_{\mathrm{obs}} N^{-1} d \tag{2.31}$$

and

$$D^{-1} = \boldsymbol{R}^{(\mathrm{s})^\dagger}\,\boldsymbol{R}^\dagger_{\mathrm{obs}} N^{-1}\boldsymbol{R}_{\mathrm{obs}}\,\boldsymbol{R}^{(\mathrm{s})} + \Xi^{-1}. \tag{2.32}$$

With this equation one is able to calculate the model posterior, eq. (2.29), and to state the most propable model $s_i$.

Figure 2.3 shows an overview of the suggested hierarchical Bayesian decision algorithm. Given the hyper parameters, the algorithm first learns the frequency $\beta_t$ and damping factor $\gamma_t$ evolution from each training data set. The logarithmic power spectrum $\tau$ for $\beta_t$ as well as for $\gamma_t$ can be regarded as a set of nuisance parameters that get reconstructed from the data to properly infer the parameters of interest $\beta_t$ and $\gamma_t$. After $s_i = (\beta_t, \gamma_t)$ was learned for a class $i$, it serves as an abstract characteristic for this model. With the knowledge of $s_i$ the algorithm is able to state how probable the previously learned model $i$ would have caused the observed data $d$. This then serves as a proxy probability for the system classification.

## 2.5. Dynamic system classifier algorithm

Inferring time dependent fields, such as a time-wise varying frequency $\beta_t$ and damping factor $\gamma_t$, from a stochastically driven oscillator is a non-trivial task. The reliability of the dynamic system classifier (DSC) algorithm strongly depends on the successful and proper reconstruction of $\beta_t$ and $\gamma_t$

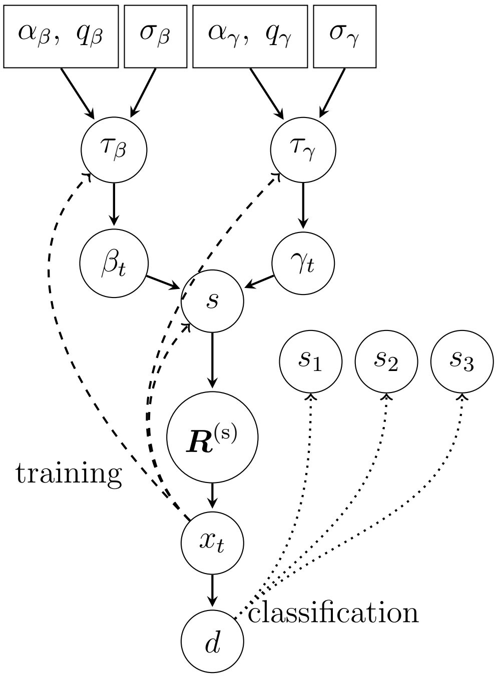


Figure 2.3.: Complete Bayesian network and algorithm. The solid arrows show graphical steps by the probabilistic dependencies. The top four boxes indicate the hyper parameters $\alpha_{\beta/\gamma}$, $q_{\beta/\gamma}$, and $\sigma_{\beta/\omega}$. Below them follow the spectral parameters $\tau_{\beta/\gamma}$ and the tuple of the frequency $\beta_t$ and the damping factor $\gamma_t$, which form together the signal $s$. Each system class $i$ is thereby characterized by its $s_i$. Beneath that comes the response operator $\boldsymbol{R}^{(\mathrm{s})}$, the training data $x_t$, and finally the observed data, according to eq. (2.28). The dashed lines on the left hand side of the figure display the workflow of the algorithm to learn the time evolution of $\beta_t$ and $\gamma_t$, including their power spectra from the training data set. On the right hand side the dotted lines display the workflow of the algorithm to state which of the previously learned training data set explains the observed data $d$ best.

as they serve as classifiers. To show the principal capabilities of our suggested algorithm we will first discuss the algorithm to infer time-dependent parameters of a dynamic system driven by a stochastic force as described in section 2.2. Subsequently, we use the inferred parameters to test the performance of the model selection algorithm as described in section 2.4.

## 2.5.1. Model learning algorithm

### Numerical application

The information Hamiltonian, eq. (2.22), is a scalar quantity defined over the configuration space of possible model parameter evolutions. In addition to the parameters $\beta_t$ and $\gamma_t$, the two spectral parameters $\tau_\beta$ and $\tau_\gamma$ also need to get inferred from a single system trajectory $x(t)$.

Hence, the algorithm faces an underdetermined inverse problem, which is also reflected in the possibility of local-minima of the non-convex Hamiltonian. Ultimately the complexity of this inverse problem goes back to the generally highly non-linear entanglement between the two parameters $\beta_t$ and $\gamma_t$. To overcome this problem we strongly advise to analyze as many realizations ($x = x_1, x_2, x_3, \dots x_l$) of the same system as possible. In section 2.5.1 we will discuss in more detail how many data realizations are necessary for an appropriate reconstruction of the parameters. The training part of the DSC-algorithm is based on an iterative optimization scheme, where certain parts of the problem get alternatively optimized instead of the whole problem simultaneously. To some degree the optimization results are sensitive to the starting values due to the non-convexity of the considered Hamiltonian. However, remaining degeneracies between $\beta_t$ and $\gamma_t$ after exploiting sufficient large training data sets are irrelevant, as these do not strongly discriminate between the members of the training set of a given system class.

Based on our experience with variations of the DSC-algorithm we propose the following scheme:

1. Initialise the algorithm with naive values, such as $\beta_t = \gamma_t = 0$ and $\tau_k = \text{const.}\,\forall k$.

2. Infer $\beta_t$ and $\gamma_t$ via an iterative downhill algorithm, such as steepest descent using the information Hamiltonian eq. (2.22), as well as its gradient eq. (2.23) and eq. (2.24). A more sophisticated minimization scheme, such as non-linear conjugate gradient, is conceivable to speed up the algorithm but it would require the full Hessian of eq. (2.22). Multiple test runs have shown that it is sufficient to evaluate a sim-

plified Hamiltonian

$$\begin{aligned}\mathscr{H}\left(s|x_1,x_2,\dots x_l\right) \propto & \frac{1}{2}\sum_{i=1,\dots} x_i^\dagger \left(\left[\boldsymbol{R}^{(\mathrm{s})}\right]^{-1}\right)^\dagger \Xi^{-1}\left(\boldsymbol{R}^{(\mathrm{s})}\right)^{-1} x_i \\ & + \frac{1}{2}\log\left(\det\left[\boldsymbol{\Omega}\right]\right) + \frac{1}{2}\beta_t^\dagger \boldsymbol{\Omega}^{-1}\beta_t \\ & + \frac{1}{2}\log\left(\det\left[\boldsymbol{\Gamma}\right]\right) + \frac{1}{2}\gamma_t^\dagger \boldsymbol{\Gamma}^{-1}\gamma_t\,,\end{aligned}$$

and its corresponding gradient. The simplified Hamiltonian neglects in particular $\log\left(\det\left[\boldsymbol{R}^{(\mathrm{s})}\right]\right)$ as it appears in eq. (2.22). In contrast to the diagonal covariance matrixes $\boldsymbol{\Omega}$ and $\boldsymbol{\Gamma}$ the evaluation of the determinant of $\boldsymbol{R}^{(\mathrm{s})}$ is computationally time consuming due to its complex structure. Numerical experiments with and without $\det \boldsymbol{R}^{(\mathrm{s})}$ did not show a significant importance of this term.

3. Use eq. (2.25) and eq. (2.26), respectively, to update the priors $\boldsymbol{\Omega}$ and $\boldsymbol{\Gamma}$.

4. Repeat step 2 and 3 until convergence. This iterative scheme will take a few cycles until the algorithm has reached its desired convergence level.

The spaces of possible parameter configurations of $\beta_t$ and $\gamma_t$ are huge. Consequently, it seems impossible to judge whether the algorithm has converged into the desired global minimum or some local minimum. It might also happen that the reconstructed fields display features which are originally caused by the exciting force $\xi$ and not by the frequency and damping factor itself. These problems can be reduced by the above demonstrated joint analysis of multiple data realizations, as we see in the following, where we discuss the numerical tests of the optimization scheme.

### Numerical tests

To test the performance of the DSC algorithm we applied it to simulated but realistic training data sets (see fig. 2.4). This data might represent physical systems whose frequency and damping factor are changing over time. For example, in astrophysics one could expect such a behavior from a gravitational wave burst caused by a supernova [11, 60].

In the following tests we used a regular grid with $10^4$ pixels and the signal inference library NIFTY [61] to implement the algorithm. Figure 2.4 shows six realizations of the same simulated system class. This means that the wave realizations, calculated according to eq. (2.9) with $\tilde{\Xi} = 5$, share the same $\beta_t$ and $\gamma_t$. Note that the waves displayed in fig. 2.4 are not just rescaled versions of the same wave template. For the described MAP re-

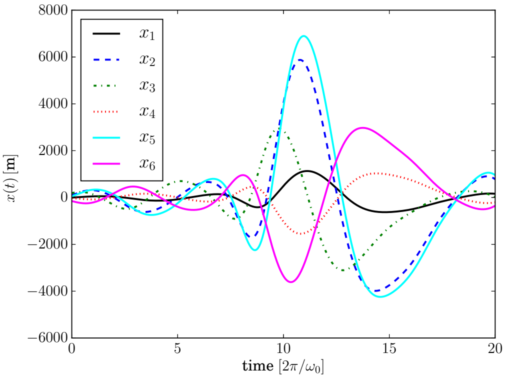
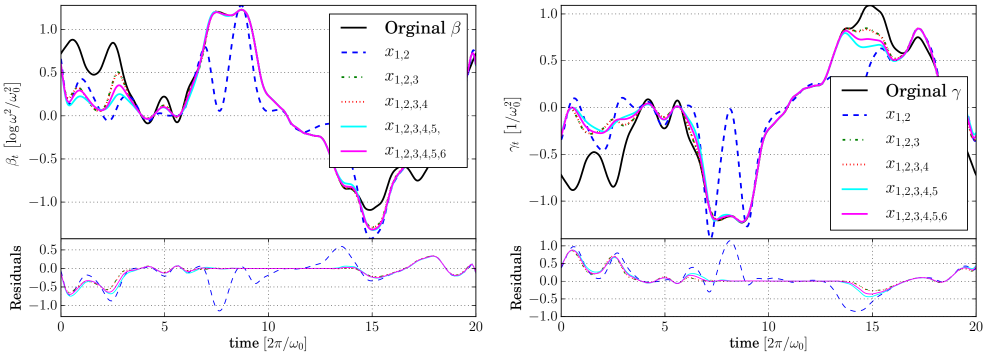


Figure 2.4.: Six wave realizations according to eq. (2.9) for the same model class, i.e. all $x_t$ share the same $\gamma_t$ and $\beta_t$. The amplifying force $\xi$ was drawn from $P(\xi|\Xi) = \mathscr{G}(\xi, \Xi)$, with a given constant covariance $\tilde{\Xi} = 5$.

construction we used $\alpha = 1$, $q = 10^{-30}$, and $\sigma = 2$ for both $\gamma_t$ and $\beta_t$. In fig. 2.5 the inferred parameters, $\beta_{\text{rec}}$ and $\gamma_{\text{rec}}$ including their residuals, as well as the original parameter evolution, $\beta_t$ and $\gamma_t$, are shown. Due to the ahead mentioned degeneracy of the inverse problem we expected that a few wave realizations are needed in order to get reliable reconstructions of the parameters. This is clearly visible in the outcome shown in fig. 2.5, as the residuals between original and reconstructed parameters are largest if one only uses two timelines. Consequently, one needs multiple wave realizations from a system class to get a proper reconstruction of the classe's parameters $\beta_t$ and $\gamma_t$. However, these two parameters are not intended to describe the exact frequency and damping factor evolution of the system. They only serve as an abstract fingerprint of a system class. In case the wave realizations $x(t)$ of a system class do not provide sufficient information, i.e. have only very small amplitudes, the inference problem becomes more and more degenerated. In this case the reconstructed $\beta_{\text{rec}}$ and $\gamma_{\text{rec}}$ do not reproduce precisely the original ones used in our test to generate the simulated waves. Nonetheless this degeneracy does not destroy the performance of the DSC-algorithm because what counts for the model decision algorithm is the ability of the two parameters to represent the covariance structure for a model class and not whether the parameters are as those generating the timelines of the model classes. In fig. 2.6 the inferred power spectrum of $\beta_t$ is shown together with the original one. The power spectrum at small $|k|$, which corresponds to large-scale correlations, is well reconstructed. In contrast, the power spectrum at large $|k|$ is underestimated, which may have various reasons. Small short term variations in $\gamma_t$ are nearly indistinguishable from random noise variations in $\xi$. To better reconstruct the



(a) (b)

Figure 2.5.: Panel (a) shows the original $\beta_t$ as well as its reconstruction for different numbers of data realizations $x_t$ ranging from two ($x_{1,2}$) up to six ($x_{1,2,3,4,5,6}$). The residuals between the original $\beta_t$ and their reconstructions are also shown. Panel (b) shows the same for $\gamma_t$ and its reconstructions. One needs at least three $x_t$ to get a proper reconstruction of the fields. Otherwise, the reconstruction shows too many features imprinted by the driving white noise.

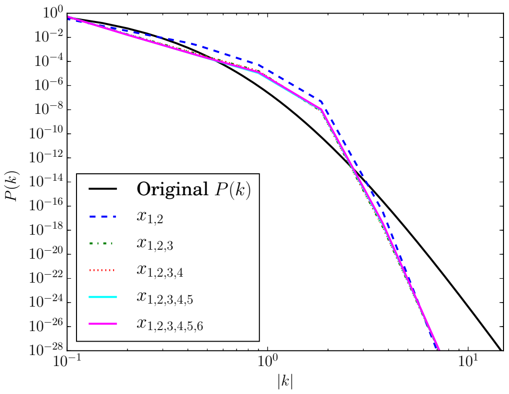
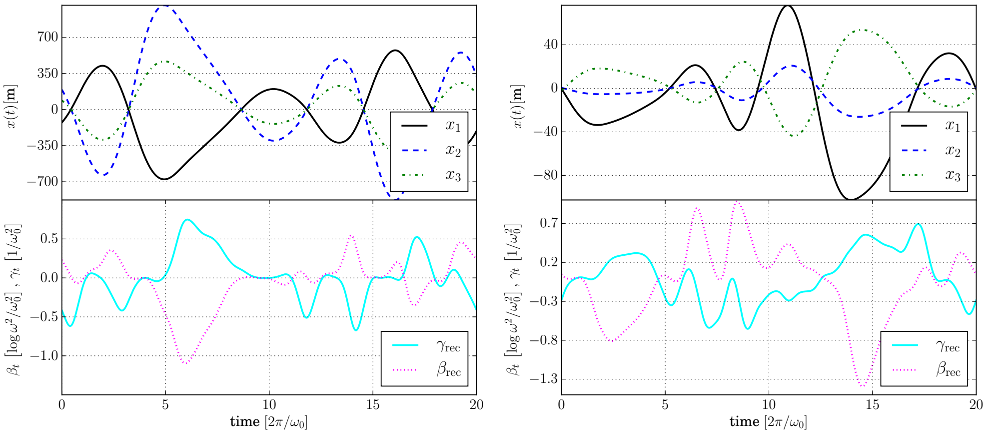


Figure 2.6.: The original power spectrum of $\beta_t$ as well as its reconstruction for different numbers of data realizations $x_t$ used in its reconstruction.



(a) (b)

Figure 2.7.: In panel (a) the training data set of $s_2$ with its reconstructed parameters, $\omega_{\rm rec}$ and $\gamma_{\rm rec}$ is shown. Panel (b) shows the same for $s_3$.

small-scale correlation function many more realizations of the system class would be needed.

In summary we conclude that the presented algorithm can reproduce time dependent parameters from a stochastic differential equation. In the next section we will use the inferred parameters, $\beta_{\rm rec}$ and $\gamma_{\rm rec}$ to discriminate between different models $s_i$.

### 2.5.2. Performance of model selection algorithm

To show that $\beta_t$ and $\gamma_t$ can indeed serve as an abstract fingerprint of classes we trained our algorithm with three different training sets, generated from three different models, specified by the parameters $s_1, s_2$, and $s_3$. Each of them had different $\beta_t$ and $\gamma_t$, and for each model we used three wave realizations using the $s_i$ to train the classificator. By $s_1$ we refer to the training data set in fig. 2.4, by $s_2$ and $s_3$ we refer to the training data sets shown in fig. 2.7. It is a trivial task to distinguish between the models $s_{1,2,3}$, if the means of $\beta_t$ and $\gamma_t$ differ by several orders of magnitude. To avoid such trivial situations, all $\beta_t$'s and $\gamma_t$'s were drawn from the same power spectrum

$$P(k) = \frac{42}{(1+|k|)^{12}} \,. \tag{2.33}$$

In a second step we drew from each $s_i$ a fourth realization, but now added noise according to eq. (2.28). The additive noise $n$ is white and Gaussian, i.e $n \hookleftarrow \mathscr{G}(n, \sigma_{\rm noise})$. $\sigma_{\rm noise}$ was tuned to a specific signal-to-noise ratio (SNR) which we define as

$$\mathrm{SNR} = \frac{\sigma^2_{x(t)}}{\sigma^2_{\mathrm{noise}}}\,. \tag{2.34}$$

Thereby $\sigma_{x(t)}$ refers to the variance of the wave realizations over the displayed period. For the observational response $\boldsymbol{R}_{\mathrm{obs}}$, see eq. (2.28) we assume the unity operator $\boldsymbol{R}_{\mathrm{obs}}(t,t') = \delta_{tt'}$. fig. 2.8 shows two different SNR scenarios for one our waves generated from $s_1$. To demonstrate the performance of the model selection algorithm we choose $P\left(s_i|\mathscr{S}\right) = \frac{1}{3}$ for all $s_i$, i.e. we did not prefer any model.

In Table 2.1 we give the differences of the log-likelihood

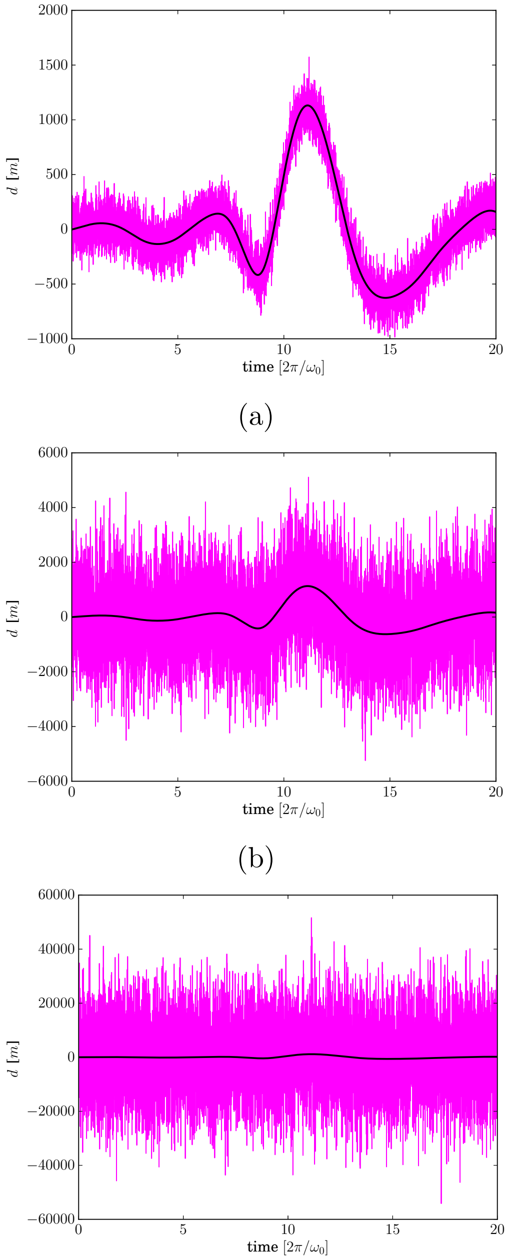
$$\begin{aligned}\Delta_{ij} &= \text{surprise of model } i - \text{surprise of correct model } j\\ &= \mathscr{H}(d|s_i) - \mathscr{H}(d|s_j)\\ &= -\log\mathcal{P}(d|s_i) + \log\mathcal{P}(d|s_j)\end{aligned} \tag{2.35}$$

with $i,j \in \{1,2,3\}$, for various SNR. Note that the information Hamiltonian, the negative log probability, can also be considered as the amount of surprise. The larger the value $\Delta_{ij}$ is, the less plausible an assumed class $i$ is compared to the correct class $j$. It would be as more surprising from the perspective of the data that the assumed model $i$ is correct compared to the correct $j$. Up to a SNR$= 0.01$ all $\Delta_{ij} \geq 0$ and all $\Delta_{ii} = 0$, which means that all datasets are correctly classified. If the SNR is worse than 0.01 the algorithm starts to give misleading classification results, however, only on the $1\sigma$ level, and therefore not with convincing significance. As one intuitively expects the algorithm performs better in case of high SNR, because the absolute differences between the likelihoods are the largest within this regime.

## 2.6. Conclusion

We have established the dynamic system classifier (DSC) algorithm for model selection between dynamic systems. The algorithm consists of two steps. First, it analyzes training data from system classes to construct abstract classifying information for each model class. For distinguishing oscillating systems, a natural basis is the systems' time dependent frequency and damping factor evolution. In the second step the algorithm confronts data with the previously learned models and states the probability, which of the learned models $s$ explains the data best. With these capabilities, DSC is a powerful tool to analyze stochastic and dynamically evolving systems. It can abstract a set of sample timelines into characteristic coefficients which encode a non-stationary correlation structure of the signals of the class.

The theoretical foundations of the first step of DSC is based on a hierarchical Bayesian parameter model within the framework of information field



(a)

(b)

(c)

Figure 2.8.: Panel (a), (b), and (c) show data realizations, eq. (2.28) of $s_1$. The SNR was tuned to 10, 0.1 and 0.001, respectively.

| SNR=0.001 | $s_1$ assumed | $s_2$ assumed | $s_3$ assumed |
|---|---|---|---|
| $s_1$ correct | 0 | 5.10 | -0.82 |
| $s_2$ correct | 7.17 | 0 | 7.62 |
| $s_3$ correct | 1.10 | 0.65 | 0 |
| SNR=.01 | | | |
| $s_1$ correct | 0 | 45.6 | 5.14 |
| $s_2$ correct | 39.4 | 0 | 39.6 |
| $s_3$ correct | 6.02 | 36.5 | 0 |
| SNR=0.1 | | | |
| $s_1$ correct | 0 | 374 | 66.4 |
| $s_2$ correct | 395 | 0 | 353 |
| $s_3$ correct | 60.4 | 209 | 0 |
| SNR=10 | | | |
| $s_1$ correct | 0 | 36100 | 11200 |
| $s_2$ correct | 10500 | 0 | 8720 |
| $s_3$ correct | 311 | 631 | 0 |

Table 2.1.: The four tables above show the performance of the model selection algorithm, for different the signal-to-noise ratio (SNR) of the analyzed data. The printed values denote the relative differences of the log-likelihood according to eq. (2.35).

theory. The model needs *a priori* very few assumptions that account for the statistics and correlations of the two components, $\beta_t$ and $\gamma_t$. Both of them are assumed to obey multivariate Gaussian statistics, whose temporal covariance is described by a power spectrum. The power spectra of both parameter fields are expected to be unknown *a priori*. Therefore, they are also reconstructed from the data, by using the critical filter [56]. This approach is based on the introduction of hyper priors as well as a spectral smoothness enforcing prior [57]. The strength of our proposed and tested DSC algorithm is that it only depends on very few parameters, which can all be motivated *a priori*, and all of them are equally important for the inference.

The classification ability of the DSC-algorithm has successfully been demonstrated in realistic numerical tests, which showed that one needs at least three data realization of each system class in order to be able to sufficiently characterize the frequency and the damping factor evolution. This is due to the high degeneracy of our problem, as we are trying to reconstruct two parameters and their correlation structures from a few timelines. After learning a number of system classes in terms of their characteristic frequency and damping factor evolvements, the algorithm properly classified realistic measurement data. Down to a SNR = 0.01 the algorithm determined for all three test models the correct underlying system class with high significance.

The DSC-algorithm should be applicable to a wide range of inference problems. Concrete examples may be found within in the field of gravitational wave physics.

# 3. Denoising, Deconvolving and Decomposing multi-dimensional Photon Observations

*This chapter as well as appendix A are used as a publication submitted to Astronomy & Astrophysics (Pumpe* et al.*, [62]).*
*I am the principal investigator of the research described in the following chapter. My contributions include the development of the novel idea, working out the algorithm, implementing and testing it. Further I wrote the chapter. Martin Reinecke helped to improve the numerical stability and scalability of the algorithm. Torsten Enßlin also fulfilled the role of a principal investigator as he is my PhD supervisor. All authors read, commented, and approved the final manuscript.*

## 3.1. Introduction

In all data sets of physical experiments, different physical phenomena superimpose each other. This is particularly true for astrophysical sky observations. Point sources, compact objects, diffuse emission, background radiation and other sky structures imprint on the data. Furthermore multiple instrumental distortions such as an imperfect instrument response and noise complicate the data interpretation. In particular the data of high energy photon and particle telescopes are subject to Poissonian shot noise which turns any smooth emission region into a granular image in the detector plane. This paper aims to give reliable estimates of the actual sky components with the help of a rigorous mathematical and statistical treatment.

Individual photon counts are subject to Poissonian shot noise, whose amplitude depends on the count rate itself. Thus, the signal-to-noise ratio (SNR) drops for low count rates. This limits the detection and discrimination of faint sources and hence poses a challenging ill-posed inverse problem. Besides, all telescopes are inexact in the sense that the exposure is inhomogeneous across their field of view. On top of that, the instrument response function may be non-linear and only known up to certain accuracy. Especially point sources on the sky are smeared out in the data plane by the instrument's point spread function, which might let them appear as extended objects in the data plane.

Supplementary to these artifacts present in the data, the observed photon flux is a superposition of multiple morphologically different emission structures that are best decomposed into their original components to better understand their causes. Typical morphologically different emitters in astrophysics can be characterised as diffuse, compact, and point sources. Diffuse objects illuminate the sky over extended areas and show distinct spatial and spectrally extended structures. Point sources, on the other hand, are local features that do not show any spatial structure by definition. They typically have well structured broad band energy spectra. Intermediate sized objects, in between the extremes of point sources and diffuse emission regions, will not be considered here as a separate morphological class. They should either be regarded as part of the diffuse or the point like flux. Furthermore, in addition to the observed photon counts coming from the sky, the recorded counts might contain events due to background radiation, such as cosmic rays and other unwanted sources. If its morphological structure is significantly different from that caused by diffuse and/or point sources, one may be able to distinguish it from sky emission. Overall, this leads to the obvious question how to denoise, deconvolve and decompose the observed data set into its original emission components. This is a hard, ill-posed problem, as there are obviously multiple ways to split the observed counts into the three components discussed here, diffuse, point-source and background radiation flux.

The most popular tool to extract point sources from diffuse emission regions is SExtractor [63], which provides a catalogue of point sources. Furthermore, the CLEAN algorithm [64], widely used in radio interferometery, assumes that the whole sky is composed of point sources, which leads to inferior reconstructions of the diffuse and background emission. Extensions of CLEAN to model the sky with Gaussian blobs to account for diffuse emission structures have improved on this [65, 66].

To treat the outlined inference problem at its roots, we investigate the relation between data and the signals. In this context the signals of interest are the different source contributions. The chosen Bayesian approach allows in a natural way to incorporate a valid data model and a priori knowledge about the source structures to break the degeneracy between the different source contributions. However, this comes at the price of higher computational costs and more complex algorithms.

First attempts in this direction have been pursued by a maximum likelihood analysis [67], followed by maximum entropy analysis [68] and $\chi^2$-methods

[69] which were applied to various astrophysical photon count data sets, such as INTEGRAL/SPI [70], COMPTEL [71] etc.. These methods have in common that they only reconstruct one single component. Within the field of sparse regularisation multiple techniques exploiting various waveforms have been proven to successfully denoise and deconvolve different types of data [72–81].

Disregarding the Poissonian statistics of photon counts, a generic method to denoise, deconvolve and decompose simulated radio data assuming Gaussian noise statistics has been developed [82, 83]. Further in the regime of Gaussian noise statistics, Giovannelli & Coulais [84] developed an algorithm to decompose point and extended sources based on the minimisation of least squares. The algorithm PowellSnakes I/II [85, 86] was successfully applied on the Planck sky data [87]. It is capable of analysing multi-frequency data and to detect point-like sources within diffuse emission regions.

A Bayesian approach close to ours has been developed to separate the background signal from the sky sources in the ROSAT data [88]. This method is based on a two-component mixture model which infers background and diffuse emission concurrently.

This work builds on the D$^3$PO algorithm [89], which was successfully applied to the FERMI LAT data [90]. Here we extend D$^3$PO towards more generic data sets based on photon counts. D$^3$PO denoises, deconvolves and decomposes photon counts into two signals, a point-like and a diffuse one, while it simultaneously reconstructs the spatial power spectrum of the latter. This is done through a hierarchical parameter model incorporating prior knowledge. Here we present an advancement that aims to break the previously assumed statistical isotropy of the reconstructed signal fields and further has the capability to decompose the denoised and deconvolved data set into more than two components, i.e. additional backgrounds. As in D$^3$PO, we derive the algorithm within the framework of information field theory (IFT, [32]), which allows in a natural way the incorporation of priors in the form of a hierarchical prior network. This prior knowledge is crucial as it is used to discriminate between the morphologically different sources via their individual statistical properties. While D$^3$PO could reconstruct the diffuse component depending purely on its location, we show how to incorporate further information present in the data (such as the energy of the photons) to get reconstructions that do not only depend on the location of the reconstruction but also on its energy.

All fluxes, be they diffuse, point-like or background, are modelled individually as signal fields. A field is a continuous quantity defined over a continuous space. A space here is the domain of one or several manifolds or sub-domains. For example the sky emissivity is regarded to be a field living over the product of the two dimensional angular manifold of the celestial

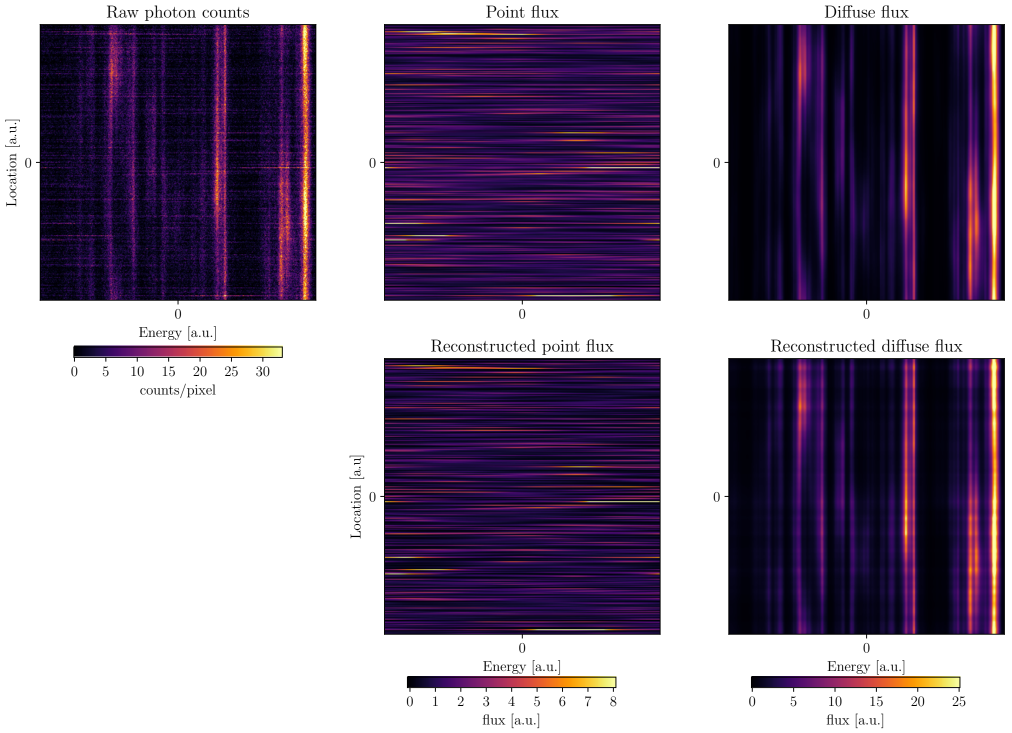


Figure 3.1.: Showcase for the D$^4$PO algorithm to denoise, deconvolve, and decompose multidimensional photon observations. This is a simplified test scenario, which disregards a potential background flux and has only a unit response function (a perfect PSF). Nevertheless it demonstrates the potential capabilities of D$^4$PO, as the reconstructed fluxes are in good agreement with their original ones. All fields are living over a regular grid with $350 \times 350$ pixels.

sphere times a one-dimensional spectral energy manifold. Diffuse continuum emission is a smooth function of both sky position and photon energy. Within each sub-manifold of a field space we assume statistical homogeneity and isotropy separately for the field. This joint field correlation over the composed space is assumed to be a direct product of the sub-domain correlations. This provides the novel possibility to reconstruct fields which do not only depend on one parameter (in this case, location), but on multiple parameters, such as location and energy or energy and time.

Figure 3.1 illustrates exactly such a reconstruction scenario, the denoising, deconvolving, and decomposition of an exemplary data set generated by an emission that consists of diffuse and point-like fluxes, each living over two different sub-domains. The correlation length of the diffuse flux in the spatial sub-domain is significantly longer compared to the correlation length of the point sources which only illuminate the sky at distinct loca-

tions. In contrast are the correlations in the energy sub domain, where as point sources shine over a broader energy range compared to diffuse emission.
The numerical implementation of the algorithm is done in NIFTy3 [61, 91]. This allows us to set up the algorithm in a rather abstract way by being independent of a concrete data set, discretisation and number of fields. Thanks to the abstractness and the flexibility of NIFTy3, reconstructions such as those shown in fig. 3.1 can easily be extended to different domains (i.e. the sphere) and to more source components, i.e. additional backgrounds.

The structure of this work is as follows: In section 3.2 we give a detailed derivation of the outlined algorithm, with an in-depth discussion of the incorporated models and priors. Section 3.3 describes different approaches to calculate the derived posterior. In section 3.4 we discuss numerically efficient implementations of the algorithm. Its performance is demonstrated in section 3.5 by its application to a realistic simulated astrophysical data set. Finally we conclude in section 3.6.

## 3.2. Inference from photon observation

### 3.2.1. Signal inference

Our goal is to image the high energy sky based on photon count data as it is provided by the astroparticle physics instruments like Fermi [92], Integral [70], Comptel [71], CTA [93], and others. The sky emissivity is not only a function of the sky position, but also of photon energy, and it can be thought of as being generated by a number of spectral and morphologically different sources. This work develops the algorithm D$^4$PO that is able to map out the sky emissivity as a function of multiple dimensions, while it simultaneously decomposes it into multiple morphologically different components, such as point sources and diffuse emission. D$^4$PO takes into account the Poisson nature of photon counts as well as the instrument's response function for each individual component.

Due to experimental constraints and practical limitations (such as limited observation time, limited energy range, and limited spatial and spectral resolution), the obtained data set of any photon count experiment cannot capture all degrees of freedom of the underlying photon flux. As a physical photon flux is a continuous scalar field that can vary with respect to various parameters, such as time, location and energy, we let all signals of interest live in a continuous space over some domain $\Omega$. Hence we are facing an underdetermined inference problem as there are infinitely many signal field configurations leading to the same finite data set. Consequently we need to use probabilistic data analysis methods, which do not necessarily provide the physically correct underlying signal field configuration but provide expectations and remaining uncertainties of the signal field.

In this context we are investigating the *a posteriori* probability distribution $P(\boldsymbol{\varphi}|\boldsymbol{d})$, which states how likely a potential signal $\boldsymbol{\varphi}$ is given the data set $\boldsymbol{d}$. This is provided by using Bayes' theorem

$$P(\boldsymbol{\varphi}|\boldsymbol{d}) = \frac{P(\boldsymbol{d}|\boldsymbol{\varphi})P(\boldsymbol{\varphi})}{P(\boldsymbol{d})}\,, \tag{3.1}$$

which is the quotient of the product of the *likelihood* $P(\boldsymbol{d}|\boldsymbol{\varphi})$ and the signal *prior* $P(\boldsymbol{\varphi})$ divided by the *evidence* $P(\boldsymbol{d})$. The likelihood describes how likely it was to observe the measured data set $\boldsymbol{d}$ given a signal field $\boldsymbol{\varphi}$. It should cover all processes that are relevant for the measurement. The prior describes all a priori knowledge on $\boldsymbol{\varphi}$ and must therefore not depend on $\boldsymbol{d}$ itself. Usually we are trying to get the *a posteriori* mean estimate $\boldsymbol{m}$ of the signal field given the data and its uncertainty covariance $\boldsymbol{D}$, which are defined as

$$\boldsymbol{m} = \langle \boldsymbol{\varphi} \rangle_{(\boldsymbol{\varphi}|d)} = \int \mathcal{D}\boldsymbol{\varphi}\, \boldsymbol{\varphi} P(\boldsymbol{\varphi}|\boldsymbol{d})\,, \quad \text{and} \tag{3.2}$$

$$\boldsymbol{D} = \langle (\boldsymbol{m}-\boldsymbol{\varphi})(\boldsymbol{m}-\boldsymbol{\varphi})^\dagger \rangle_{(\boldsymbol{\varphi}|\boldsymbol{d})}\,, \tag{3.3}$$

where † denotes adjunction and $\langle\cdot\rangle_{(\boldsymbol{\varphi}|\boldsymbol{d})}$ the expectation value with respect to the posterior probability distribution $P(\boldsymbol{\varphi}|\boldsymbol{d})$.

In the following sections we will gradually derive the posterior of the physical flux distribution of multiple superimposed photon fluxes given in a data set. This will partly follow and build on the existing D$^3$PO algorithm by [89].

### 3.2.2. Poissonian likelihood

A typical photon count instrument provides us with a data vector $d$ consisting of integer photon counts that are spatially binned into $N_{\mathrm{PIX}}$ pixels. The photon flux $\boldsymbol{\rho} = \rho(x, E)$, which caused the photon counts, is defined for each continuous sky position $x$ and energy $E$. Since high energy astrophysical spectra cover orders of magnitude in energy, it is convenient to introduce the logarithmic energy coordinate $y = \log(E/E_0)$ with some reference energy $E_0$. The flux is a function of the combined coordinates $z = (x, y)$, such that $\rho(z) = \rho(x, y)$ and lives over the combined domain $\Omega = \mathbb{S} \otimes \mathbb{K}$, the product of spatial and spectral domains. $\mathbb{S} = \mathcal{S}^2$ if the full sky is treated, $\mathbb{S} = \mathbb{R}^2$ if a patch of the sky is described as in the flat sky approximation, or $\mathbb{S} = \mathbb{R}$ as in the mock examples, $\mathbb{K} = \mathbb{R}$ as logarithmic energies can be positive and negative. The flux $\boldsymbol{\rho}$ is a superposition of two morphologically different signal fields, such as a diffuse and point-like flux. Hence

$$\begin{aligned}\boldsymbol{\rho}\,(z) &= \boldsymbol{\rho}\,(z)_{\text{diffuse}} + \boldsymbol{\rho}\,(z)_{\text{point-like}} \\ &= \rho_0 \left(e^{s(z)} + e^{u(z)}\right) \end{aligned} \tag{3.4}$$

where we introduced the dimensionless fields $s(z) =: \boldsymbol{s}$ and $u(z) =: \boldsymbol{u}$, to represent the logarithmic diffuse and point-source fluxes over the signal domain $\Omega$. Further we introduced the convention that a scalar function is applied to a field value by values on its natural domain, $(f\,(s))\,(z) = f\,(s\,(z))$ if $z \in \Omega$. The constant $\rho_0$ is absorbing the physical dimensions of the photon flux.

The imaging device observing the celestial photon flux $\boldsymbol{\rho}$ is expected to measure a number of events $\boldsymbol{\lambda}$, informing us about $\boldsymbol{s}$ and $\boldsymbol{u}$, as well as other event sources like cosmic ray hits or radioactivity, which we will refer to as background. This dependence of $\boldsymbol{\lambda}$ on the sky emissivity $\boldsymbol{\rho}$ can be modelled via a linear instrument response operator $\boldsymbol{R}_0$ acting on it:

$$\begin{aligned}\boldsymbol{\lambda} &= \boldsymbol{R}_0\boldsymbol{\rho} + \boldsymbol{R}_0'\boldsymbol{\rho}' \\ &= \boldsymbol{R}\left(e^{\boldsymbol{s}} + e^{\boldsymbol{u}}\right) + \boldsymbol{R}'e^{\boldsymbol{b}}\,, \end{aligned} \tag{3.5}$$

with $\boldsymbol{R} = \boldsymbol{R}_0\rho_0$ and $\boldsymbol{R}' = \boldsymbol{R}_0'\rho_0'$. The response operator $\boldsymbol{R}_0$ describes all aspects of the measurement processes which relate the sky brightness to the average photon count. To describe the background counts we further introduce the background event emissivity $\boldsymbol{\rho}' = \rho_0'e^{\boldsymbol{b}}$, the background event instrument response $\boldsymbol{R}_0'$, as well as the abbreviation $\boldsymbol{R}' = \boldsymbol{R}_0'\rho_0'$. The background field is a function of the two coordinates exposure time $t$ and the log-energy $y$, such that $z' = (t, y)$ and its domain $\Omega' = \mathbb{T} \otimes \mathbb{K}$ with $\mathbb{T} \in \mathbb{R}$. In case the sky is scanned once location by location, the observing time interval and the sky may be identified with each other. In our mock examples, we will assume this for better visualisation of the data. In reality, the sky is often scanned multiple times, providing redundancies in the data, which facilitates the separation of sky and background. The background response $\boldsymbol{R}'$ describes the instruments sensitivity to background processes, such as cosmic ray events which produce counts but are not of interest and need to be filtered out. For each pixel in the detector we get

$$\lambda_i = \int_\Omega dz\, R_i(z)\left(e^{s(z)} + e^{u(z)}\right) + \int_{\Omega'} dz'\, R_i'(z')e^{b(z')}\,. \tag{3.6}$$

The individual events are assumed to follow a Poisson distribution $\mathcal{P}(\boldsymbol{d}_i|\boldsymbol{\lambda}_i) = \boldsymbol{\lambda}_i^{\boldsymbol{d}_i}e^{-\boldsymbol{\lambda}_i}/(\boldsymbol{d}_i!)$ in each data bin. Hence the likelihood of the data vector $\boldsymbol{d} = (d_1, \ldots, d_{N_{\text{PIX}}})$ given an expected number of photons $\boldsymbol{\lambda}$ is a product of statistically independent Poisson processes,

$$P(\boldsymbol{d}|\boldsymbol{\lambda}) = \prod_i \mathcal{P}(d_i, \lambda_i) = \prod_i \frac{1}{d_i!} \lambda_i^{d_i} e^{-\lambda_i} \,. \tag{3.7}$$

In total, the likelihood of photon count data given two superimposed morphologically different photon fluxes and some background fluxes is consequently described by eq. (3.6) and eq. (3.7). To simplify and clarify the notations in the following, we introduce the three component vector $\boldsymbol{\varphi} = (\boldsymbol{s}^\dagger, \boldsymbol{u}^\dagger, \boldsymbol{b}^\dagger)^\dagger$ consisting of the diffuse, point-like and the background flux. Further we introduce the combined response $\boldsymbol{\mathcal{R}} = (\boldsymbol{R}, \boldsymbol{R}, \boldsymbol{R}')$. This allows us to state the information Hamiltonian, defined as the negative logarithm of $P(\boldsymbol{d}|\boldsymbol{\varphi})$, compactly as

$$\begin{aligned} H(\boldsymbol{d}|\boldsymbol{\varphi}) &= -\log P(\boldsymbol{d}|\boldsymbol{\varphi}) \\ &= H_0 + \boldsymbol{1}^\dagger \boldsymbol{\mathcal{R}} e^{\boldsymbol{\varphi}} - \boldsymbol{d}^\dagger \log\left(\boldsymbol{\mathcal{R}} e^{\boldsymbol{\varphi}}\right) \,, \end{aligned} \tag{3.8}$$

where we absorbed all terms that are constant in $\boldsymbol{\varphi}$ into $H_0$ and introduced the scalar product of concatenated vectors

$$\begin{aligned} \boldsymbol{\varphi}^\dagger \boldsymbol{\varphi} =& \boldsymbol{s}^\dagger \boldsymbol{s} + \boldsymbol{u}^\dagger \boldsymbol{u} + \boldsymbol{b}^\dagger \boldsymbol{b} \\ =& \int_\Omega dz \left( \overline{s\,(z)} s'\,(z) + \overline{u\,(z)} u'\,(z) \right) + \int_{\Omega'} dz' \, \overline{b\,(z)} b'\,(z) \,. \end{aligned} \tag{3.9}$$

$\boldsymbol{1}^\dagger$ is a constant data vector being 1 everywhere on the data space.

As an aside it may be noted that the likelihood, eq. (3.7), is also valid for $n$ signals in data space.

$$\boldsymbol{\lambda} = \sum_{\mathrm{i}=1}^{\mathrm{n}} R_{\mathrm{i}} e^{s_{\mathrm{i}}(z)} \,. \tag{3.10}$$

This would only extend $\boldsymbol{\varphi}$ and $\boldsymbol{\mathcal{R}}$ while eq. (3.8) stays untouched. For this reason, D$^4$PO is designed such that it can reconstruct $n$ such fields, each living over its own space, and all contributing to the expected counts through an individual response function.

### 3.2.3. Prior assumptions

As we are seeking to decompose the photon counts into background flux and two sky flux components we have to introduce priors. Otherwise the inverse problem would be completely degenerate since it would be possible to explain the full data set just by the diffuse, just by the point-like signal alone or even purely by a background or any combination thereof. In order to break this degeneracy we introduce priors on the fields $\boldsymbol{s}, \boldsymbol{u}$, and $\boldsymbol{b}$.

Hence we will discuss the priors which describe and implicitly define the typical expected morphology of the three different expected types of fluxes, the diffuse, point-like and background flux.

### The diffuse component

The diffuse photon flux $\rho^{(\boldsymbol{s})} = \rho_0 e^{\boldsymbol{s}}$ is a strictly positive quantity which might vary over several orders of magnitude. Its morphology may be described by cloud-like and smoothly varying patches on the sky. Hence the diffuse flux shows spatial correlations. Furthermore, we concentrate here on radiation processes due to cosmic ray interactions in the interstellar medium like the inverse Compton scattering and $\pi^0$ production and decay. These have smooth, often power-law like spectra, which therefore also show some considerable correlation in the log-energy dimension. According to the principle of maximum entropy the log-normal model can be regarded as a minimalistic description of our a priori knowledge on $\rho^{(\boldsymbol{s})}$ [57, 94]. The log-normal model has shown to be suitable within various observational [90, 95, 96] and theoretical considerations [33, 89, 97–105]. Hence, we adopt a multivariate Gaussian distribution as a prior for the logarithmic $\boldsymbol{s}$:

$$\mathcal{G}(\boldsymbol{s}, \mathcal{S}) = \frac{1}{\sqrt{2\pi|\mathcal{S}|}} \exp\left(-\frac{1}{2}\boldsymbol{s}^\dagger \mathcal{S}^{-1} \boldsymbol{s}\right) \tag{3.11}$$

with a covariance $\mathcal{S} = \langle \boldsymbol{s}\boldsymbol{s}^\dagger \rangle_{(\boldsymbol{s}|\mathcal{S})}$. The covariance $\mathcal{S}$ describes the strength of spatial correlation in the energy $y = \log E/E_0$ and space domain $x$ of $\boldsymbol{s}$. As these two correlations need to be modelled individually, we choose the following ansatz

$$\mathcal{S}_{zz'} = \mathcal{X}^{(s)}(|x - x'|)\mathcal{Y}^{(s)}(|y - y'|)\,, \tag{3.12}$$

being a direct product of the two correlation functions, $\mathcal{X}^{(s)}_{x,x'} = \mathcal{X}^{(s)}\left(|x - x'|\right)$ and $\mathcal{Y}^{(s)}_{yy'} = \mathcal{Y}^{(s)}(|y - y'|)$, which only depend on the relative differences in position $x$ and log-energy $y$. This is equivalent to the assumption that $\boldsymbol{s}$ is statistically homogeneous and isotropic on the celestial sphere and statistically homogeneous in the log-energy space. Statistical homogeneity in log-energy models the fact that typical high energy astrophysics spectra exhibit similar features on a log-log perspective. Thanks to the assumed statistical homogeneity we can find a diagonal representation of $\mathcal{X}$ and $\mathcal{Y}$

in their harmonic bases, such that

$$\mathcal{X} = \sum_l e^{\tau_{\mathcal{X}}(l)} \mathbb{P}_l \,, \quad \text{and} \tag{3.13}$$

$$\mathcal{Y} = \sum_k e^{\tau_{\mathcal{Y}}(k)} \mathbb{P}_k \,. \tag{3.14}$$

Here $\tau_{\mathcal{X}}(l)$ and $\tau_{\mathcal{Y}}(k)$ are spectral parameters determining the logarithmic power spectra in the spatial and spectral domain, respectively, with respect to the chosen harmonic basis denoted by $l$ and $k$ for their corresponding harmonic spaces. $\mathbb{P}$ is a projection operator onto spectral bands, and we assume a similar level of variance for the fields within each of these bands separately. The projection operator is given by:

$$\mathbb{P}_k \equiv \sum_{k' \in b_k} F_{k'} F_{k'}^{\dagger} \tag{3.15}$$

where $F_{k'y} = e^{ik'y}$ is the Fourier basis (or spherical harmonic basis if the subdomain is $\mathcal{S}^2$) and $b_k$ denotes the set of Fourier modes belonging to the corresponding Fourier band $k$ and $l$, respectively. The inverses of the two covariances are

$$\mathcal{X}^{-1} = \sum_l e^{-\tau_{\mathcal{X}}(l)} \mathbb{P}_l \,, \tag{3.16}$$

$$\mathcal{Y}^{-1} = \sum_k e^{-\tau_{\mathcal{Y}}(k)} \mathbb{P}_k \,. \tag{3.17}$$

As the spectral parameters $\tau_{\mathcal{X}}$ and $\tau_{\mathcal{Y}}$ are in general unknown *a priori* and therefore need to be reconstructed from the same data as the signal we have to introduce another prior for their covariances.

In the following paragraphs we will introduce two constraints on the spectral parameters $\tau$ . These hyperpriors on the prior itself lead to a hierarchical parameter model. To shorten and clarify notations we will only discuss $\tau_{\mathcal{X}}(l)$ in full detail; the expressions for $\tau_{\mathcal{Y}}(k)$ are analogous.

As a further aside it may be noted that $\boldsymbol{s}$ may not only depend on location and energy but also on further parameters such as time etc. In this case the covariance $\mathcal{S}$ would become

$$\begin{aligned} \mathcal{S}_{zz'} &= \prod_{\mathrm{i}} \mathcal{X}_{\mathrm{i}} \left( |x_{\mathrm{i}} - x'_{\mathrm{i}}| \right) \quad \text{with} \\ z &= (x_1, x_2, \ldots, x_n)^{\dagger} \\ z' &= (x'_1, x'_2, \ldots, x'_n)^{\dagger} \end{aligned} \tag{3.18}$$

leading in consequence to multiple spectral parameters that need to be inferred from the data if not known a priori. D$^4$PO is also prepared to handle such cases.

### Unknown magnitude of the power spectrum

As the spectral parameters $\tau_{\mathcal{X}}(l)$ might vary over several orders of magnitude, this demands for a logarithmically uniform prior for each element of the power spectrum and in consequence for a uniform prior $P_{\text{un}}$ for each spectral parameter $\tau_{\mathcal{X}}(l)$. In accordance with [56, 58] and [89] we initially assume inverse-Gamma distributions for each individual element,

$$P_{\text{un}}(e^{\tau}|\alpha_l, q_l) = \prod_l \frac{q_l^{\alpha_l - 1}}{\Gamma(\alpha_l - 1)} e^{-(\alpha_l \tau_l + q_l e^{-\tau_l})} \tag{3.19}$$

and hence,

$$\begin{aligned} P_{\text{un}}\left(\tau|\alpha_l,\, q_l\right) = \prod_l \frac{q_l^{\alpha_l - 1}}{\Gamma\left(\alpha_l - 1\right)} \\ \times\, e^{-\left(\alpha_l \tau_l + q_l e^{-\tau_l}\right)} \left|\frac{de^{\tau_k}}{d\tau_l}\right| \,, \end{aligned} \tag{3.20}$$

where $\alpha_l$ and $q_l$ denote shape and scale parameters for the spectral hyperpriors, and $\Gamma$ the Gamma function. The form of eq. (3.20) shows that for $\alpha_l \to 1$ and $q_l \to 0, \forall l > 0$, the inverse gamma distribution becomes asymptotically flat on a logarithmic scale and therefore for these parameter values do not provide any constraints on the magnitudes of $\tau$.

### Smoothness of power spectrum

Up to now we have only treated each element of the power spectrum separately, permitting the power to change strongly as a function of scale $l$ (and $k$). However, similar spatial (or energetic) scales should exhibit similar amounts of variance for most astrophysical emission processes in the interstellar medium. Thus we assume that the power spectrum is smooth on a logarithmic scale of $l$ and $k$, respectively. In accordance with [56] and [57] this can be enforced by introducing a smoothness enforcing prior $P_{\text{sm}}$

$$\begin{aligned} P_{\text{sm}}(\tau|\sigma) &\propto \exp\left(-\frac{1}{2\sigma^2}\int d(\log l)\,\left(\frac{\partial^2 \tau_l}{\partial\left(\log l\right)}\right)^2\right) \\ &\propto \exp\left(-\frac{1}{2}\tau^{\dagger}\boldsymbol{T}\tau\right)\,, \end{aligned} \tag{3.21}$$

which is based on the second logarithmic derivative of the spectral parameters $\tau$. The parameter $\sigma$ specifies the expected roughness of the spectrum. In the limit $\sigma \to \infty$ spectral roughness is not suppressed in contrast to $\sigma \to 0$ which enforces a smooth power-law power spectrum. For a more detailed explanation and demonstration of the influence of $\sigma$ on the reconstruction of $\tau$ we refer to [96].

In total the resulting priors for the diffuse flux field are determined by the spectral parameters $\tau_{\mathcal{X}^{(\boldsymbol{s})}}$ for spatial correlation and $\tau_{\mathcal{Y}^{(\boldsymbol{s})}}$ for correlations in the log-energy domain. These spectral parameters are constrained by the product of the priors discussed above

$$P\left(\tau_i|\alpha_i, q_i, \sigma_i\right) = P_{\mathrm{sm}}\left(\tau_i|\sigma_i\right) P_{\mathrm{un}}\left(\tau_i|\alpha_i, q_i\right) \,, \tag{3.22}$$

with $i \in \{\mathcal{X}^{(\boldsymbol{s})}, \mathcal{Y}^{(\boldsymbol{s})}\}$.

### The point-like component

Due to the large distances of many astrophysical sources to earth, they appear as point-sources despite their actual extension. The photon flux contributions of neighboring point sources can be assumed to be statistically independent of each other if we decide to ignore knowledge on source clustering. Although two point sources might appear to be very close to each other on the sky, their physical distance might be enormous. Consequently, statistically independent priors for the photon flux contribution of each point-source are introduced in the following.

As the point-like flux $\rho^{(\boldsymbol{u})} = \rho_0 e^{\boldsymbol{u}}$ is also a strictly positive quantity, we mainly follow the same arguments as for the diffuse flux to derive its prior for correlations within in the log energy domain. Following the same arguments as in section 3.2.3 we adopt for the spectral correlation in the energy domain $y$ of $\boldsymbol{u}$ a multivariate Gaussian distribution $\mathcal{G}(\boldsymbol{u}, \mathcal{Y}^{(u)})$, with $\mathcal{Y}^{(u)} = \langle \boldsymbol{u}\boldsymbol{u}^\dagger \rangle_{(\boldsymbol{u}|\boldsymbol{d})}$. As we may again assume statistical homogeneity, we can find a diagonal representation of the yet unknown $\mathcal{Y}^{(u)}$, such that

$$\mathcal{Y}^{(u)} = \sum_k e^{\tau_{\mathcal{Y}^{(\boldsymbol{u})}}(k)} \mathbb{P}_k \,, \tag{3.23}$$

with $\mathbb{P}$ being the projection operator according to eq. (3.15), $\tau_{\mathcal{Y}^{(\boldsymbol{u})}}(k)$ being the spectral parameter determining the logarithmic power spectrum of $\mathcal{Y}^{(u)}$. Since $\tau_{\mathcal{Y}^{(\boldsymbol{u})}}(k)$ is usually unknown we introduce again a hierarchical parameter model as in sections 3.2.3 and 3.2.3. This gives us

$$\begin{aligned} P(\tau_{\mathcal{Y}^{(\boldsymbol{u})}}|\alpha_{\mathcal{Y}^{(\boldsymbol{u})}}, q_{\mathcal{Y}^{(\boldsymbol{u})}}, \sigma_{\mathcal{Y}^{(\boldsymbol{u})}}) =& P_{\mathrm{sm}}(\tau_{\mathcal{Y}^{(\boldsymbol{u})}}|\sigma_{\mathcal{Y}^{(\boldsymbol{u})}}) \\ & \times P_{\mathrm{un}}(\tau_{\mathcal{Y}^{(\boldsymbol{u})}})|\alpha_{\mathcal{Y}^{(\boldsymbol{u})}}, q_{\mathcal{Y}^{(\boldsymbol{u})}}) \,, \end{aligned} \tag{3.24}$$

with $\alpha_{\mathcal{Y}(\boldsymbol{u})}$ and $q_{\mathcal{Y}(\boldsymbol{u})}$ being the scale and shape parameters of the inverse gamma distribution and $\sigma_{\mathcal{Y}(\boldsymbol{u})}$ the parameter to specify the expected roughness of the spectrum.

To derive a suitable prior for the spatial dimension of $\boldsymbol{u}$, one may look at the following basic considerations. Let us assume that the universe hosts a homogenous distribution of point sources. Therefore the number of point sources would scale with the observable volume, i.e. with distance cubed. But the apparent brightness of a source is reduced by the spreading of the light rays, i.e. decreases with distances squared. Hence one may expect a power law behaviour between the number of point sources and their brightness with a slope of $\beta = 3/2$. As such a power-law is not necessarily normalisable, since it diverges at zero we further impose an exponential cut-off slightly above 0. This yields an inverse Gamma distribution, which has been shown to be a suitable prior for point-like photon fluxes [85, 86, 88–90]. The spatial prior for $\boldsymbol{u}$ is therefore given by a product of independent inverse-Gamma distributions

$$
\begin{aligned}
P\left(\boldsymbol{u}_x|\beta,\eta\right) &= \prod_x \mathcal{I}\left(e^{\boldsymbol{u}_x},\beta,\eta\right)\left|\frac{de^{\boldsymbol{u}_x}}{d\boldsymbol{u}_x}\right| && (3.25)\\
&\propto \exp\left(-\left(\boldsymbol{\beta}-1\right)^{\dagger}\boldsymbol{u}_x - \boldsymbol{\eta}^{\dagger}e^{\boldsymbol{u}_x}\right)\,, && (3.26)
\end{aligned}
$$

where $\boldsymbol{\beta} = \beta, \forall x$ and $\boldsymbol{\eta} = \eta, \forall x$ are the shape and scale parameters of the inverse gamma distribution $\mathcal{I}$.

In total the prior for the point-like flux becomes

$$
P(\boldsymbol{u}) = P\left(\boldsymbol{u}_y|\tau_{\mathcal{Y}(\boldsymbol{u})}\right)P\left(\tau_{\mathcal{Y}(\boldsymbol{u})}|\alpha_{\mathcal{Y}(\boldsymbol{u})},q_{\mathcal{Y}(\boldsymbol{u})},\sigma_{\mathcal{Y}(\boldsymbol{u})}\right)P\left(\boldsymbol{u}_x|\beta,\eta\right)\,, \qquad (3.27)
$$

### Modeling the background

Since the background $\boldsymbol{b}$ is also a strictly positive but unknown field we assume the prior $P(\boldsymbol{b})$ to have the same structure as the one for $\boldsymbol{s}$, except for the fact that it might be defined over different spaces. Hence we may again follow section 3.2.3 to build a hierarchical prior model for $P(\boldsymbol{b})$.

### Parameter model

Fig. 3.2 shows the complete hierarchical Bayesian network introduced in section 3.2. The goal is to model the data $\boldsymbol{d}$ by our quantities $\boldsymbol{s}$ and $\boldsymbol{u}$ and the background described by $\boldsymbol{b}$. The three logarithmic power spectra $\tau^{(\boldsymbol{s})} = (\tau^{\dagger}_{\mathcal{X}(\boldsymbol{s})},\tau^{\dagger}_{\mathcal{Y}(\boldsymbol{s})})^{\dagger}$, $\tau^{\dagger}_{\mathcal{Y}(\boldsymbol{u})}$, and $\tau^{(\boldsymbol{b})} = (\tau^{\dagger}_{\mathcal{X}(\boldsymbol{b})},\tau^{\dagger}_{\mathcal{Y}(\boldsymbol{b})})^{\dagger}$, can be reconstructed from the data. In case additional information sources are available to further constraint these parameters one may certainly use those and adjust

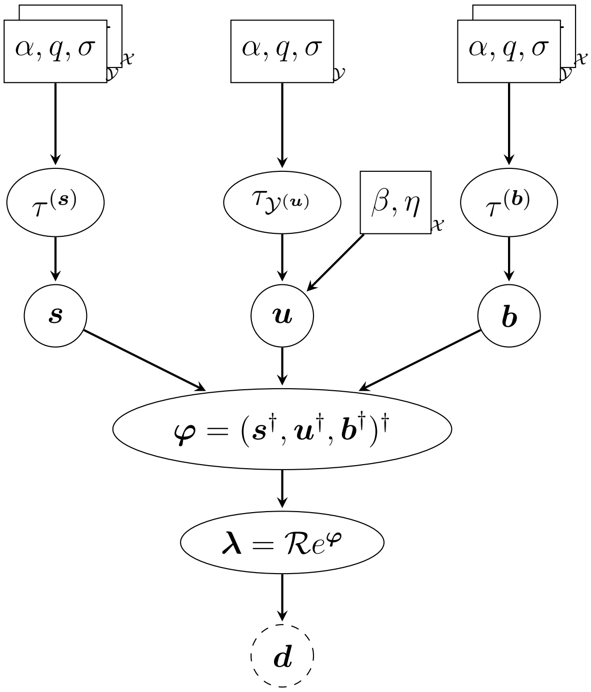


Figure 3.2.: Graphical model of the hierarchical Bayesian network introduced in section 3.2. Shown are the model parameters $\alpha, q, \sigma, \eta$ and $\beta$, in rectangular boxes, as they have to be specified by the user. The logarithmic spectral parameters $\tau^{(\boldsymbol{s})} = (\tau^\dagger_{\mathcal{X}(\boldsymbol{s})}, \tau^\dagger_{\mathcal{Y}(\boldsymbol{s})})^\dagger, \tau^\dagger_{\mathcal{Y}(\boldsymbol{u})}$ and $\tau^{(\boldsymbol{b})} = (\tau^\dagger_{\mathcal{X}(\boldsymbol{b})}, \tau^\dagger_{\mathcal{Y}(\boldsymbol{b})})^\dagger$, the diffuse signal field $\boldsymbol{\varphi}$, and the expected number of photons $\boldsymbol{\lambda}$, are inferred by the algorithm and shown in black solid circles. The observed photon count data $\boldsymbol{d}$, is marked by a dashed circle at the bottom.

the known logarithmic power spectra accordingly. However, in total the suggested algorithm is steered by the parameters $\alpha, q$, and $\sigma$, for which we can partly provide canonical values: $\alpha_{k,i} = \sigma_{k,i} = 1 \; \forall \, k = 0$, $\beta = 3/2$, as well as $\eta, q_i \gtrsim 0$ for $i \in \{\mathcal{X}^{(\boldsymbol{s})}, \mathcal{Y}^{(\boldsymbol{s})}, \mathcal{Y}^{(\boldsymbol{u})}, \mathcal{X}^{(\boldsymbol{b})}, \mathcal{Y}^{(\boldsymbol{b})}\}$.

## 3.3. The inference

The likelihood constructed in section 3.2 and the prior assumptions for the diffuse, point-like, and background signal field contain together all information available to tackle this inference problem. The resulting posterior is given by

$$\begin{aligned} P(\boldsymbol{\varphi}, \boldsymbol{\tau}|\boldsymbol{d}) = & \frac{P(\boldsymbol{d}|\boldsymbol{\varphi}, \boldsymbol{\tau})}{P(\boldsymbol{d})} \\ & \times \prod_{i \in \{\boldsymbol{s}, \boldsymbol{u}, \boldsymbol{b}\}} P\left(\tau_{\mathcal{Y}^{(i)}} | \alpha_{\mathcal{Y}^{(i)}}, q_{\mathcal{Y}^{(i)}}, \sigma_{\mathcal{Y}^{(i)}}\right) \\ & \times \prod_{i \in \{\boldsymbol{s}, \boldsymbol{b}\}} P\left(\tau_{\mathcal{X}^{(i)}} | \alpha_{\mathcal{X}^{(i)}}, q_{\mathcal{X}^{(i)}}, \sigma_{\mathcal{X}^{(i)}}\right) \\ & \times P\left(\boldsymbol{\varphi}_{\boldsymbol{u}_x} | \beta, \eta\right) \end{aligned} \tag{3.28}$$

where we have introduced $\boldsymbol{\tau} = \left(\tau_{\mathcal{X}^{(\boldsymbol{s})}}, \tau_{\mathcal{Y}^{(\boldsymbol{s})}}, \tau_{\mathcal{Y}^{(\boldsymbol{u})}}, \tau_{\mathcal{X}^{(\boldsymbol{b})}}, \tau_{\mathcal{Y}^{(\boldsymbol{b})}}\right)$.

In an ideal case with unlimited computational power, we would now calculate the mean and its variances according to eq. (3.2) and (3.3) for $\boldsymbol{\varphi}$, by integrating over all possible combinations of $\boldsymbol{\varphi}$ and $\boldsymbol{\tau}$. This would also provide us with all spectral parameters $\boldsymbol{\tau}$. But due to the complexity of the posterior probability distribution this is not worth pursuing.

We are relying on numerical approaches. Phase space sampling techniques like Markov chain Monte Carlo methods [106–110] are hardly applicable to our inference problem due to the extremely large phase space to be sampled over. Consequently we have to find suitable approximations to tackle the problem.

### 3.3.1. Maximum a posteriori

In case the posterior distribution is single peaked and symmetric, its maximum and mean coincide. At least in first order approximation this holds for eq. (3.28), which allows us to use the so called *maximum a posteriori* approach. This method can be enforced by introducing either a $\delta-$function at the posterior's mode,

$$\langle \boldsymbol{\varphi} \rangle_{(\boldsymbol{\varphi}|\boldsymbol{d})} \overset{\text{MAP-}\delta}{\approx} \int \mathcal{D}\boldsymbol{\varphi} \, \boldsymbol{\varphi} \, \delta(\boldsymbol{\varphi} - \boldsymbol{\varphi}_{\text{mode}}) \tag{3.29}$$

or by using a Laplace approximation, with its uncertainty covariance $D$, estimated from the curvature around the maximum. A field expectation value given by

$$\langle f\left(\boldsymbol{\varphi}\right)\rangle_{(\boldsymbol{\varphi}|\boldsymbol{d})} \overset{\text{MAP-}\mathcal{G}}{\approx} \int \mathcal{D}\boldsymbol{\varphi}\, f\left(\boldsymbol{\varphi}\right)\, \mathcal{G}\left(\boldsymbol{\varphi}-\boldsymbol{\varphi}_{\text{mode}}, D\right)\,. \tag{3.30}$$

In either case we are required to find the mode which is the maximum of the posterior distribution (3.28). Rather than maximizing the full posterior, it is convenient to minimize the information Hamiltonian, defined by its negative logarithm

$$\begin{aligned} H(\boldsymbol{\varphi},\boldsymbol{\tau}|\boldsymbol{d}) &= -\log P(\boldsymbol{\varphi},\boldsymbol{\tau}|\boldsymbol{d}) \\ &= H_0 + \boldsymbol{1}^\dagger \boldsymbol{\mathcal{R}} e^{\boldsymbol{\varphi}} - \boldsymbol{d}^\dagger \log\left[\boldsymbol{\mathcal{R}} e^{\boldsymbol{\varphi}}\right] \\ &\quad + \frac{1}{2}\left[\log(\det\Phi) + \boldsymbol{\varphi}^\dagger \Phi^{-1}\boldsymbol{\varphi}\right] \\ &\quad + \sum_{i\in\mathcal{I}} (\boldsymbol{\alpha}_i - 1)^\dagger \tau_i + \boldsymbol{q}_i^\dagger e^{-\tau_i} + \frac{1}{2}\tau_i^\dagger \boldsymbol{T}_i \tau_i \\ &\quad + (\boldsymbol{\beta}-1)^\dagger\, \boldsymbol{\varphi}_{\boldsymbol{u}_x} + \boldsymbol{\eta}^\dagger e^{-\boldsymbol{\varphi}_{\boldsymbol{u}_x}} \end{aligned} \tag{3.31}$$

with $\Phi = \text{diag}(\mathcal{S},\mathcal{U},\mathcal{B})^{\text{T}}$ and $\mathcal{I} \in \{\mathcal{X}^{(\boldsymbol{s})},\mathcal{Y}^{(\boldsymbol{s})},\mathcal{Y}^{(\boldsymbol{u})},\mathcal{X}^{(\boldsymbol{b})},\mathcal{Y}^{(\boldsymbol{b})}\}$. We have absorbed all terms that are constant in $\boldsymbol{\varphi}$ and $\boldsymbol{\tau}$ into $H_0$. The MAP ansatz seeks for the minimum of eq. (3.31), which is equivalent to maximizing the posterior eq. (3.28). This minimum can be found by taking the first partial derivatives of eq. (3.31) with respect to all components of $\boldsymbol{\varphi}$ and $\boldsymbol{\tau}$, respectively and equalling them to zero. The resulting filtering formulas for the diffuse and point-like flux read as

$$\begin{aligned} \left.\frac{\partial H}{\partial \boldsymbol{\varphi}}\right|_{\text{min}} &= \left(1 - \frac{\boldsymbol{d}}{\boldsymbol{l}}\right)^\dagger \boldsymbol{R} * e^{\boldsymbol{\varphi}} + \Phi^{(*)-1}\boldsymbol{\varphi} \overset{!}{=} 0 \\ \left.\frac{\partial H}{\partial \boldsymbol{\varphi}_{\boldsymbol{u}_x}}\right|_{\text{min}} &= \left.\frac{\partial H}{\partial \boldsymbol{\varphi}}\right|_{\text{min}} + (\boldsymbol{\beta}-1) - \boldsymbol{\eta} * e^{-\boldsymbol{\varphi}_{\boldsymbol{u}_x}} \overset{!}{=} 0 \end{aligned} \tag{3.32}$$

with

$$
\begin{aligned}
\boldsymbol{l} &= \boldsymbol{\mathcal{R}} e^{\boldsymbol{\varphi}}\,, && (3.33)\\
\Phi^{(*)} &= \operatorname{diag}\left(\mathcal{S}^{(*)}, \mathcal{U}^{(*)}, \mathcal{B}^{(*)}\right)^{\mathrm{T}} && (3.34)\\
\mathcal{S}^{(*)} &= \sum_k e^{\tau^{(*)}_{\mathcal{Y}^{(\boldsymbol{s})}}{}^{(k)}} \mathbb{P}_k \sum_l e^{\tau^{(*)}_{\mathcal{X}^{(\boldsymbol{s})}}{}^{(l)}} \mathbb{P}_l\,, && (3.35)\\
\mathcal{U}^{(*)} &= \sum_k e^{\tau^{(*)}_{\mathcal{Y}^{(\boldsymbol{u})}}{}^{(k)}} \mathbb{P}_k\,, && (3.36)\\
\mathcal{B}^{(*)} &= \sum_k e^{\tau^{(*)}_{\mathcal{Y}^{(\boldsymbol{b})}}{}^{(k)}} \mathbb{P}_k \sum_k e^{\tau^{(*)}_{\mathcal{X}^{(\boldsymbol{b})}}{}^{(k)}} \mathbb{P}_k\,. && (3.37)
\end{aligned}
$$

By $*$ and $\frac{\boldsymbol{d}}{\boldsymbol{l}}$ we refer to component wise multiplication and division, respectively. The filtering formulas for the power spectra, which are also derived by taking the first partial derivatives with respect to the components of $\boldsymbol{\tau}$ read as

$$
e^{\tau_{\mathcal{Y}^{(\boldsymbol{i})}}} = \frac{q_{\mathcal{Y}^{(\boldsymbol{i})}} + \frac{1}{2}\left(\operatorname{Tr}\left[\boldsymbol{i}\boldsymbol{i}^{\dagger}\mathbb{P}^{\dagger}\mathcal{X}^{-1}\right]\right)}{\alpha_{\mathcal{Y}^{(\boldsymbol{i})}} - 1 + \frac{1}{2}\left(\operatorname{Tr}\left[\mathbb{P}\mathbb{P}^{\dagger}\mathcal{X}\right]\right)_k + \boldsymbol{T}_{\mathcal{Y}^{(i)}}\tau_{\mathcal{Y}^{(\boldsymbol{i})}}}\,, \tag{3.38}
$$

and

$$
e^{\tau_{\mathcal{X}^{(\boldsymbol{j})}}} = \frac{q_{\mathcal{X}^{(\boldsymbol{j})}} + \frac{1}{2}\left(\operatorname{Tr}\left[\boldsymbol{j}\boldsymbol{j}^{\dagger}\mathbb{P}^{\dagger}\mathcal{Y}^{-1}\right]\right)}{\alpha_{\mathcal{X}^{(\boldsymbol{j})}} - 1 + \frac{1}{2}\left(\operatorname{Tr}\left[\mathbb{P}\mathbb{P}^{\dagger}\mathcal{Y}\right]\right)_l + \boldsymbol{T}_{\mathcal{X}^{(j)}}\tau_{\mathcal{X}^{(\boldsymbol{j})}}}\,, \tag{3.39}
$$

with $i \in \{\boldsymbol{s}, \boldsymbol{u}, \boldsymbol{b}\}$ and $j \in \{\boldsymbol{s}, \boldsymbol{b}\}$. These filtering formulas for the spectral parameters $\boldsymbol{\tau}$ are in accordance with [56, 57]. Unfortunately the eqs. (3.32), (3.38) and (3.39) lead to eight implicit equations rather than one explicit. Hence these equations need to be solved by an iterative minimization of eq. (3.31) using a minimisation algorithm such as steepest descent. The second derivative of the Hamiltonian, i.e. the Hessian around the minimum may serve as a first order approximation of the uncertainty covariance,

$$
\left.\frac{\partial^2 H}{\partial\boldsymbol{\varphi}\partial\boldsymbol{\varphi}^{\dagger}}\right|_{\min} \approx D^{(\boldsymbol{\varphi})-1}\,. \tag{3.40}
$$

A detailed derivation and closed form of $D^{(\boldsymbol{\varphi})-1}$ may be found in appendix A.1.

It has been shown that MAP-estimating of a field and its power spectrum simultaneously is suboptimal [56]. The reason is that the joinet posterior is far from being symmetric and exhibits long tails in directions correlated in the field and in its spectrum. The resulting scheme can strongly underestimate the field variance in low signal-to-noise situations. To overcome this difficulty, a number of approaches have been pursued like renormalization techniques. They lead to improved schemes which are closely related to each other. The most transparent of these approaches is the mean field approximation that involves the construction and minimization of an action, the Gibbs free energy [58]. A detailed derivation this approach can be found in appendix A.2.

The filtering formulas in a Gibbs free energy approach for $\boldsymbol{\varphi}$ only change within the exponent, while the one for $\boldsymbol{\tau}$ yields

$$e^{\tau_{\mathcal{Y}^{(i)}}} = \frac{q_{\mathcal{Y}^{(i)}} + \frac{1}{2}\left(\mathrm{Tr}\left[\left(\boldsymbol{i}\boldsymbol{i}^{\dagger} + D^{(j)}\right)\mathbb{P}^{\dagger}\boldsymbol{\mathcal{X}}^{-1}\right]\right)}{\alpha_{\mathcal{Y}^{(i)}} - 1 + \frac{1}{2}\left(\mathrm{Tr}\left[\mathbb{P}\mathbb{P}^{\dagger}\mathcal{X}\right]\right)_k + \boldsymbol{T}_{\mathcal{Y}^{(i)}}\boldsymbol{t}_{\mathcal{Y}^{(i)}}}\,, \tag{3.41}$$

and

$$e^{\tau_{\mathcal{X}^{(j)}}} = \frac{q_{\mathcal{X}^{(j)}} + \frac{1}{2}\left(\mathrm{Tr}\left[\left(\boldsymbol{j}\boldsymbol{j}^{\dagger} + D^{(j)}\right)\mathbb{P}^{\dagger}\mathcal{Y}^{-1}\right]\right)}{\alpha_{\mathcal{X}^{(j)}} - 1 + \frac{1}{2}\left(\mathrm{Tr}\left[\mathbb{P}\mathbb{P}^{\dagger}\mathcal{Y}\right]\right)_l + \boldsymbol{T}_{\mathcal{X}^{(j)}}\boldsymbol{t}_{\mathcal{X}^{(j)}}}\,, \tag{3.42}$$

with $i \in \{\boldsymbol{s}, \boldsymbol{u}, \boldsymbol{b}\}$ and $j \in \{\boldsymbol{s}, \boldsymbol{b}\}$. These formulas are in close accordance with the critical filtering technique [57]. Now they give a reconstruction of the spectral parameters of a field which does not necessarily need to be statistically isotropic and homogeneous over its combined domains. The appearing correction term $D$ in the trace term of the numerator of eqs. (3.41) and (3.42) compared to the MAP-solutions eqs. (3.38) and (3.39) is positive definite. Hence it introduces a positive contribution to the logarithmic power spectrum and therefore lowers a potential perception threshold [56].

## 3.3.2. The physical solution

All previously described methods only recover logarithmic fluxes, but the actual quantities of interest are the physical fluxes $\boldsymbol{\rho}$. Furtunately it is straight forward to calculate these given one selected approximate scheme,

$$\langle\boldsymbol{\rho}\rangle \overset{\text{MAP}-\delta}{\approx} \langle\boldsymbol{\rho}\rangle_{\delta} = \rho_0 e^{m_{\text{mode}}} \tag{3.43}$$

$$\overset{\text{MAP-G}}{\approx} \langle\boldsymbol{\rho}\rangle_{\mathcal{G}} = \rho_0 e^{m_{\text{mode}} + \frac{1}{2}D_{\text{mode}}}\,. \tag{3.44}$$

The uncertainty of the reconstructed fields may be approximated by

$$\sigma^2_{\mathcal{G}} = \left\langle\boldsymbol{\rho}^2\right\rangle_{\mathcal{G}} - \left\langle\boldsymbol{\rho}\right\rangle^2_{\mathcal{G}} \overset{\text{MAP-G}}{\approx} \left\langle\boldsymbol{\rho}\right\rangle^2_{\mathcal{G}}\left(e^{D} - 1\right)\,, \tag{3.45}$$

with its square root being the relative uncertainty.
The full Gibbs approach described in appendix A.2 would require to know $D^{(\varphi)}$ at all times during the minimisation of eq. (A.19), which is numerically not feasible, we only consider the MAP-$\mathcal{G}$ approach here to infer the signal fields. D$^3$PO has shown that such an approach does produce accurate signal field reconstructions. For the reconstruction of the spectral parameters $\boldsymbol{\tau}$,

we propose to use the full Gibbs approach as given by eqs. (3.41) and (3.42), as these take higher order correction terms into account. These have proven to lower a potential perception threshold, leading in total consequence to more detailed signal field reconstructions.
It must be noted that the mode approximation only holds for strictly convex problems and can perform poorly if this property does not hold. The precise form of the posterior is neither analytically nor numerically fully accessible due to the potentially extremely large phase space of the degrees of freedom. However in first order approximation the posterior Hamiltonian may be assumed to be convex close to its minimum.

## 3.4. The inference algorithm

To denoise, deconvolve, and decompose photon observations, while simultaneously learning the statistical properties of the fields, i.e. their power spectra, is a highly relevant but non-trivial task.
The derived information Hamiltonian eq. (3.31) and Gibbs free energy eq. (A.19) are scalar quantities defined over a potentially huge phase space of $\boldsymbol{\varphi}$ and $\boldsymbol{\tau}$. Even within an ideal measurement scenario the inference has to estimate three numbers plus the spectral parameters for each location and energy of the field from just one data value. Hence the inference can become highly degenerate if the data or priors do not sufficiently constrain the reconstruction. Such a scenario would likely lead to multiple local minima in a non-convex manifold of the landscape of the information Hamiltonian and Gibbs free energy, respectively. In total the complexity of the inference has its main roots at the non-linear coupling between the individual fields and spectral parameters to be inferred.
After numerous numerical tests we can propose an iterative optimisation scheme, which divides the global minimisation into multiple, more easily solvable subsets. By now, the following guide has given the best results:

1. Initialise the algorithm with naive starting values, i.e. $\boldsymbol{\varphi} = 0$ and $e^{\boldsymbol{\tau}_k} \propto k^{-4}$. If more profound knowledge is at hand you may certainly use this to construct a suitable initial field prior in order to speed up the inference and to overcome the outlined issues about the non-convexity of the minimisation.

2. Optimise the diffuse signal field, by minimising the information Hamiltonian eq. (3.31) or the Gibbs free energy eq. (A.19), respectively, by a method of your choice. As the gradients are always analytically accessible, we recommend using methods which make use of this, such as steepest descent.

3. Optimise the point-like and background signal field, accordingly to the diffuse signal field in the step before.

4. Update both spectral parameters of the diffuse flux field by again minimising eq. (3.31) and eq. (A.19), respectively, with respect to $\tau_{\mathcal{X}}^{(s)}$ and $\tau_{\mathcal{Y}}^{(s)}$. This optimisation step may be executed via a quasi Newton method.

5. Optimise the diffuse flux field, analog to step two.

6. Optimise the spectral parameters and maps of the point-like and background signal field according to step four and five, analog to the diffuse signal field.

7. Iterate between the steps four to six, until you have reached a global optimum. This may take a couple cycles. In order to get rough estimates of the signal in early cycles it is not necessary to let each minimisation run until it has reached its global desired convergence criteria. These may be retightened gradually as one gets closer to the global minimum.

8. At the reached minimum, calculate $D^{(\varphi)}$ in order to obtain the physical fluxes eq. (3.44).

It must be noted that the outlined iterative minimisation scheme has proven in multiple numerical tests to lead to the global minimum. However due to the extremely large phase space it is almost impossible to judge whether the algorithm has truly converged to the global optimum in case one works with real astrophysical data sets. Nevertheless this should not matter too much, in case the local minimum and the global optimum are close to each other and therefore do not differ substantially.

## 3.5. An inference example

In order to demonstrate the performance of the inference algorithm we apply the D$^4$P0 algorithm to a realistic but simulated astrophysical data set. In this mock example the algorithm is required to reconstruct the diffuse flux, the point flux, and the background flux. Additionally we request to infer all statistical properties of the diffuse flux, i.e. $\tau_{\mathcal{X}}^{(s)}$ and $\tau_{\mathcal{Y}}^{(s)}$, and $\tau_{\mathcal{Y}}^{(u)}$ of the point flux. The statistical properties of the background radiation are assumed to be known, otherwise the inference problem would be completely degenerate as no prior information would separate the background from the diffuse flux. The mock data set originates from a hypothetical observation with a field of view of $350 \times 350$ pixels and a resolution of $0.005 \times 0.01$ [a.u.]. All signal fields were drawn from Gaussian random fields with different

Table 3.1.: Parameters to define correlation structure of $\boldsymbol{s}$, $\boldsymbol{u}$, and $\boldsymbol{b}$

| | $e^{\tau_{\mathcal{X}}^{(\boldsymbol{s})}}$ | $e^{\tau_{\mathcal{Y}}^{(\boldsymbol{s})}}$ | $e^{\tau_{\mathcal{Y}}^{(\boldsymbol{u})}}$ | $e^{\tau_{\mathcal{X}}^{(\boldsymbol{b})}}$ | $e^{\tau_{\mathcal{Y}}^{(\boldsymbol{b})}}$ |
|---|---|---|---|---|---|
| $\theta$ | 4.5 | 0.3 | 0.1 | 0.75 | 2.0 |
| $\kappa$ | 2.5 | 0.2 | 2.5 | 0.005 | 0.014 |

correlation structures. The functional form of all correlation structures is

$$e^{\boldsymbol{\tau}(k)} \propto \frac{\theta^2 \kappa}{\left(1 + \left(\frac{2\pi k \kappa}{4}\right)^2\right)^2}, \tag{3.46}$$

but the correlation length $\kappa$ and the variance $\theta$ differ for each field and their sub-domains. The chosen parameters are given by table 3.1.

The assumed instrument's response incorporates a convolution with a Gaussian-like point spread function (PSF) with a FWHM of two times the pixelation size in each direction and an inhomogeneous exposure. The logarithmic exposure, the logarithmic PSF, the logarithmic photon counts, as well as the raw photon counts are shown in the top panel of fig. 3.3.

Further rows of fig. 3.3 show all signal fields in terms of logarithmic fluxes, i.e. $\boldsymbol{s}, \boldsymbol{u}$, and $\boldsymbol{b}$. For each field we show the ground truth, i.e. the drawn Gaussian random field, its reconstruction, the error between reconstruction and truth flux, i.e. $\left|\boldsymbol{\rho}^{\text{rec}} - \boldsymbol{\rho}^{\text{truth}}\right|$, as well as the uncertainty $\sigma_{\mathcal{G}}$ of the reconstruction provided by D$^4$PO, according to eq. (3.45). For the reconstruction we used the following parameter setup, $\alpha_{\text{i}} = 1, q_{\text{i}} = 10^{-12}, \sigma_{\text{i}} = 1, \text{i} \in \{\mathcal{X}^{(\boldsymbol{s})}, \mathcal{Y}^{(\boldsymbol{s})}, \mathcal{Y}^{(\boldsymbol{u})}\}$, $\beta = \frac{3}{2}$, and $\eta = 10^{-4}$ in a MAP-$\mathcal{G}$ approach as this has proven to give the best results within a reasonable amount of computing time [89].

Looking more closely at the diffuse flux field, the original and its reconstruction are in good agreement. The strongest deviation may be found in regions with low amplitudes, which is not surprising as we are using an exponential ansatz to enforce positivity for all our fields. Hence small errors in $\boldsymbol{s} \to (1 \pm \epsilon)\, \boldsymbol{s}$ factorise in the physical photon flux field, $\boldsymbol{\rho}^{(\boldsymbol{s})} \to e^{\boldsymbol{s}} e^{\pm\epsilon}$ that scales exponentially with the amplitude of the diffuse flux field. Further, in almost all regions the absolute error shows that the reconstruction is in very good agreement with the original one. Only in areas with a relatively weak point flux and a rather strong diffuse flux the decomposition seems to run into a fundamental problem, as the priors and the likelihood can no longer break the degeneracies between the different sources.

From fig. 3.4 it becomes apparent that the reconstructed power spectra of $\boldsymbol{s}$ track all large scale modes in good agreement up to $\nu \lesssim 20$. At higher harmonic modes the reconstructed power spectra start to deviate from the

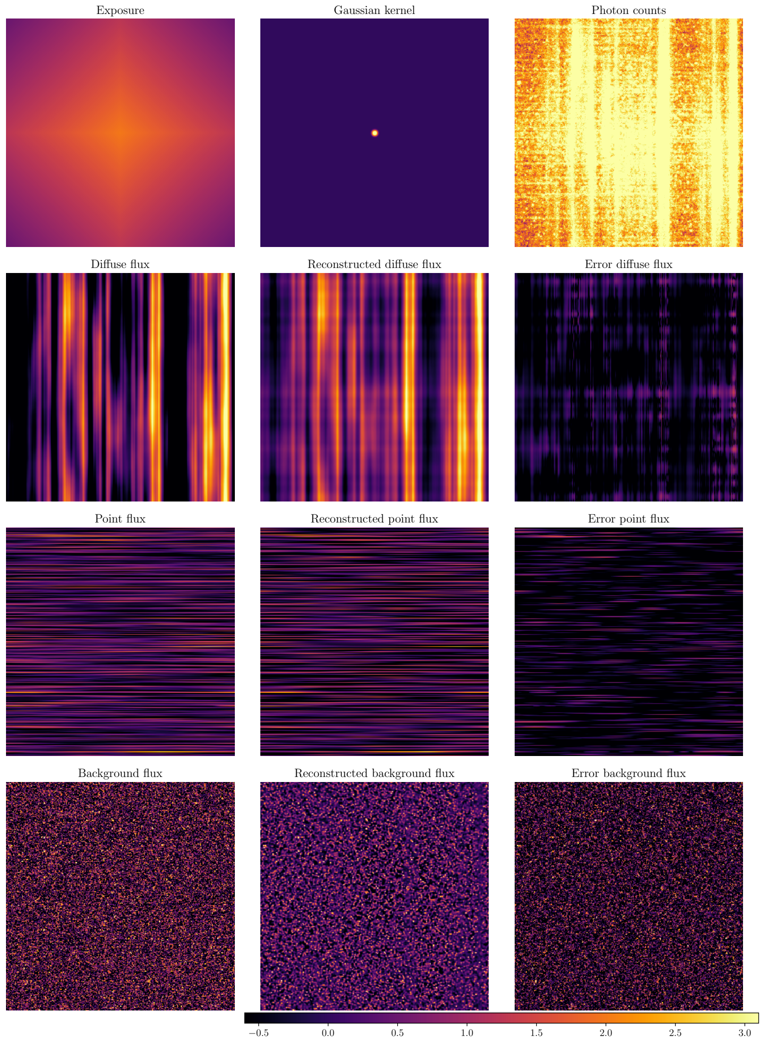
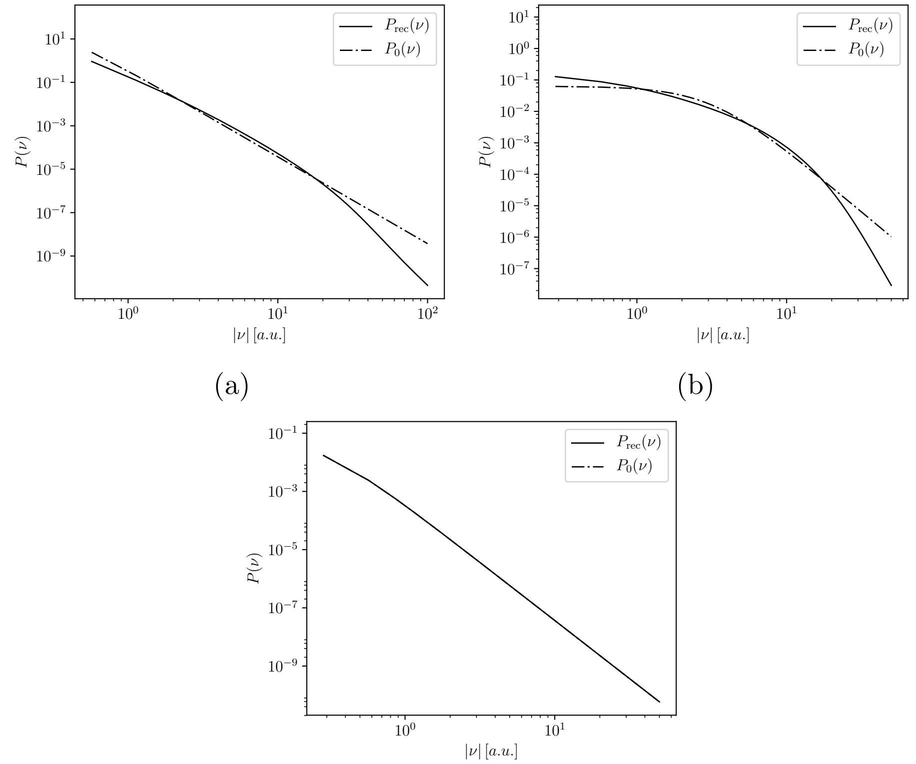


Figure 3.3.: Demonstration of the full capabilities of D$^4$PO based on a simulated but realistic data set gathered from a potential astrophysical high energy telescope. The spatial dimension is oriented vertically and the spectral/energy dimension horizontally. Point sources appear therefore as the horizontal lines. All fields are living over a regular grid of $350 \times 350$ pixels. The top panel shows the assumed instrument's exposure map, its Gaussian convolution kernel and the obtained data set, once on logarithmic scale and once the raw photon counts. The panels below display the diffuse, point-like, and background flux. For all signal fields we show the ground truth on the left hand side, followed by its reconstruction, the error and the uncertainty of the reconstruction. For illustration purposes all fluxes are on logarithmic scale and clipped between $-0.6$ and $3.1$, except the 'Raw photon counts' which are shown on their native scale.



(a) (b)

(c)

Figure 3.4.: Illustration of the reconstructed power spectrum of $\boldsymbol{s}$ in its spatial (fig. 3.4a) and spectral sub-domain (fig. 3.4b) and $\boldsymbol{u}$ in its spectral domain (fig. 3.4c). The dashed black line indicates the default spectrum from which the Gaussian random fields, shown in Figure 3.3, were drawn, while the solid black lines show its reconstruction. In case of $\boldsymbol{\tau}_{\mathcal{Y}(\boldsymbol{u})}$, both lines are in such close agreement that they are visually indistinguishable.

reference and fall more steeply. The drop off point at $\nu \approx 20$ roughly corresponds to the support of the PSF of the instrument's response. As the power spectrum still shows a smooth shape at $\nu \gtrsim 20$, the action of the smoothness enforcing prior starts to set in. Were $\sigma$ significantly smaller, the spectra would start to scatter wildly, which we do not expect in astrophysical spectra. Hence the smoothness enforcing prior allows some kind of superresolution up to a certain threshold.

Having a closer look at the logarithmic point like flux field (fig. 3.3), we observe a similar situation as for diffuse flux field. This is supported by the reconstructed power spectrum, $\boldsymbol{\tau}_{\mathcal{Y}(\boldsymbol{u})}$ and its original one which match perfectly (fig. 3.4c). Up to where the reconstruction is mainly driven by the data, may not be stated any more as the algorithm recovered all modes correctly. This is of course due to an appropriate setup of the smoothness enforcing prior.

Nevertheless it must be noted that $\sigma$ has to be set accurately as it can have significant influence on the reconstructed power spectrum. For a detailed discussion about its influence we refer to [96].

The results for the spatial reconstruction performance of the point-like sources are plotted in fig. 3.5. For all energies we show individually the match between original and reconstructed flux at all locations. As a point-source illuminates the sky over a broad range of energies at a fixed location these serpentine pattern appear in fig. 3.5. In total they are in good agreement within the $2\sigma$ confidence interval. This confidence interval corresponds to a diffuse and background free data set, it only illustrates the expected photon shot noise of point sources. The higher flux point sources tend to be reconstructed more accurately as a better SNR allows a sharper decomposition of the different sources. As a natural consequence the accuracies of the reconstruction becomes worse for regions with low count rates as the SNR becomes severe. The calculated absolute error supports these findings.

As the correlation length of the background flux is similar to the support of the assumed PSF of the instrument's response, not all of its small features could be reconstructed even though its statistical properties, i.e. the power spectrum, were provided and not inferred from the data. Hence the reconstruction is smeared out as one would expect. Therefore the calculated absolute error is more finely grained and its absolute magnitude is smaller compared to the reconstruction.

The calculated uncertainty estimates of the reconstructions, $\boldsymbol{s}$, $\boldsymbol{u}$, and $\boldsymbol{b}$ are as one would expect them to be, given the Poissonian shot noise and the Gaussian convolution of the response operator. The absolute magnitudes

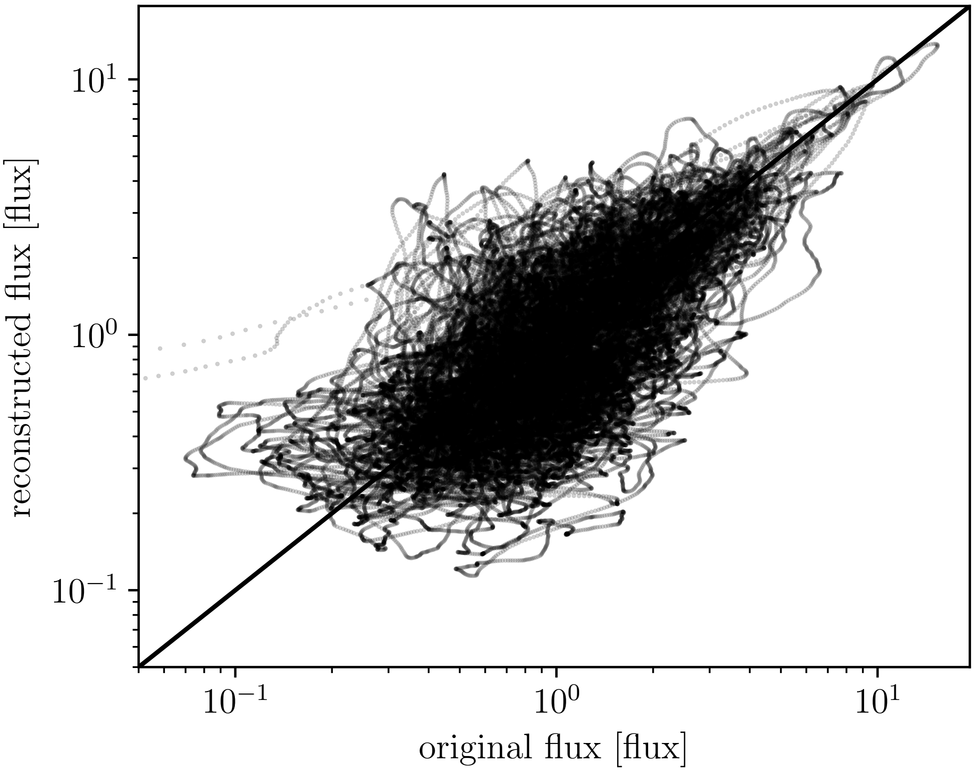


Figure 3.5.: Reconstructed versus original physical point like flux in its spatial domain for all energies. The grey contour indicates the $2\sigma$ confidence interval of the Poissonian shot noise.

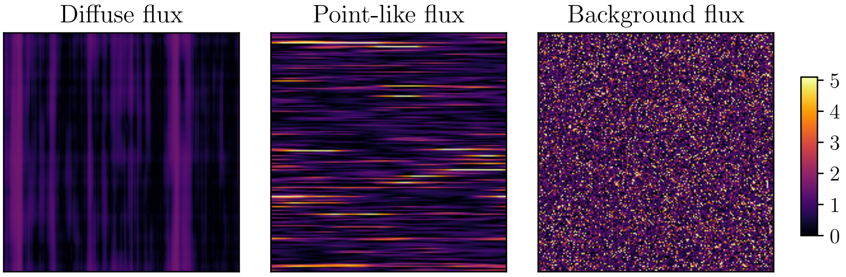


Figure 3.6.: Illustrations show the differences between $\left|\boldsymbol{\rho}^{\text{error}} - \boldsymbol{\rho}^{\text{uncertainty}}\right|$ on their native scale, by devision through the uncertainty estimates. The image-mean of the normalised errors is $\approx 0.6$ for the diffuse flux, $\approx 0.85$ for the point-like flux, and $\approx 1.5$ for the background flux. Hence they show, that some uncertainties are overestimated and others underestimated, respectively, due to the used Laplace approximation.

of the uncertainties are in all cases smaller than the amplitude of the reconstruction itself. Figure 3.6 shows that uncertainty estimates are useful, however one has to keep in mind that approximation to the non-Gaussian posterior were done. Consequently for heavy tails of the posterior PDF the approximated uncertainty do not account for.

The inference algorithm is steered by the model parameters, shown in rectangular boxes in Figure 3.2. In this mock data demonstration run we set them all to physically motivated values. Changing these, especially $\sigma$ of the smoothness enforcing prior and $\beta$, the shape parameter of the spatial prior for the point-like flux can have a significant influence on the reconstruction. A detailed parameter study of $\alpha, q, \beta$, and $\eta$ can be found in [89]. It may be summarised that the advancements of D$^4$PO to denoise, deconvolve, and decompose multidimensional photon observations into multiple morphologically different sources, such as diffuse, point-like and background sources, while simultaneously learning their statistical properties in each of the field domains independently, have shown to give reliable estimates about the physical fluxes for data of sufficient qualtity.

## 3.6. Conclusion

We derived the D$^4$PO algorithm, to denoise, deconvolve and decompose multidimensional photon observations, into multiple morphologically different sources. In this context we focused on the decomposition of astrophysical high energy photon count data into three different types of sources, diffuse, point-like and background radiation. Each of these components lives over a continuous space over multiple domains, such as energy and location. In addition to the simultaneous reconstruction of all components the algorithm can infer the correlation structure of each field over each of its sub-domains. Thereby D$^4$PO takes accurate care of the instruments response function and the induced photon shot noise. Finally the algorithm can provide a posteriori uncertainty estimates of the reconstructed fields.

The introduced algorithm is based on D$^3$PO [89], which was successfully applied to the FERMI LAT data [90] and giant magnetar flare observations [96]. Here we show an advancement towards multidimensional fields and how to incorporate further components. The D$^4$PO algorithm is based on a hierarchical Bayesian parameter model within the framework of IFT [32]. The model incorporates multiple a priori assumptions for the signal fields of interest. These assumptions account for the a priori known statistical properties of the fields in order to decompose the data set into their different sources. As some of these statistical identities are often not known a prior, the algorithm can learn them from the data set itself. Thereby we assume that each field follows a log-normal distribution over each of its sub-domains, except for the spatial correlations of point-sources. The correlations over these sub-domains may then be encoded via a power spectrum, implicitly assuming statistical homogeneity and isotropy over each sub-domain. As the point-like flux is implicitly defined to be statistically independent in its spatial sub-domain, we motivated an independent Inverse-Gamma prior, which implies a power law behaviour of the amplitudes of the flux.

To denoise and deconvolve astrophysical counts we took detailed care of an adequate likelihood modelling. The derived likelihood is a Poisson distribution to denoise the photon shot noise and incorporates all instruments artefacts, which are imprinted in the data set.

In total the hierarchical Bayesian parameter model is steered by only five parameters, for which we can provide well motivated a priori values. None of these parameters drives the inference dominantly as they may all be set to values where they provide minimal additional information to the inference problem.

In a simulated high energy photon count data set we demonstrated the performance of D$^4$PO. The algorithm successfully denoised, deconvolved, and decomposed the raw photon counts into diffuse, point-like and background radiation. Simultaneously it recovered the power spectrum of the diffuse flux in its spatial and spectral sub-domain. The correlation structure in the spectral sub-domain of the point-like flux was also inferred from the data. In total the analysis yielded detailed reconstructions and uncertainty estimates which are in good agreement with the simulated input.

The introduced algorithm is applicable to a wide range of inference problems. Its main advancement to reconstruct fields over multiple manifolds, each with different statistical identities, should find use in various data inference problems. Besides the most obvious applications in high energy astrophysics, such as FERMI, XMM, Chandra, Comptel, etc. to reconstruct the diffuse and point-like flux dependent on energy and location on the sphere, one may also infer energy dependent light curves of gamma ray bursts. Hence D$^4$PO has a broad range of applications.

# 4. Search for quasi-periodic signals in magnetar giant flares

*This chapter, as well as appendix B, are used additionally as a journal publication in Astronomy & Astrophysics (Pumpe* et al.*, [111]).*
*I am the principal investigator of the research described in the following chapter. My contributions include the application of the developed algorithm D$^4$PO on the data set provided by Michael Gabler (MG). I wrote the main parts of the chapter, while MG wrote approximately 20% of sections 4.1 and 4.4 about the detailed structure of magnetars. Theo Steininger played a pivotal role by continuously debugging NIFTy, the software package on which D$^4$PO is primarily based on. Torsten Enßlin also fulfilled the role of a principal investigator as he is my PhD supervisor. All authors read, commented, and approved the final manuscript.*

## 4.1. Introduction

The discovery of quasi-periodic oscillations (QPOs) in the giant flare of the magnetar SGR 1806-20 by [29] may have been the first detection of neutron star oscillations and triggered a wealth of theoretical work explaining the reported frequencies. The giant flare was likely caused by a large-scale reconnection or an interchange instability of the magnetic field [112]. Large amounts of energy are released as an expanding $e^{\pm}$-pair plasma, observable as the initial spike of the giant flare. Parts of this plasma are trapped by the ultra-strong magnetic field and form a so-called trapped fireball [112], which then slowly evaporates on a timescale of up to a few 100 seconds. The QPOs were detected in this decaying tail of the giant flare. Other groups have not only confirmed this detection, they even found additional oscillation frequencies in different magnetars: 18, 26, 29, 92, 150 , 625, and 1840 Hz in the giant flare of SGR 1806-20, and 28, 53, 84, and 155 Hz in the giant flare of SGR 1900+14 [113–116]. With different methods, more oscillation frequencies were found in the giant flare of SGR 1806-20 by Hambaryan *et al.* [117], 17, 21, 36, 59, and 116 Hz. The number of giant flares at the time of writing is limited to three events. The more frequent but less energetic bursts of several magnetars have therefore also been investigated for frequencies, and some candidates were found: 57 Hz in SGR 1806-20 [118], and at 93, 127, and 260 Hz in SGR J1550-5418 [119].

It was soon realized that these frequencies are probably related to oscillations of the neutron star, and several groups tried to identify them as elastic oscillations of the crust [113, 120–129], Alfvén oscillations [130–132], or coupled magneto-elastic oscillations [133–140]. The theoretical models based on the observed frequencies are very elaborate and may be able to constrain properties of high-density matter as found in the interior of neutron stars. Some of the models, for instance, require a superfluid component in the core of the star [139–146]. Different models depend sensitively on the identification of the fundamental oscillation frequency, and may not explain all of the observed frequencies. Even when the fundamental frequency is identified, the interpretation and parameter estimation is not yet straightforward because of degeneracies in the parameter space. However, keeping other stellar parameters fixed, some general trends of the fundamental oscillation frequency can be summarized as follows [see 145, 146, for a detailed discussion]: i) The frequency scales linearly with the magnetic field strength. ii) It decreases with increasing compactness [131]. The compactness is related to the hardness of the equation of state (EOS): Material with a stiff equation of state is harder to compress, leading to larger radii and hence lower compactnesses. iii) It can only reach the surface for significantly strong magnetic fields $\bar{B} \gtrsim \bar{B}_{\mathrm{outbreak}}(\sqrt{c_s})$, whose thresholds depend on the square root of the shear velocity [146].

It is of great importance to understand which of the frequencies are in the signal. In previous attempts at identifying possible frequencies in the light curve, the statistical noise of the detectors was not modelled consistently. To improve on this, we employ a Bayesian method that properly takes the photon shot noise as the largest contributor into account.

In this work, we re-analyse the data for the two giant flares of SGR1806-20 and SGR1900+14, which were obtained with the proportional counter array (PCA) of the Rossi X-ray Timing Explorer (RXTE). The data are available online at the High Energy Astrophysics Science Archive Research Center (HEASARC). In section 4.2 we briefly discuss our numerical approach to estimating the influence of the noise and to reconstructing the likely signal with a reduced contribution from the noise. The next section is devoted to the reconstruction and investigation of the light curves of the two giant flares of SGR 1806-20 and SGR 1900+14, which are then discussed in section 4.4. We conclude our analysis in section 4.5.

## 4.2. Inference of photon observations

Our goal is to reconstruct the light curve and possible frequencies of QPOs in giant X-ray flares for the neutron stars SGR 1806-20 and SGR 1900+14 using the data taken by RXTE. Owing to experimental constraints such as limited data storage, finite time resolution of each photon count, finite detector size, and sensitivity, RXTE cannot record all physical relevant and available information of a continuous photon flux. Most importantly, the recorded photon counts contain significant photon shot noise. Hence, we have to use probabilisitic data analysis methods to obtain an estimate of the time-dependent, continuous photon flux $\boldsymbol{\phi}(t)$ , including its uncertainties. Since the flux varies on a logarithmic scale, we reconstructed the logarithm of the flux $s(t) = \log(\boldsymbol{\phi}(t)/\phi_0)$ and its temporal power spectrum directly from the data.

In this context, it is natural to build the inference upon the Bayes theorem, allowing us to investigate the a posteriori probability distribution $P(\boldsymbol{\phi}|\boldsymbol{d})$, stating how likely a given photon flux field $\boldsymbol{\phi}$ is given the data set $\boldsymbol{d}$ . For the two data sets of SGR 1806-20 and SGR 1900-14, we assumed the data $\boldsymbol{d}$ to be the result of a Poisson process whose expectation value $\boldsymbol{\lambda}$ is given by the photon flux $\boldsymbol{\phi}(t)$ of the photon burst convolved with the response operator $\boldsymbol{R}$,

$$\boldsymbol{\lambda} = \boldsymbol{R}\boldsymbol{\phi}(t) = \boldsymbol{R}\phi_0 e^{s(t)}\,. \tag{4.1}$$

The response operator encodes all instrument specifications and states how $\boldsymbol{\phi}$ imprints itself on $\boldsymbol{\lambda}$. Here, we only considered the readout deadtime periods of the instrument and neglected all other instrument responses. In eq. (4.1) the constant $\phi_0$ absorbs numerical constants and the physical units of the time dependent log- flux signal field $s(t) = \boldsymbol{s}$. As the signal describes the logarithmic flux, we naturally ensured the positive definiteness of the photon flux. Since we analysed QPOs, it is expected that $\boldsymbol{\phi}$ exhibits unknown but spatial correlations. Hence, we did not enforce any concrete spatial correlations. We only assumed $\boldsymbol{\phi}$ to follow a multivariate log-normal statistic, with an a priori unknown covariance. Assuming stationary statistics, the underlying covariance is fully determined by a power spectrum $P_s(\nu)$, described by $P_s(\nu) = P_0 e^{\boldsymbol{\tau}(\nu)}$, to ensure positivity of the power spectrum. We inferred this power spectrum as well as $\boldsymbol{\phi}$ from the data themselves by setting up a hierarchical prior model [33, 56, 90, 102, 147]. Hence the posterior of our Bayesian inference is given by

$$P(\boldsymbol{\phi}, \boldsymbol{\tau}|\boldsymbol{d}) \propto \frac{P(,\boldsymbol{d}|\boldsymbol{\phi}, \boldsymbol{\tau})\, P(\boldsymbol{\tau}|\sigma_{\mathrm{sm}})}{P(\boldsymbol{d})}\,, \tag{4.2}$$

where $P(\boldsymbol{d}|\boldsymbol{\phi}, \boldsymbol{\tau})$ is the likelihood describing how a potential photon flux $\boldsymbol{\phi}$ including a certain covariance structure described by $\boldsymbol{\tau}$ imprints itself in a potential data set. In addition to strong spectral peaks in the power spectrum induced by the almost discrete frequencies of QPOs, we enforce some kind of spectral smoothness on the logarithmic scale, favouring power-law spectra. This behaviour is enforced and tuned by the prior term $P\left(\boldsymbol{\tau}|\sigma_{\mathrm{sm}}\right)$ and its parameter $\sigma_{\mathrm{sm}}$ ([57]). If $\sigma_{\mathrm{sm}} \to 0$, infinite spectral smoothness is enforced, while $\sigma_{\mathrm{sm}} \to \infty$ does not enforce any spectral smoothness. As we wish to be sensitive to spectral lines in the power spectrum, the influence of $\sigma_{\mathrm{sm}}$ on the inference is discussed in greater detail in appendix B.1.

For the detailed mathematical rigorous derivation and discussion of the used D$^3$PO- algorithm, we refer to [89] and [147]. The D$^3$PO- algorithm was successfully applied on the Fermi LAT data [90]. In order to handle high power spectrum resolutions as is needed to infer spectral lines in the power spectrum, we reimplemented the algorithm in NIFTY 3 [148]. The capabilities of the algorithm to analyse QPOs are demonstrated in the appendices B.1 to B.3.

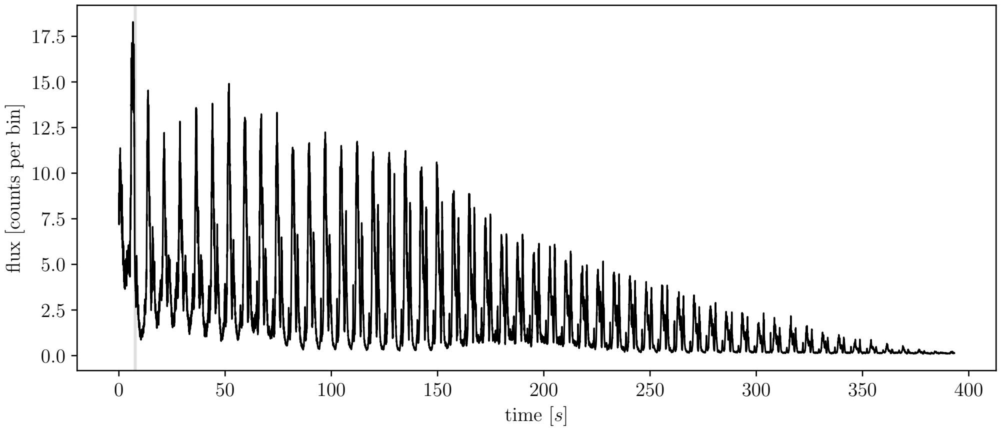
## 4.3. Results

### 4.3.1. SGR 1806-20

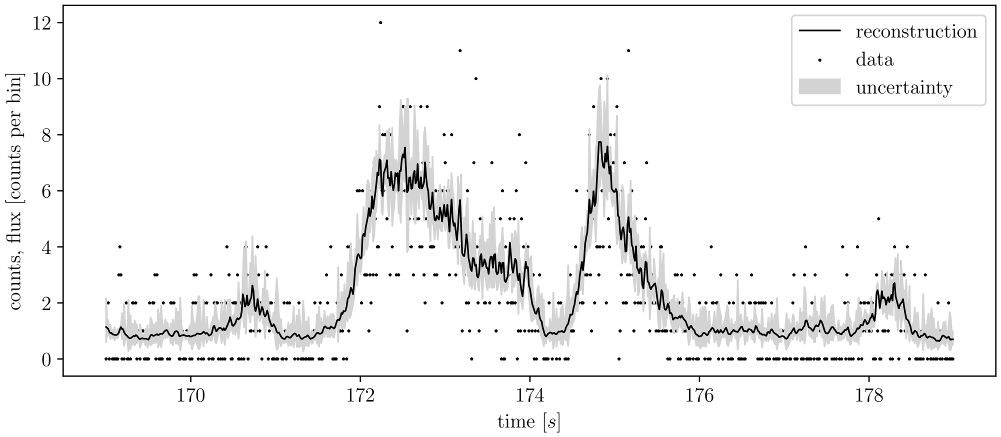
We re-analysed the archival RXTE data of the giant flare of SGR 1806-20, which occurred on 2004 December 27. Owing to the high photon flux, the instrument telemetry was saturated, causing several data gaps in the first seconds of the flare. We therefore neglected the very beginning of the flare in the current analysis and started our investigation roughly 4.5 s after the initial rise, immediately after the third deadtime interval. However, we accounted for the only remaining operational down time of the instrument in our data between $7.3865\,\mathrm{s} \leq \mathrm{t} \leq 7.9975\,\mathrm{s}$. For the analysis, we binned the data into pixels with a volume of 1/800 s. To overcome the periodic boundary conditions introduced by the fast-Fourier transformation, which we used to switch between signal space and its harmonic space, we performed the signal inference on a regular grid with $2^{19}$ pixel, each also with a volume of 1/800 s, by adding sufficient buffer time.

In fig. 4.1 we plot the inferred $\boldsymbol{\phi}$ for the entire duration of the giant flare and for a selected period of time for a smoothness parameter $\sigma_{\mathrm{sm}} = 5 \times 10^5$. Our algorithm is able to significantly reduce the scatter of the light curve that is caused by the photon shot noise. The thus reconstructed light curves can now be further analysed for potential periodic signals.

fig. 4.2 shows the reconstructed profiles of one pulse rotation period. There, the mean pulse profile is given as a thick black line, and all individual pulses are plotted in different colours. Dark blue are the first pulses that have a significantly lower second maxima around 5.5 s than the mean pulse. The major maximum (time $\sim 3$ s) is very close to the mean maximum. At intermediate times (green lines), both maxima of the pulse take their maximum values, roughly 40% and 130% more than at the beginning for the first and second maximum, respectively. At late times (red lines)



(a)



(b)

Figure 4.1.: Reconstructed light curve of the giant flare of SGR 1806-20 using a smoothness-enforcing prior, with $\sigma_{\mathrm{sm}} = 5 \times 10^5$ is shown in fig. 4.1a. The grey narrow rectangle indicates the operational down time of the instrument. For better visibility, we plot the light curve between $\approx 164\,\mathrm{s}$ and $\approx 174\,\mathrm{s}$ in fig. 4.1b. In addition to the raw photon counts (black dots), the black line indicates the reconstruction of the expected photon counts, i.e. $\lambda$ as well as its one-$\sigma$ confidence interval. Owing to the high resolution of the photon flux, only every fourth point of the regular grid is plotted. In both plots each pixel has a duration of $1/800\,\mathrm{s}$ .

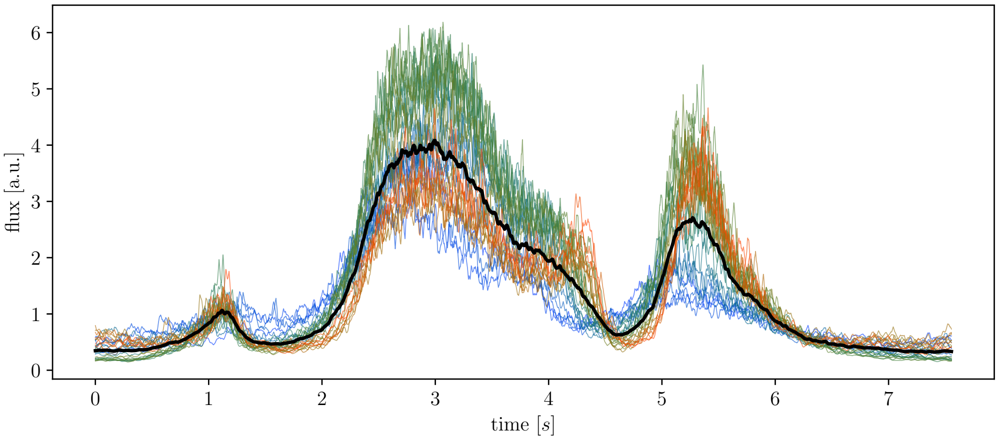
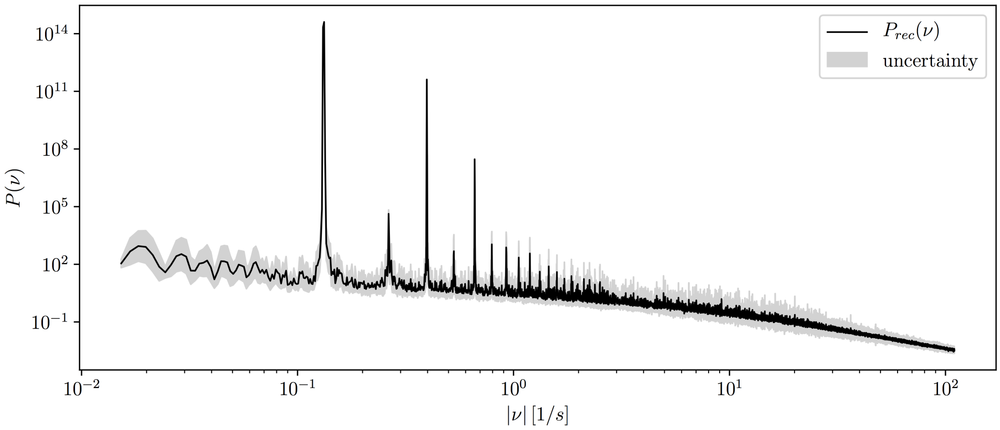


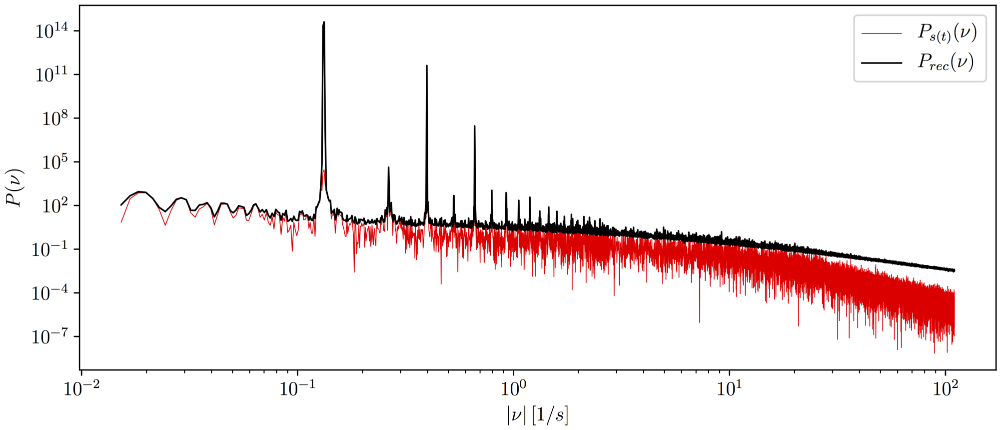
Figure 4.2.: Pulse profiles of different rotation periods overplotted as given in fig. 4.1a. The temporal evolution of the pulse profile is visible from blue to green to red. The mean pulse is shown as a thick black line. All pulses are shifted such that their log-means are zero.

the main peak declines by 40%, while the second maximum almost stays constant and has an amplitude similar to the main peak. This behaviour indicates a complex evolution of the fireball, or of the fireballs, if there are more than one.

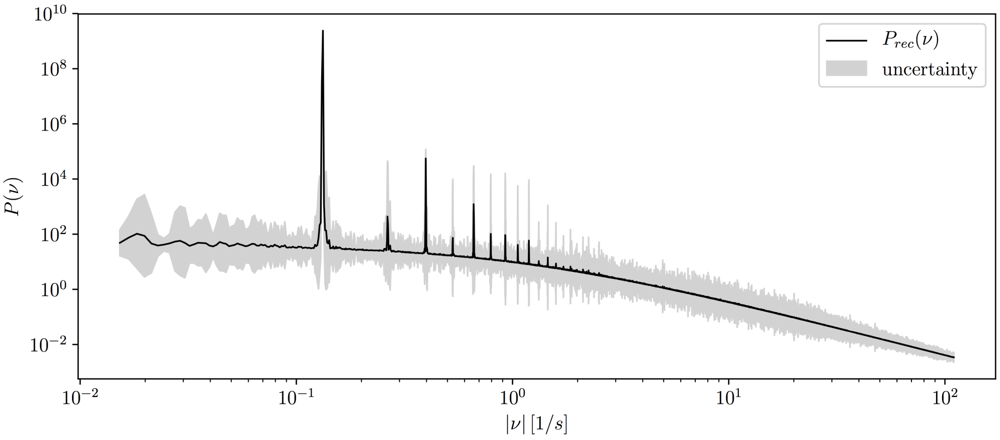
The power spectrum of the entire flare is plotted in fig. 4.3a. The rotation period of the magnetar is recovered with the frequency of the first main peak $\nu_0 = 0.1323\,\mathrm{Hz}$, which is very close to $\nu_0 = 0.13249\,\mathrm{Hz}$, as given in [149]. We are able to find up to the 31st overtone of this frequency at $\nu = 4.245\,\mathrm{Hz}$. In fig. 4.3b we show the reconstructed power spectrum $P_{\mathrm{rec}}$ from fig. 4.3a in black together with the power spectrum obtained from the reconstructed light curve (fig. 4.1a) $P_{s(t)}(\nu)$ in red. At low frequencies, that is, $\nu \lesssim 3\,[\mathrm{Hz}]$, the two spectra are in good agreement, as the reconstruction is well constrained by the data on large scales. Between $3\,[\mathrm{Hz}] \lesssim \nu \lesssim 20\,[\mathrm{Hz}]$, the inference algorithm enters the regime of a lower signal-to-noise ratio (S/N), which in principle leads to noisier $P_{\mathrm{rec}}$. However, this natural behaviour is counteracted by the smoothness-enforcing prior. At higher $\nu \gtrsim 20$ Hz, the shapes of both spectra start to deviate significantly. The reason for this is that in the noise-dominated frequency regime, D$^3$PO filters out the photon shot noise. From a naive perspective, small-scale features in the signal therefore need to be significantly strong in order to be detectable after a pure photon shot noise filtering operation on the data set. However, D$^3$PO accounts for the power loss of this filtering when it reconstructs the power spectrum from the data themselves. Thus, fig. 4.3b indicates that above 20 Hz the data are noise dominated, and spectral features there have to


$P_{rec}(\nu)$
uncertainty
$10^{14}$
$10^{11}$
$10^{8}$
$10^{5}$
$10^{2}$
$10^{-1}$
$P(\nu)$
$10^{-2}$
$10^{-1}$
$10^{0}$
$10^{1}$
$10^{2}$
$|\nu|\ [1/s]$

(a)


$P_{s(t)}(\nu)$
$P_{rec}(\nu)$
$10^{14}$
$10^{11}$
$10^{8}$
$10^{5}$
$10^{2}$
$10^{-1}$
$10^{-4}$
$10^{-7}$
$P(\nu)$
$10^{-2}$
$10^{-1}$
$10^{0}$
$10^{1}$
$10^{2}$
$|\nu|\ [1/s]$

(b)



(c)

Figure 4.3.: Reconstructed power spectra of the giant flare of SGR 1806-20: For fig. 4.3a and fig. 4.3c we used smoothness-enforcing priors with $\sigma_{\mathrm{sm}} = 5 \times 10^5$ and $\sigma_{\mathrm{sm}} = 10^5$, respectively. The uncertainty intervals are given as grey shaded areas. In fig. 4.3b we show the reconstructed power spectrum again from the data themselves as in fig. 4.3a, along with the power spectrum of the logarithmic reconstructed light curve of fig. 4.1a. Because of the high resolution of the reconstructed power spectra, only every fourth point of the regular grid is plotted. Each pixel has a volume of 1/655 Hz.

Table 4.1.: All frequencies above 3.5 Hz with $\chi_0 > 11$ and their multiplicity $n$ of the rotation period $\nu_0 = 0.13249$ [Hz] for SGR 1806-20.

| $\nu$ [Hz] | $\chi_0$ | $n$ [$\nu_0$] | $\nu$ [Hz] | $\chi_0$ | $n$ [$\nu_0$] |
|---|---|---|---|---|---|
| 3.86 | 11.071 | 29.134 | 11.171 | 11.762 | 84.316 |
| 4.602 | 11.141 | 34.735 | 15.768 | 11.999 | 119.013 |
| 4.95 | 13.592 | 37.361 | 16.272 | 12.772 | 122.817 |
| 6.81 | 11.212 | 51.4 | 19.034 | 12.632 | 143.664 |
| 9.187 | 18.293 | 69.341 | | | |

be very strong to be recognisable. To test the dependence of our method on the chosen smoothness prior $\sigma_{\rm sm}$ as discussed also in appendix B.1, we additionally calculated the reconstructed light curve and its corresponding power spectrum for $\sigma_{\rm sm} = 10^5$. The latter is given in fig. 4.3c. Obviously, a smaller $\sigma_{\rm sm}$ leads to a smoothing of the spectrum and the algorithm suppresses the detection of periodic signals at higher frequencies. We found $\sigma_{\rm sm} = 5 \times 10^5$ to be the optimal value to still observe power in the Fourier transform at higher frequencies. For higher values of $\sigma_{\rm sm}$ , we qualitatively obtain similar but more noisy results for the reconstructed light curve and the corresponding power spectra.

In addition to the obvious peaks that are related to the rotation period and the corresponding overtones, there are still other features in the reconstructed power spectrum of fig. 4.3a that seem to have higher powers than the noise. To estimate the significance of these spectral peaks, we calculated a residual $\chi$ between the inferred log-spectrum $\tau$ and its local median $\bar{\tau}$ weighted with the local variance $\sigma$,

$$\chi = \frac{\tau - \bar{\tau}}{\sigma} \, . \tag{4.3}$$

The local median and local variance were calculated over a window of 401 pixels, corresponding to a frequency window of approximately 1 Hz. In the top panel of fig. 4.5 we plot the histogram of $\chi_0$, where the index 0 refers to the fundamental frequency, that is, $\chi$ at the respective frequency. The resulting distribution deviates significantly from a Gaussian, as there is a significant excess for large $\chi_0$. These counts can easily be identified with the highest spectral peaks in fig. 4.3a as integer multiples of the neutron star rotation frequency of $\nu_0 = 0.1323$ Hz. The fat tails of the distribution make it hard to identify whether a peak sticks out of the tail, in particular for $\chi_0 \gtrsim 10$. For $\chi_0 \gtrsim 15$ all peaks can be identified and are related to the neutron star rotation period, except for one at $\nu \sim 9.187$ Hz. In table 4.1 we show all frequencies that have $\chi_0 > 11$ to have a selection of possible oscillation candidates. If there are two or more neighbouring frequencies with $\chi_0 > 11$, we list the highest value. All other candidates in table 4.1

Table 4.2.: Maximum $\chi_0$ at $\nu_{\max}$ in a 5% interval around previously observed oscillation frequencies $\nu$ for SGR 1806-20. We also show the local variance $\sigma_\chi$ of $\chi_0$ at this interval.

| $\nu$ [Hz] | $\nu_{\max}$ [Hz] in ±5% interval | $\chi_{0,\max}$ | $\sigma_\chi$ |
|---|---|---|---|
| 16.9 | 16.27 | 12.772 | 6.623 |
| 18.0 | 18.265 | 10.789 | 5.229 |
| 21.4 | 20.609 | 9.191 | 4.194 |
| 26.0 | 26.686 | 8.709 | 3.917 |
| 29.0 | 28.87 | 7.121 | 3.234 |
| 36.8 | 36.412 | 8.565 | 2.82 |
| 59.0 | 56.592 | 5.146 | 2.258 |
| 61.3 | 63.162 | 5.027 | 2.114 |
| 92.5 | 90.364 | 4.699 | 1.859 |
| 116.3 | 118.346 | 4.363 | 1.817 |
| 150 | 152.252 | 4.316 | 1.675 |

have significantly lower $\chi$ than the oscillation at 9.187 Hz and are consistent with being in the tail of the distribution that is shown in the top panel of fig. 4.5. This indicates that these are artefacts of the Poisson noise.

We also checked the $\chi_0$ values of previously reported frequencies. None of them reach more than $\chi_0 \gtrsim 5$. We therefor extended our search in a ±5% interval around the frequencies in table 4.2. The only frequency higher than $\chi_0 = 11$ is at $\nu = 16.27$ Hz. Thus all reported lines are consistent with noise. However, we already see an interesting pattern emerging: In table 4.2 we locally (within the ±5% interval) find the highest powers at 18.265 and 36.412 Hz. These are almost twice and four times the only significant frequency at $\nu = 9.187$ Hz that our method detects beyond the rotational frequency and its first 31 harmonics. In fig. 4.4 we plot the reconstructed power spectrum around the $\nu = 9.187$ Hz candidate oscillation (black line) and its first overtone (red line). The amplitude at $\nu = 9.187$ Hz is significantly larger (factor 2) than other amplitudes in the shown frequency range, while the amplitude at $\nu = 18.265$ Hz is comparable with other spectral peaks. As discrete frequencies in a power spectrum are likely to show spectral peaks at integer multiplies of a ground frequency, we also display the two-dimensional histogram of the calculated weighted residual $\chi_0$ at some ground frequency on the x-axis and its first harmonics $\chi_1$ on the y-axis in the bottom panel of fig. 4.5. We marked all counts with $\chi_1 \geq 5$ and $\chi_0 \geq 10$ with their corresponding frequency in Hz. We find about ten frequencies that satisfy this criterium. Obviously, all but the frequencies around $\nu \sim 9.186$ Hz are integer multiples of the rotation frequency $\nu_0$.

This further increases our confidence that $\nu \sim 9.186$ Hz is a candidate for an additional periodic signal in the data. We do not find any significant

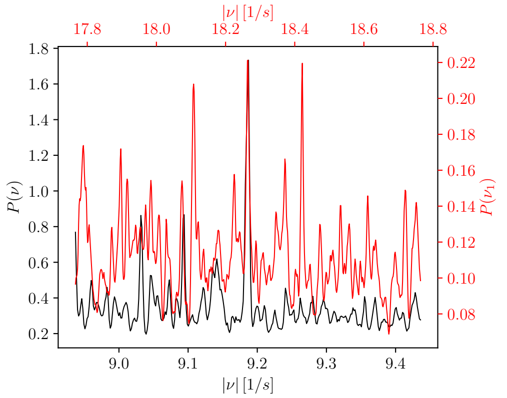


Figure 4.4.: Zoomed-in view of the reconstructed power spectra of giant flare SGR 1806-20 around $\nu \sim 9.186\,\text{Hz}$ in black and its first overtone around $\nu \sim 18.265\,\text{Hz}$ in red. The spectra correspond to $\sigma_{\text{sm}} = 5 \times 10^5$ as in fig. 4.3a

features for frequencies higher than the corresponding overtones that the algorithm recovered at $\nu \sim 18.265\,\text{Hz}$ and $\nu \sim 36.412\,\text{Hz}$.

### 4.3.2. SGR 1900+14

Analogously to SGR 1806-20, we also re-analysed the archival RXTE data of the giant flare of SGR 1900+14 that occurred on 1998 August 27. The resolution of the signal field is again 1/800 s, but here we only used $2^{18}$ pixels, as the giant flare did not last as long as that of SGR 1806-20. The operational down times[1] of RXTE during the observation were taken into account for the inference.

As for SGR1806-20, we were able to recover the frequency corresponding to the rotation as $\nu_0 = 0.1938\,\text{Hz}$, which is consistent with the reported $\nu_0 = 0.1923\,\text{Hz}$ [149]. For SGR 1900+14, the corresponding spectrum with the same smoothness prior $\sigma_{\text{sm}} = 5 \times 10^5$ is more noisy because of the various operational down times of the instrument, which are marked grey in fig. 4.6a. In fig. 4.6b, we show a snapshot of the entire reconstructed light curve for a more detailed view. Even though no data were recorded between $88.245\,\text{s} \leq \text{t} \leq 95.1375\,\text{s}$, the algorithm was able to infer a light curve by extrapolating from the data-constrained regions. This extrapolation is mainly driven by periodic features appearing in the data. As a further natural consequence, the one-$\sigma$ confidence interval increases drastically during down times. In total, these multiple operational down times of the instrument

---

[1] At the following intervals in seconds, the instrument did not record any data: $1.89 \leq \text{t} \leq 15.12; 20.84625 \leq \text{t} \leq 31.1175; 37.54125 \leq \text{t} \leq 47.11875; 55.405 \leq \text{t} \leq 63.12; 71.87875 \leq \text{t} \leq 79.12; 88.245 \leq \text{t} \leq 95.1375; 106.86 \leq \text{t} \leq 111.12; 124.285 \leq \text{t} \leq 127.1225; 141.935 \leq \text{t} \leq 143.1225$

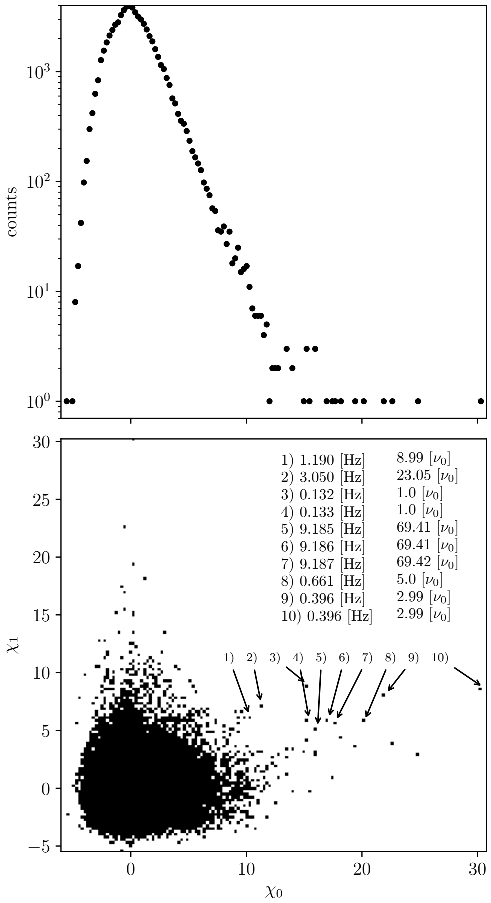
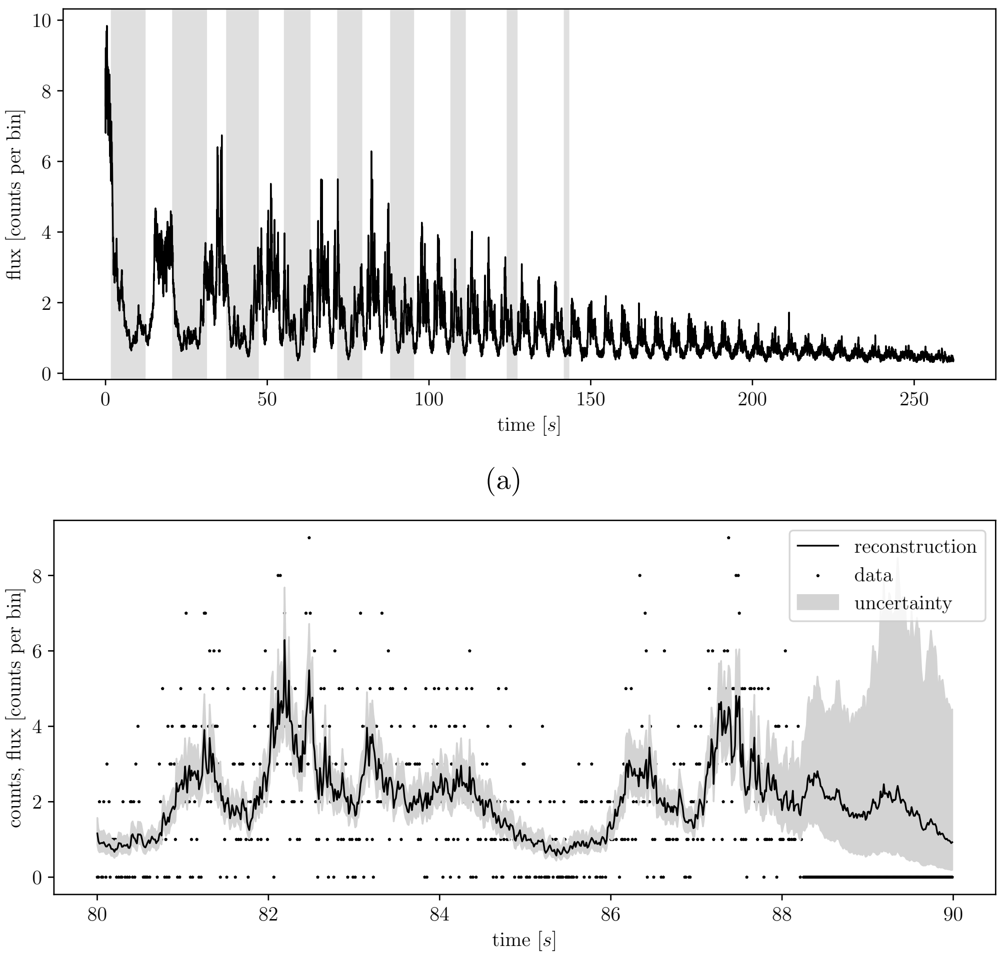
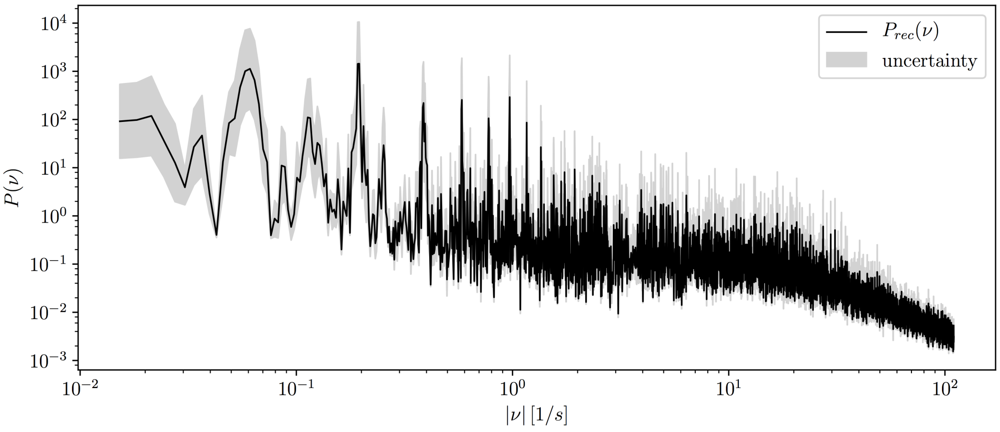


Figure 4.5.: Top panel: Histogram of the residuals between the logarithmic power spectrum of fig. 4.3a and its local median, weighted with its local variance $\sigma$. Bottom panel: Two-dimensional histogram of the same quantity at some frequency at the x-axis and its first harmonic at the y-axis. In addition, we indicate for all counts with $\chi_1 \geq 5$ and $\chi_0 \geq 10$ the corresponding frequency in Hertz and its multiple of the neutron star frequency, $\nu_0 = 0.1323\,\mathrm{Hz}$. To generate the histograms, we used the power spectrum shown in fig. 4.3a, i.e. $\sigma = 5 \times 10^5$.

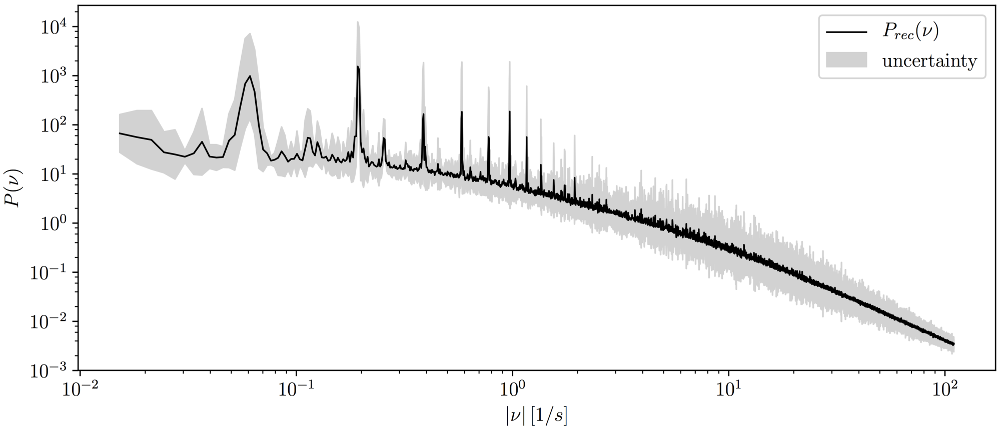

flux [counts per bin]
10
8
6
4
2
0
0
50
100
150
200
250
time [$s$]
counts, flux [counts per bin]
8
6
4
2
0
80
82
84
86
88
90
time [$s$]
reconstruction
data
uncertainty

(a)

(b)



(c)



(d)

Figure 4.6.: Figure 4.6a shows the reconstructed light curve $\phi$ for SGR 1900+14. The grey rectangles indicate the operational down times of the instrument. fig. 4.6b shows a snapshot of the entire reconstruction between $\approx 80$ sec and $\approx 91$ sec. For $t \gtrsim 88$ the instrument had an operational down time, leading to zero counts in this time interval. Figure 4.6c and fig. 4.6d shows the reconstructed power spectra as well as their uncertainties, using a smoothness-enforcing prior with $\sigma_{\mathrm{sm}} = 5 \times 10^5$ and $\sigma_{\mathrm{sm}} = 10^5$, respectively. The plots have the same volumes and sampling rate as in fig. 4.1.

Table 4.3.: All frequencies with $\chi_0 > 11$ and their multiplicity $n$ of the rotation period $\nu_0 = 0.1938$ [Hz] for SGR 1900+14.

| $\nu$ [Hz] | $\chi_0$ | $n$ [$\nu_0$] | $\nu$ [Hz] | $\chi_0$ | $n$ [$\nu_0$] |
|---|---|---|---|---|---|
| 7.693 | 13.214 | 39.7 | 11.768 | 13.718 | 60.7 |

Table 4.4.: Maximum $\chi_0$ at $\nu_{\max}$ in a 5% interval around previously observed oscillation frequencies $\nu$ for SGR 1900+14. We also list the local variance $\sigma_\chi$ of $\chi_0$ at this interval.

| $\nu$ [Hz] | $\nu_{\max}$ [Hz] in ±5% interval | $\chi_{0,\max}$ | $\sigma_\chi$ |
|---|---|---|---|
| 28.0 | 27.341 | 5.392 | 2.931 |
| 53.5 | 54.781 | 5.24 | 2.318 |
| 84 | 86.328 | 4.39 | 2.228 |
| 155.1 | 161.719 | 3.777 | 1.582 |

lead to fewer observed photons and in consequence to a weaker constrained reconstruction of the light curve as well as its power spectrum. Since a smoothness-enforcing prior $\sigma_{\mathrm{sm}} = 5 \times 10^5$ does not sufficiently denoise the light curve, we used a slightly stronger prior, $\sigma_{\mathrm{sm}} = 10^5$ (fig. 4.6d), for our further analysis.

With the same detection threshold as before, $\chi_0 > 11$, we find two candidate frequencies at 7.693 and 11.768 Hz, see table 4.3. In table 4.4 we give the maximum $\chi_0$ around previously observed frequencies. As for SGR 1806-20, we cannot confirm any of the previously detected frequencies, the highest significance we see is $\chi_0 > 5.4$ for 28.0 Hz. We further investigated our combined criterium, but now reduced to the giant flare of SGR 1900+14 compared to the flare of SGR 1806-20. Figure 4.9 shows only one additional frequency at $\nu = 7.695$ Hz, which strengthens our confidence in the previous finding in table 4.3 at $\nu = 7.693$ Hz. In fig. 4.7 we overplot the reconstructed power spectra around $\nu = 7.693$ Hz (black line) and its first overtone (red line). These two peaks are the largest in the given frequency range.

We plot the time evolution of the different pulses of SGR 1900+14 in fig. 4.8. As before, blue indicates the beginning of the flare, green the middle, and red the end. Here, the data are of inferior quality compared to the SGR 1806-20 and the parts of the curves without significant short time-variability are purely reconstructed by our algorithm. For SGR 1900+14 we observed four maxima, which also evolved differently compared to each other. For example, the weakest maximum (time ∼ 1 s) declines almost monotonically with time, while the others remain rather constant with far less variability in time.

In order to facilitate secondary studies based on our reconstructions and

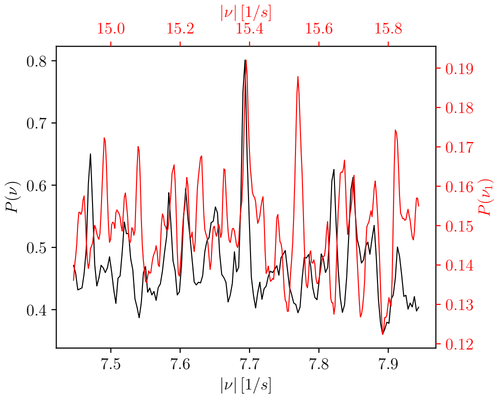


Figure 4.7.: Zoomed-in view of the reconstructed power spectra of the giant flare SGR 1900+14 around $\nu \sim 7.693\,\mathrm{Hz}$ in black and its first overtone around $\nu \sim 15.386\,\mathrm{Hz}$ in red. The plot corresponds to $\sigma_{\mathrm{sm}} = 1 \times 10^5$ as in fig. 4.6d.

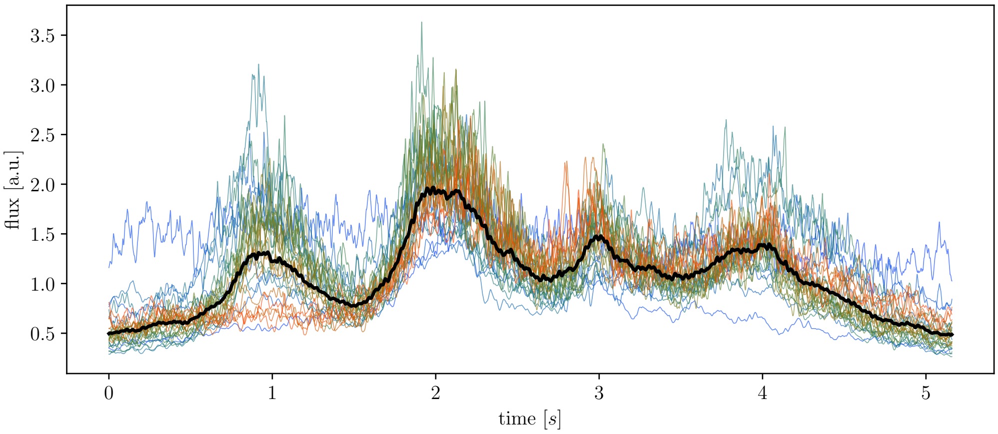


Figure 4.8.: Pulse profiles of different rotation periods overplotted for SGR 1900+14. The temporal evolution of the pulse profile is visible from blue to green to red. Smooth curves without significant short time-variability are reconstructed during instrument deadtimes.

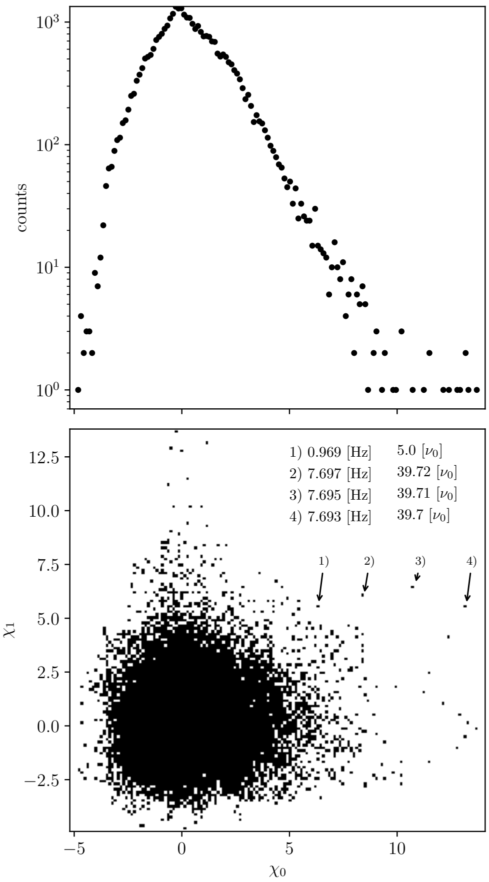


Figure 4.9.: Same as in fig. 4.5 for the SGR 1900+14 event, with $\sigma = 10^5$. We indicate for all counts with $\chi_0 \geq 5$ and $\chi_1 \geq 5$ the corresponding frequency in Hertz and its multiple of the neutron star frequency $\nu_0 = 0.1938\,\text{Hz}$.

spectra, we provide these data online at `http://cdsweb.u-strasbg.fr/cgi-bin/qcat?J/A+A/`. At `www.mpa-garching.mpg.de/ift/data/QPO` we also provide interactive plots of figs. 4.1, 4.3 and 4.6 for more detailed views.

## 4.4. Discussion

We recovered the rotation period of the two magnetars $\nu^{1806} = 0.1323$ Hz and $\nu^{1900} = 0.1938\,\mathrm{Hz}$, and up to the 31th overtone for SGR 1806-20 ($\chi_0 > 11$) and to the 18th for SGR 1900+14 ($\chi_0 > 7.5$). In addition, we only found potential periodic signals at 9.2 Hz (SGR1806-20) and 7.7 Hz (SGR1900+14), which are significant in $\chi_0$ and in the combination of $\chi_0$ and $\chi_1$ according to our respective criterium. For SGR 1900+14 and with the single criterium on $\chi_0$, we found another candidate frequency at $\nu = 11.8\,\mathrm{Hz}$, with a similar $\chi_0$ as the signal at $7.7\,Hz$, but without detected overtones. Therefore we only consider $\nu = 7.7\,\mathrm{Hz}$ as a potential signal. The most robust feature is at 9.2 Hz for SGR 1806-20, it has the highest $\chi_0$ , and we detect some of its overtones as local maxima in $\chi_0$ at 18.3 and 36.4 Hz, respectively. The two latter frequencies are similar to previously reported frequencies at 17.9 [114] and 36.8 Hz [117].

Although the reimplemented Bayesian inference algorithm D$^3$PO proved (see appendices B.1 to B.3) its basic ability to reconstruct QPO in photon bursts in many tests, even in the low S/N regime, we cannot confirm the previously reported frequencies at $17, 21, 26, 29, 59, 92.5, 116$, and 150 Hz for SGR 1806-20 and $28, 53, 84$, and 155 Hz for SGR 1900+14. In particular, we show in appendix B.3 that our algorithm is able to recover sufficiently strong quasi-periodic signals around 90 Hz with a width of $\sim 0.6\,\mathrm{Hz}$.

Our analysis and therefore all reconstructed power spectra and reconstructed photon fluxes depend on the particular chosen smoothness-enforcing prior $\sigma_{\mathrm{sm}}$ . Selecting a smaller $\sigma_{\mathrm{sm}}$ allows the algorithm to denoise the power spectrum even at low S/N. However, at high frequencies, the small $\sigma_{\mathrm{sm}}$ leads to a lower sensitivity for spectral lines. Therefore, a trade-off needs to be found for $\sigma_{\mathrm{sm}}$ between better denoising (small $\sigma_{\mathrm{sm}}$) and better sensitivity (large $\sigma_{\mathrm{sm}}$). Owing to the sharply deteriorating S/N for frequency ranges around 625 Hz and 1840 Hz, which would require a very small $\sigma_{\mathrm{sm}}$, we are currently not able to investigate these frequencies for QPOs.

We assume that our findings hold and determine the effect that the results have on the interpretation of the theoretical model of magnetar oscillations. First of all, the new candidate frequencies at 9.2 and 7.7 Hz for SGR1806-20 and SGR1900+14, respectively, are much lower than the frequencies reported so far. Here, our method has a clear advantage over previous studies, as we can analyse the entire data set at once, meaning that no parts of the data set are left out. In principle, this leads to a more robust reconstruction and also counteracts statistical artefacts that may be introduced by a windowed analysis. Our method is fully noise-aware and has in principle no problems with observational dead times. If these two frequencies are related to neutron star oscillations, the parameters of current models change significantly. When we assume that the oscillations are pure crustal shear modes, the identification of the low frequency with the $n = 0, l = 2$ mode would favour models with rather low shear speeds and therefore EoS with a fast increase of symmetry energy [125]. However, pure shear models are not very likely because of the strong interaction with the magnetic field in the interior of the star [133, 134, 136, 137]. If the magnetic field is not neglected, coupled magneto-elastic oscillations need to be considered. In this case, the much lower fundamental oscillation frequency indicates a magnetic field of the order of $\bar{B} \sim 6 \times 10^{13} - 3 \times 10^{14}$ G, lower by a factor of $\sim 3$ than our estimates in [143]. With such weak magnetic fields, the oscillations would also remain confined to the core for models with very strong shear modulus, and there would be no chance to observe the oscillations exterior to the star. Therefore, similarly to the case of pure crustal shear oscillations, the low frequencies of 9.2 or 7.7 Hz favour EoSs that give a small shear modulus. However, the estimation of the magnetic field strength has some serious problems : i) The spin-down estimate is only accurate to a factor of a few. ii) The particular magnetic field configuration inside a magnetar is unknown, meaning that even for dipolar-like fields in the exterior, there are different possible realization in the interior. iii) In addition to the degeneracies of the dipolar magnetic field strength and the magnetic field configuration, which lead to comparable frequencies, the compactness of the neutron star and the superfluid properties of the core matter also influence the frequencies significantly [146]. In order to advance and to lift some of these degeneracies, we need more observations of QPOs, and if possible, observations for different sources.

In respect of the potential of improving our method by modelling the spectra as an independent combination of a continuum and lines, we postpone a more thorough discussion of these consequences to future work. We also plan to improve our algorithm by taking the deadtime after each photon detection into account and include further instrument responses. However, we expect these improvements to increase our detection limits only by about a few per cent.

We have furthermore shown the capability of our code D$^3$PO to denoise and recover the shape of the light curve. This may have interesting applications in the field of modelling the pulses of X-ray bursters.

## 4.5. Conclusion

We have applied a new Bayesian method, D$^3$PO to analyse time-series data of the giant flares of SGR 1806-20 and SGR 1900+14. Thereby, we ensured that the photon shot noise, operational down times of the instrument, and the a priori unknown power spectrum of the photon flux were taken into account. In order to denoise and reconstruct the logarithmic photon flux and its power spectrum simultaneously from the data, we had to assume some kind of spectral smoothness on the power spectrum to counteract data variance. Our analysis cannot confirm previous findings of QPOs in the giant flare data, but we found new candidates for periodic signals at 9.2 Hz for SGR 1806-20 and 7.7 Hz for SGR1900+14. If these are real and related to the lowest frequency oscillation of the magnetar, our results favour high-compactness, weaker magnetic fields than were assumed before, and low shear moduli. The necessary application of a smoothness-enforcing prior results in a reduced sensitivity to the spectral line in the power spectrum. Hence, we propose for future work to decompose the power spectrum into two components, one smooth background spectrum, and one consisting of spectral lines. In this way, the previously reported lines may be confirmed or refuted with higher confidence, or additional and as yet unknown frequencies in QPOs might be discovered.

# 5. Further work

*This chapter lists all further publications I have been part of. As my contributions to those are marginal these publications are not shown in their full extend, only their abstracts.*

## 5.1. NIFTy 3 - Numerical Information Field Theory - A Python framework for multicomponent signal inference on HPC clusters

*This section is used additionally as a publication submitted to PLOS one (Steininger* et al.*, [91]).*

NIFTy, "Numerical Information Field Theory", is a software framework designed to ease the development and implementation of field inference algorithms. Field equations are formulated independently of the underlying spatial geometry allowing the user to focus on the algorithmic design. Under the hood, NIFTy ensures that the discretization of the implemented equations is consistent. This enables the user to prototype an algorithm rapidly in 1D and then apply it to high-dimensional real-world problems. This paper introduces NIFTy 3, a major upgrade to the original NIFTy framework. NIFTy 3 allows the user to run inference algorithms on massively parallel high performance computing clusters without changing the implementation of the field equations. It supports n-dimensional Cartesian spaces, spherical spaces, power spaces, and product spaces as well as transforms to their harmonic counterparts. Furthermore, NIFTy 3 is able to treat non-scalar fields. The functionality and performance of the software package is demonstrated with example code, which implements a real inference algorithm from the realm of information field theory. NIFTy 3 is open-source software available under the GNU General Public License v3 (GPL-3).

# 6. Conclusion

## 6.1. Summary

This thesis aimes to bridge between astrophysical observations and theory. To gain insights into complex dynamic system, we developed the dynamic system classifier algorithm (DSC). It might be of particular interest for analysing the detection of a gravitational wave caused by a stellar core collapse. In order to advance the mathematical toolkit to infer information from photon count observations, as needed to get reliable estimates about different superimposed astrophysical photon fluxes, we generalised the D$^3$PO code towards more generic and multi-dimensional photon count data sets. Finally we applied this D$^4$PO code to analyse the giant magnetar flares SGR 1806-20 and SGR 1900+14.

The first part of this thesis focuses on the development of the DSC, a tool to analyse stochastic and dynamically evolving systems. DSC consists of two steps. First it abstracts classifying information from just a few training samples of a class of systems. The classifying information is encoded via time dependent coefficients of stochastic differential equations. In the second step the algorithm confronts real noise contaminated observations with the previously learned model classes. Thereby it provides the probability which class most likely explains the observation.

We introduced a hierarchical Bayesian inference model to minimize the influence of any a prior assumptions. In numerical tests we showed that one needs at least three data realisations of each system class to sufficiently characterise each of them. This is mainly due to the high degeneracy of our problem, as we try to abstract each stochastically evolving class, simply by its frequency and damping factor evolution. After learning a number of system classes, the algorithm properly identified the correct classes in noise dominated data sets up to an SNR of $\approx 0.01$ per measurement.

The strength of the proposed and tested DSC approach is that it depends on very few prior assumptions and that only a hand full of training data is needed. Hence the algorithm should find use whenever one cannot afford to forward simulate a complex dynamical system over its entire possible physical phase space due to high computational costs. This holds in particular for gravitational wave signals from core collapse supernova as it is up to date not possible to simulate such systems in their full extent, due to their computational complexity.

For the large variety of space based telescopes that observe the celestial sphere, via the registration of individual photons we developed D$^4$PO. D$^4$PO denoises, deconolves, and decomposes multidimensional photon observations into morphologically different sources. It aims to decompose the high energy photon count data into three different components: diffuse, point-like, and background radiation. Detailed care has been taken to accurately built the likelihood model, which incorporates a Poissonian distribution and many instrument artefacts. In contrast to its predecessor, D$^4$PO allows to reconstruct emission components which may live over the product space of multiple manifolds, such as energy and time. This permits to reconstruct fields, which do not necessarily need to be statistical isotropic and homogeneous. In addition to the simultaneous reconstruction of all components and the provision of their uncertainty estimates, the algorithm infers the correlation structure for each component over all of its sub-manifolds. The full capabilities of the algorithm have been shown in a simulated high energy photon count data set.

Finally we applied the new D$^4$PO algorithm[1] to analyse the data of SGR1806-20 and SGR 1900+14. Our analysis could not confirm previously reported QPOs, despite exhibiting the required sensitivity in the signal injection experiments. However we found new QPO candidates at $\approx 9.2$ Hz for SGR 1806-20 and $\approx 7.7$ Hz for SGR 1900+14. As it is statistically not straight forward to state which of these frequencies correspond to real physical signals of these magnetars or whether these are just noise artefacts, one has to be careful with scientific conclusions.
However, if our findings hold, our data analysis would favour a magnetic field of $\bar{B} \approx 6 \times 10^{13} - 3 \times 10^{14}$G. This is a factor of $\approx 3$ smaller compared to previous estimates [150]. With such weak magnetic fields, the oscillations would remain to be confined to the core and there would be no chance to observe these oscillations at the exterior of the star.

## 6.2. Outlook

As this thesis aims at advancing the mathematical toolkits to explore astrophysical data set in order to get reliable estimates about the underlying physical phenomena all developed methods have a wide range of applications. How one may apply these methods and how they may be refined is listed below:

---

[1] The D$^3$PO algorithm would have the principle capabilities, but its implementation could not handle the required high resolution in the spectral domain

- **Dynamic system classifier-** Having shown the principle capabilities of the DSC algorithm the next logical step is to apply it to real data. As outlined the DSC would be ideal to classify a potential gravitational wave signal caused by a core collapse supernova. This implies that the DSC can not only classify observations into different classes but also needs to search for them in very long timelines. Hence one has to develop a technique to confront the abstracted information from the small training data sets to real noise contaminated data sets, where one does not know if and when the GW signal is present in the data.

- **D$^4$PO-** As the algorithm is implemented in a rather abstract way on the basis of NIFTy3.5 [148] it is in principle applicable to any photon count data observation. In the context of astrophysics the first and most obvious application fields would be to analyse the space based X-ray and $\gamma$-ray observatories, such as
  - Rosat (0.042 keV - 2 keV),
  - Chandra X-ray Observatory (0.09 keV - 10 keV),
  - XMM-Newton (0.15 keV- 15keV),
  - eRosita (0.3 keV - 10 keV, planned launch 2018),
  - Integral (15 keV - 10 MeV),
  - Compton gamma ray observatory (20 keV - 30 GeV),
  - Fermi Gamma-ray Space Telescope (20 MeV - 300 GeV).

  Each of these satellite mission provides spatial and spectral information for each detected photon count and therefore can be used to reconstruct the sky as a function of energy and location. As D$^4$PO is designed for exactly this task, one may even go one step further and do a joint reconstruction of the sky in four dimensions, including time as a further parameter. This involves to analyse the data simultaneously and would provide a reconstructed sky from $\approx$ 0.0042 keV up to $\approx$ 300 GeV, if the full range of high energy telescopes would be analysed jointly. Such an inference would allow to propagate spatial and spectral information across the entire reconstruction through the correlation structure of each manifold. The imprinted redundancies of the physical signals in the joint data set would support further a very efficient background estimation and therefore give more reliable estimates about the diffuse and point-like photon flux. Higher energies than reached by the Fermi Gamma-ray Space Telescope may be included into such an analysis by the usage of Cherenkov telescopes, such as the Cherenkov telescope array (CTA).

  In the context of the analysis of the events SGR 1806-20 and SGR

1900+14 one may continue to analyse the observed $\gamma$-ray burst events by the Fermi Gamma-ray Space telescope. This can potentially lead to further constraints on the EoS of neutron stars.

# A. Supplement to D$^4$PO

## A.1. Covariances and curvatures

The covariance $D$ of a Gaussian $\mathcal{G}\left(\phi - \bar{\phi}, D\right)$ describes the uncertainty associated with the mean of the distribution. It may be computed via the inverse Hessian of the corresponding information Hamiltonian or Gibbs free energy respectively,

$$\begin{aligned} \left.\frac{\partial^2 G}{\partial \phi \partial \phi^\dagger}\right|_{\phi=\bar{\phi}} &= \left.\frac{\partial^2}{\partial \phi \partial \phi^\dagger}\left(\frac{1}{2}\left(\phi - \bar{\phi}\right) D^{-1} \left(\phi - \bar{\phi}\right)\right)\right|_{\phi=\bar{\phi}} \\ &= D^{-1}\,. \end{aligned} \tag{A.1}$$

The uncertainty covariances of the derived information Hamiltonian eq. (3.31) are,

$$D^{(\varphi)-1} = \Phi_{\boldsymbol{t}}^{-1} + \left(1 - \frac{\boldsymbol{d}}{\boldsymbol{l}}\right)^\dagger \boldsymbol{\mathcal{R}} * \widehat{e^{\boldsymbol{m}}} + \boldsymbol{\mathcal{R}}^\dagger \widehat{e^{\boldsymbol{m}}} \frac{\widehat{d}}{l^2} \boldsymbol{\mathcal{R}}^\dagger \widehat{e^{\boldsymbol{m}}}\,, \tag{A.2}$$

$$D^{(\varphi_{\boldsymbol{u}_x})-1} = D^{(\varphi)-1} + \boldsymbol{\eta} e^{-\boldsymbol{m}_{\boldsymbol{u}_x}}\,, \tag{A.3}$$

$$D^{(\boldsymbol{\tau})-1} = \boldsymbol{T} + \left(q + \frac{w}{2}\right) e^{-\boldsymbol{t}} \tag{A.4}$$

with

$$\boldsymbol{l} = \boldsymbol{\mathcal{R}} e^{\boldsymbol{\varphi}}\,. \tag{A.5}$$

The corresponding covariances of the chosen Gibbs approach in eq. (A.19) are,

$$D^{-1} = \begin{pmatrix} D^{(\boldsymbol{\varphi})^{-1}} & 0 \\ 0 & D^{(\boldsymbol{\tau})^{-1}} \end{pmatrix}, \quad \text{with} \tag{A.6}$$

$$\begin{aligned} D^{(\varphi)-1} &= \mathcal{T}\left( \Phi_{\boldsymbol{t}'}^{-1} + \left(1 - \frac{\boldsymbol{d}}{\boldsymbol{l}}\right)^{\dagger} \boldsymbol{\mathcal{R}} * e^{\widehat{\boldsymbol{m}+\frac{1}{2}\widehat{D}^{(\varphi)}}} \right. \\ &\left. \quad + \boldsymbol{\mathcal{R}}^{\dagger} e^{\widehat{\boldsymbol{m}+\frac{1}{2}\widehat{D}^{(\varphi)}}} \frac{\widehat{d}}{l^2} \boldsymbol{\mathcal{R}}^{\dagger} e^{\widehat{\boldsymbol{m}+\frac{1}{2}\widehat{D}^{(\varphi)}}} \right), \end{aligned} \tag{A.7}$$

$$D^{(\varphi_{\boldsymbol{u}_x})-1} = D^{(\varphi)-1} + \boldsymbol{\eta} \exp\left( -\boldsymbol{m}_{\boldsymbol{u}_x} + \frac{1}{2} \widehat{D}^{(\varphi_{\boldsymbol{u}_x})} \right), \tag{A.8}$$

$$D^{(\boldsymbol{\tau})-1} = \mathcal{T}\left( \boldsymbol{T} + \left( q + \frac{w}{2} \right) e^{-\boldsymbol{t}'} \right) \tag{A.9}$$

with

$$\boldsymbol{l} = \boldsymbol{\mathcal{R}} e^{\boldsymbol{m}+\frac{1}{2}\widehat{D}^{(\varphi)}}. \tag{A.10}$$

Up to the term $\frac{1}{2}\widehat{D}$ in the exponential functions, these covariances are identical to the ones derived via the maximum a posteriori ansatz eq. (A.4). This shows that these higher order corrections terms change the uncertainty covariances, however their influence is hard to judge as they introduce terms that couple to all elements of $D$. It must be noted that the inverse Hessian only describes the curvature of the potential, hence it may only be regarded as the uncertainty of a reconstruction if the potential is quadratic. However, numerous numerical experiments showed that this assumption holds in most cases.

## A.2. Deriving the Gibbs free energy

Due to the complex structure of the information Hamiltonian in eq. (3.31), we are seeking for an approximation to the true posterior eq. (3.28). To this end we adapt a Gaussian distribution for our posterior approximation and require it to be close in an information theoretical sense to the correct posterior. The correct information distance is the Kullback-Leibler divergence, which is equivalent, up to irrelevant constants, to the Gibbs free energy [58]. To do so, we directly adopt the final functional form of the posterior in the construction of the Gibbs free energy. Here we use an approximate Gaussian ansatz for the posterior eq. (3.28) of our signal vector $\phi = \left(\boldsymbol{\varphi}^\dagger, \boldsymbol{\tau}^\dagger\right)^\dagger$:

$$P(\phi|\boldsymbol{d}) = \mathcal{G}\left(\phi - \bar{\phi}, D\right) \quad \text{with} \tag{A.11}$$

$$\bar{\phi} = \left(\boldsymbol{m}\dagger, \boldsymbol{t}^\dagger\right)^\dagger , \tag{A.12}$$

the posterior mean, and

$$D = \begin{pmatrix} D^{(\boldsymbol{\varphi})} & 0 \\ 0 & D^{(\boldsymbol{\tau})} \end{pmatrix} \tag{A.13}$$

the posterior uncertainty covariance. The posterior mean $\bar{\phi}$ consists of the mean field $\boldsymbol{m} = \langle \boldsymbol{\varphi} \rangle_{(\phi|d)}$, as well as the mean log power spectrum $\boldsymbol{t} = \langle \boldsymbol{\tau} \rangle_{(\phi|d)}$. The signal covariance $D = \langle (\phi - \bar{\phi})(\phi - \bar{\phi})^\dagger \rangle_{(\phi|d)}$ consists of $2 \times 2$ block matrices, where the off diagonal terms are set to zero to reduce the complexity of the resulting algorithm. The non zero blocks are the signal uncertainty

$$D^{(\boldsymbol{\varphi})} = \left\langle (\boldsymbol{\varphi} - \boldsymbol{m})\, (\boldsymbol{\varphi} - \boldsymbol{m})^\dagger \right\rangle_{(\phi|d)} , \tag{A.14}$$

and the log-spectrum uncertainty

$$D^{(\boldsymbol{\tau})} = \left\langle (\boldsymbol{\tau} - \boldsymbol{t})\, (\boldsymbol{\tau} - \boldsymbol{t})^\dagger \right\rangle_{(\phi|d)} . \tag{A.15}$$

In terms of these parameters the Gibbs free energy is given by

$$G(\bar{\phi}, D|d) = U(\bar{\phi}, D|d) - \mathcal{T}\mathscr{S}(\bar{\phi}, D|d) , \tag{A.16}$$

with

$$U(\bar{\phi}, D|d) = \langle H\left(\boldsymbol{\varphi}, \boldsymbol{\tau}|d\right) \rangle_{(\phi|d)} , \tag{A.17}$$

being the internal energy, describing the full non-Gaussian Hamiltonian eq. (3.31), averaged by the approximated posterior eq. (A.11). The entropy of the approximated Gaussian posterior is

$$\mathscr{S}(\bar{\phi}, D|d) = -\int \mathcal{D}\phi\, P(\phi|d) \log P(\phi|d) . \tag{A.18}$$

Up to an irrelevant sign and an additive constant the Gibbs free energy eq. (A.16) is equal to the Kullback-Leibler distance of the posterior approximation if $\mathcal{T} = 1$. Unless stated differently, we therefore usually set $\mathcal{T} = 1$. If $\mathcal{T} = 0$, the approximate posterior given by eq. (A.11) becomes a delta function at the maximum of the correct posterior, eq. (3.28), which might perform poorly in case nuisance parameters such as $\boldsymbol{\tau}$ also need to be reconstructed. In total the Gibbs free energy becomes

$$
\begin{aligned}
G =& G\left(\bar{\phi}, D|\boldsymbol{d}\right) \\
=& G_0 + \mathbf{1}^\dagger \boldsymbol{l} - \boldsymbol{d}^\dagger \log \boldsymbol{l} \\
&+ \frac{1}{2}\Big[\operatorname{Tr}\left[\log \Phi\right] + \boldsymbol{m}^\dagger \Phi^{-1} \boldsymbol{m} + \operatorname{Tr}\left(\Phi^{-1} D^{(\boldsymbol{\varphi})}\right)\Big] \\
&+ (\alpha - 1)^\dagger \boldsymbol{t} + q^\dagger \exp\left(-\boldsymbol{t} + \frac{1}{2}\widehat{D}^{(\boldsymbol{\tau})}\right) \\
&+ \frac{1}{2}\boldsymbol{t}^\dagger \boldsymbol{T} \boldsymbol{t} + \frac{1}{2}\operatorname{Tr}\left[\boldsymbol{T} D^{(\boldsymbol{\tau})}\right] \\
&+ (\beta - 1)^\dagger \boldsymbol{m}_{\boldsymbol{u}_x} + \boldsymbol{\eta}^\dagger \exp\left(-\boldsymbol{m}_{\boldsymbol{u}_x} + \frac{1}{2}\widehat{D}^{(\varphi_{\boldsymbol{u}_x})}\right) \\
&- \frac{\mathcal{T}}{2}\operatorname{Tr}\left[\mathbb{1} + \ln\left(2\pi D\right)\right] ,
\end{aligned}
\tag{A.19}
$$

with $\Phi = \operatorname{diag}\left(\mathcal{S}, \mathcal{U}, \mathcal{B}\right)$, $G_0$ is absorbing all constants, and

$$
\boldsymbol{l} = \boldsymbol{\mathcal{R}} e^{\boldsymbol{m} + \frac{1}{2}\widehat{D}^{(\varphi)}} \, . \tag{A.20}
$$

Comparing eq. (A.19) with the information Hamiltonian eq. (3.31), there are a number of correction terms appearing which properly account for the uncertainty of the inferred map $\boldsymbol{m}$. In particular $\boldsymbol{l}$ differs, comparing eq. (3.33) with eq. (A.20), which in the framework of Gibbs free energy describes the expectation value of $\boldsymbol{\lambda}$ over the approximate posterior eq. (A.11). Minimizing the Gibbs free energy with respect to $\boldsymbol{m}, \boldsymbol{t}$ and $D$ would optimize the inference under the assumed Gaussian posterior.

In the following sections we will gradually derive the Gibbs free energy.

### The Entropy

Due to the Gaussian ansatz eq. (A.11) the entropy eq. (A.18) is independent of $\bar{\phi}$,

$$
\mathscr{S}(\bar{\phi}, D|d) = \frac{1}{2}\operatorname{Tr}\left[\mathbb{1} + \ln(2\pi D)\right] \tag{A.21}
$$

and therefore its gradients with respect to $\boldsymbol{m}$ and $\boldsymbol{t}$ vanish,

$$
\frac{\partial \mathscr{S}}{\partial \boldsymbol{m}} = 0 \, , \quad \frac{\partial \mathscr{S}}{\partial \boldsymbol{t}} = 0 \, . \tag{A.22}
$$

### Internal energy of the hyperprior

The internal energy of the hierarchical Gaussian prior $P(\boldsymbol{\tau}|\sigma, \alpha, q)$ is

$$\begin{aligned} U^{(\boldsymbol{\tau})}\left(\bar{\phi}, D|\boldsymbol{d}\right) =& \langle H(\boldsymbol{\tau})\rangle_{(\phi,D|\boldsymbol{d})} \qquad \text{(A.23)} \\ \simeq& \frac{1}{2}\boldsymbol{t}^{\dagger}\boldsymbol{T}\boldsymbol{t} + \frac{1}{2}\operatorname{Tr}\left[\boldsymbol{T}D^{(\boldsymbol{\tau})}\right] \\ &+ (\alpha-1)^{\dagger}\,\boldsymbol{t} + q^{\dagger}e^{-\boldsymbol{t}+\frac{1}{2}\widetilde{D}^{(\boldsymbol{t})}}\,, \qquad \text{(A.24)} \end{aligned}$$

with $\boldsymbol{T} = (T_{\mathcal{Y}^{(s)}}, T_{\mathcal{X}^{(s)}}, T_{\mathcal{Y}^{(u)}}, T_{\mathcal{Y}^{(b)}}, T_{\mathcal{I}^{(b)}})^{\dagger}$. A hat on a tensor denotes the diagonal vector in the position basis, $\widehat{D}_x = D_{xx}$, while a hat on a vector refers to a tensor with the vector as its diagonal, $\widehat{m}_{xy} = \delta_{xy}m_x$. Similarly we define a tilde on a tensor as the diagonal vector in the band harmonic basis, $\widetilde{\theta}_k = \theta_{kk}$, and a tilde on a vector denotes a tensor with the vector on its diagonal in the band harmonic basis, $\widetilde{t}_{kl} = \delta_{kl}t_k$.

The corresponding non-vanishing gradients of the hyperprior's internal energy eq. (A.24) are

$$\begin{aligned} \frac{\partial U^{(\boldsymbol{\tau})}}{\partial \boldsymbol{t}} =& \boldsymbol{T}\boldsymbol{t} + (\alpha-1) - qe^{-\boldsymbol{t}'}\,, \qquad \text{(A.25)} \\ \frac{\partial U^{(\boldsymbol{\tau})}}{\partial D^{(\boldsymbol{\tau})}} =& \frac{1}{2}\left[\boldsymbol{T} + \widetilde{qe^{-\boldsymbol{t}'}}\right]\,, \qquad \text{(A.26)} \end{aligned}$$

with $\boldsymbol{t}' = t - \frac{1}{2}\widetilde{D}^{(\boldsymbol{\tau})}$.

### Internal energy of the prior for $\varphi$

The priors for $\boldsymbol{\varphi}$ provide the internal energy

$$\begin{aligned} U^{(\boldsymbol{\varphi})}(\bar{\phi}, D|\boldsymbol{d}) =& \frac{1}{2}\operatorname{Tr}\left[\log \Phi_{\boldsymbol{t}'}\right] \\ &+ \frac{1}{2}\left[\boldsymbol{m}^{\dagger}\Phi_{\boldsymbol{t}'}^{-1}\boldsymbol{m} + \operatorname{Tr}\left[\Phi_{\boldsymbol{t}'}^{-1}D^{(\boldsymbol{\varphi})}\right]\right]\,. \qquad \text{(A.27)} \end{aligned}$$

The corresponding gradients are

$$\begin{aligned}
\frac{\partial U^{(\boldsymbol{\varphi})}}{\partial \boldsymbol{m}} =& \Phi_{\boldsymbol{t}'}^{-1} \boldsymbol{m} \,, && \text{(A.28)} \\
\frac{\partial U^{(\boldsymbol{\varphi})}}{\partial D^{(\boldsymbol{\varphi})}} =& \frac{1}{2} \Phi_{\boldsymbol{t}'}^{-1} && \text{(A.29)} \\
\frac{\partial U^{(\boldsymbol{\varphi})}}{\partial t_{\mathcal{Y}}} =& \frac{1}{2} \left( \rho_{\mathbb{P}} - w_{t'_{\mathcal{Y}}} e^{-t'_{\mathcal{Y}}} \right) \,, && \text{(A.30)} \\
\frac{\partial U^{(\boldsymbol{\varphi})}}{\partial t_{\mathcal{X}}} =& \frac{1}{2} \left( \rho_{\mathbb{P}} - w_{t'_{\mathcal{X}}} e^{-t'_{\mathcal{X}}} \right) \,, && \text{(A.31)} \\
\frac{\partial U^{(\boldsymbol{\varphi})}}{\partial D^{(t_{\mathcal{X}})}} =& \frac{1}{4} \left( \widetilde{w_{t'_{\mathcal{X}}} e^{-t'_{\mathcal{X}}}} \right) \,, && \text{(A.32)} \\
\frac{\partial U^{(\boldsymbol{\varphi})}}{\partial D^{(t_{\mathcal{Y}})}} =& \frac{1}{4} \left( \widetilde{w_{t'_{\mathcal{Y}}} e^{-t'_{\mathcal{Y}}}} \right) \,, && \text{(A.33)}
\end{aligned}$$

with

$$\begin{aligned}
w_{t'_{\mathcal{Y}}} &= \mathrm{Tr} \left[ \mathcal{X}^{(\boldsymbol{\varphi})^{-1}} \mathbb{P} \left( \boldsymbol{m}^{\dagger} \boldsymbol{m} + D^{(\boldsymbol{\varphi})} \right) \right] \,, && \text{(A.34)} \\
w_{t'_{\mathcal{X}}} &= \mathrm{Tr} \left[ \mathcal{Y}^{(\boldsymbol{\varphi})^{-1}} \mathbb{P} \left( \boldsymbol{m}^{\dagger} \boldsymbol{m} + D^{(\boldsymbol{\varphi})} \right) \right] && \text{(A.35)}
\end{aligned}$$

and with the multiplicities $\rho_{\mathbb{P}} = \mathrm{Tr}\, \mathbb{P}$ of the spectral bands $k$ and $l$, which are sharing the same harmonic eigenvalue.

### Internal energy of the Inverse-Gamma prior

The internal energy of the inverse-gamma prior on $\boldsymbol{\varphi}_{\boldsymbol{u}_x}$ is

$$U^{(\boldsymbol{\varphi}_{\boldsymbol{u}_x})}(\bar{\phi}, D|\boldsymbol{d}) = (\beta - 1)^{\dagger} \, \boldsymbol{m}_{\boldsymbol{u}_x} + \boldsymbol{\eta}^{\dagger} \exp \left( -\boldsymbol{m}_{\boldsymbol{u}_x} + \frac{1}{2} \widehat{D}^{(\varphi_{\boldsymbol{u}_x})} \right) \,, \qquad \text{(A.36)}$$

with its corresponding gradient

$$\frac{\partial U^{(\boldsymbol{\varphi}_{\boldsymbol{u}_x})}}{\partial \boldsymbol{m}_x} = \boldsymbol{\beta} - 1 - \boldsymbol{\eta} * \exp \left( -\boldsymbol{m}_{\boldsymbol{u}_x} + \frac{1}{2} \widehat{D}^{(\varphi_{\boldsymbol{u}_x})} \right) \,. \qquad \text{(A.37)}$$

### Internal energy of the likelihood

The internal energy of the likelihood eq. (3.8) is

$$\begin{aligned} U(\bar{\phi}, D|\boldsymbol{d}) &= \langle H(d|\boldsymbol{\varphi}, D)\rangle_{(\phi, D|\boldsymbol{d})} \\ &\simeq G_0 + \mathbf{1}^\dagger \boldsymbol{l} \\ &\quad - \boldsymbol{d}^\dagger \left\{ \log(\boldsymbol{l}) - \sum_{\nu=2}^{\infty} \frac{(-1)^\nu}{\nu} \left\langle \left( \frac{\boldsymbol{\lambda}}{\boldsymbol{l}} - 1 \right)^\nu \right\rangle \right\}, \end{aligned} \tag{A.38}$$

with

$$\boldsymbol{\lambda} = \boldsymbol{\mathcal{R}} e^{\boldsymbol{\varphi}}, \tag{A.39}$$

$$\boldsymbol{l} = \boldsymbol{\mathcal{R}} e^{\boldsymbol{m} + \frac{1}{2}\widehat{D}^{(\varphi)}}, \tag{A.40}$$

where we absorbed all terms that are constant in $\boldsymbol{\varphi}$ into $G_0$. The evolution of the internal energy would require to know all entries of $D^{(\varphi)}$ explicitly. As this is computationally infeasible and the term $\left\langle \left(\frac{\boldsymbol{\lambda}}{\boldsymbol{l}} - 1\right)^\nu \right\rangle \approx 0$ at the mode where $\lambda \approx l$, this sum can be neglected. As a consequence all cross correlations, such as $D^{(\boldsymbol{su})}$ are implicitly set to zero.

The gradient of eq. (A.38) can be derived by taking the first derivative of $U(\bar{\phi}, D|\boldsymbol{d})$ with respect to the approximate mean estimates,

$$\frac{\partial U}{\partial \boldsymbol{m}} = \left(1 - \frac{\boldsymbol{d}}{\boldsymbol{l}}\right)^\dagger \boldsymbol{\mathcal{R}} * e^{\boldsymbol{m} + \frac{1}{2}\widehat{D}^{(\varphi)}}. \tag{A.41}$$

# B. Supplement to QPO signals in magnetar giant flares

## B.1. Influence of the smoothness-enforcing parameter $\sigma$

To demonstrate the performance of the Bayesian inference algorithm, we challenged the new implementation of the algorithm with mock data. To do so, we drew Poisson samples from $\lambda = \boldsymbol{R}e^{\boldsymbol{s}}$. For simplicity, we assumed $\boldsymbol{R}$ to be the identity operator. The signal field $\boldsymbol{s}$ was drawn from a Gaussian random field $\boldsymbol{s} \hookleftarrow \mathscr{G}(s, S)$, with

$$S_{\nu\nu'} = \frac{27}{\left(1 + \nu/\nu_0\right)^2} \delta_{\nu\nu'} \,, \tag{B.1}$$

with $\nu_0 = 1$ absorbing the physical units. To test whether the algorithm can reconstruct discrete frequencies as well, we further injected four $\delta$-peaks into $S_{\nu\nu'}$ at randomly chosen positions. The peak heights decrease at higher frequencies to be as realistic as possible. For the test we used a one-dimensional regular grid with with $2^{16}$ pixels, each with a volume of 1/1000 seconds.

In fig. B.1 we tested the principle capabilities of the algorithm under the above-mentioned setting while varying the strength of the smoothness enforcing parameter $\sigma_{\mathrm{sm}}$. In the scenario of $\sigma_{\mathrm{sm}} = 10^4$, shown in figs. B.1a and B.1b, the algorithm was able to recover the signal as well as the power spectrum well within the one-$\sigma$ confidence interval. However, the two highest injected frequencies could not be recovered, as their injected strength and therefore their effect on the data was too weak to be distinguished from the Poisson shot noise.

In comparison to fig. B.1b, fig. B.1c shows the reconstructed power spectrum using a significantly weaker smoothness enforcing $\sigma_{\mathrm{sm}} = 10^5$. As a first obvious consequence, the local variance of the spectrum increases by multiple orders of magnitude. However, the same two frequencies as with the stronger $\sigma_{\mathrm{sm}}$ were clearly recovered. As a further consequence of the smaller $\sigma_{\mathrm{sm}}$ , the power spectrum does not fully recover the shape of $S_{\nu\nu'}$ as it picks up more noise features from the recovered map and data. For a detailed discussion and mathematical derivation of this smoothness-enforcing parameter, we refer to [57].

## B.2. Recovery of spectral lines at high frequencies

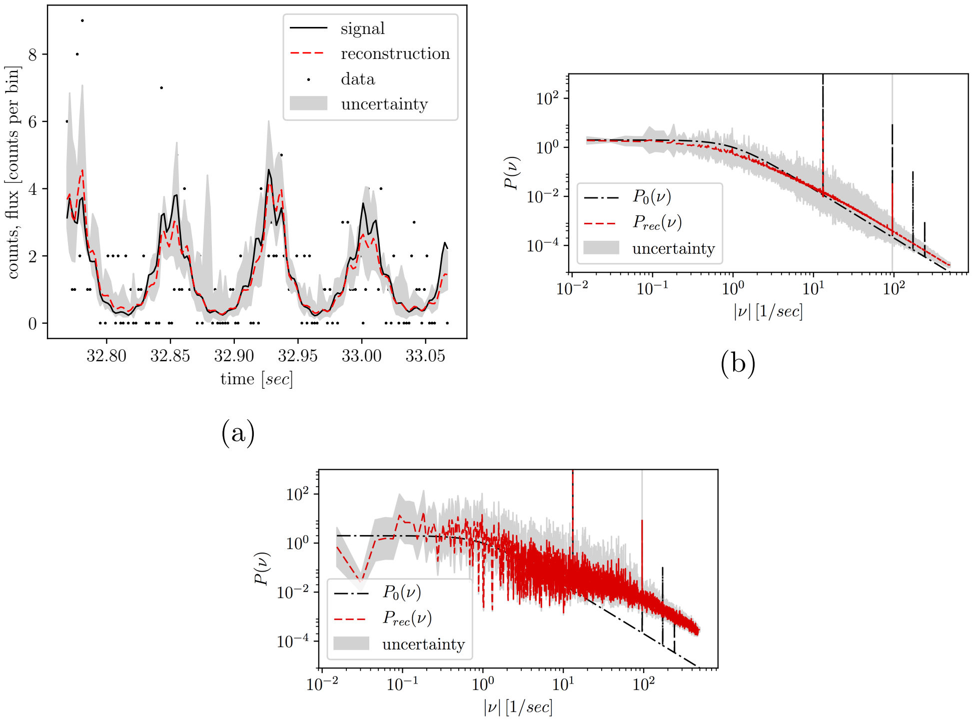
Now we estimate how strong periodic oscillations must be in order to be detectable in the reconstructed power spectrum. To be as close as possible to a realistic scenario, we manipulated the reconstructed photon flux of event SGR 1806-20. We added a periodic signal with discrete frequencies at $18.0, 26.0, 30.0$, and 90.0 Hz to the reconstructed $\phi$ shown in fig. 4.1a. The relative strength of this additional photon flux was varied, between $10^2$ and $10^5$ times stronger than the local power of the injected frequencies. From these manipulated fluxes, we again drew Poisson samples and let the algorithm recover the power spectrum and the flux.

fig. B.2 shows the reconstructed power spectra. In the setting of the top panel, we were only able to recover the injected frequency at 18 Hz, the other three could not be inferred as their strength of $10^2$ compared to the local power was too weak to be identified as part of the signal and not the Poisson shot noise. However, as the one-$\sigma$ confidence interval of the reconstructed spectrum at $18\,\mathrm{Hz} \lesssim \nu \lesssim 40\,\mathrm{Hz}$ increases, the algorithm is just at the S/N threshold at which it is able to reconstruct discrete frequencies. If the injected strength of the frequencies, that is, the amplitude of the injected signal, becomes larger, as in fig. B.2b, the algorithm can infer all frequencies.

Hence, we may conclude that if the data show periodic signals at high frequencies with sufficient strength, they can be reconstructed by the algorithm. The strength of the periodic signal must therefore be at least $10^2$ times stronger than the local power of the signal in order to give a significant imprint on the reconstructed power spectrum.

## B.3. Recovery of quasi-periodic oscillations

Here, we estimate how strong QPOs have to be for our algorithm to still find a significant signal in the reconstructed power spectrum. Similar to appendix B.2, we used the reconstructed photon flux, fig. 4.1a of event SGR 1806-20, and added a quasi-periodic signal at around 20 Hz and 90 Hz. We assumed that the power spectrum of a QPO has an approximately Gaussian shape with variance $\sigma \approx 0.6$. From this manipulated flux, we drew Poisson samples and let the algorithm recover the power spectrum and the flux. Figure B.3 shows the power spectrum of the Poissonian photon counts for a pure QPO-like injected signal, smoothed with a Gaussian convolution kernel whose variance is the same as that of the injected QPO. The reconstructed

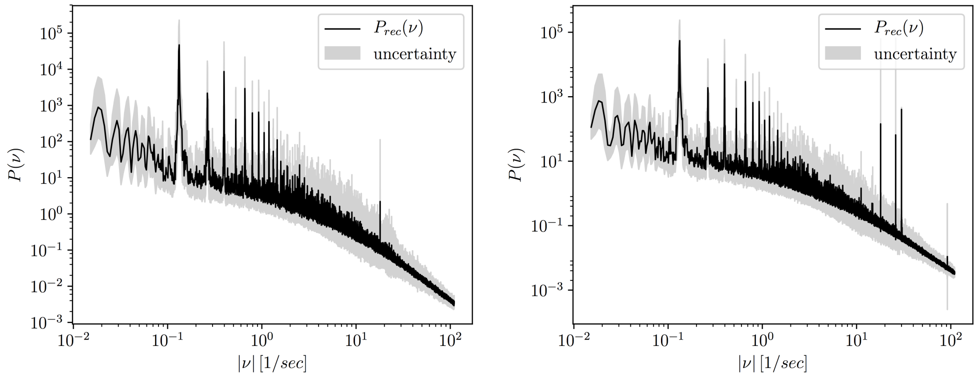


(a)

(b)

(c)

Figure B.1.: Reconstruction, i.e. map and power spectra from mock data, created according to appendix B.1. For clarity, fig. B.1a only shows a snapshot of the events between 32.8 sec $\lesssim t \lesssim$ 33 sec. In addition to the raw photon counts, the black line shows the original signal $\boldsymbol{s}$, as well as the reconstruction in red, including its uncertainty. Figure B.1b shows the reconstructed power spectra as well as its uncertainty, as in fig. B.1a, using a smoothness-enforcing prior with $\sigma_{\text{sm}} = 10^4$. Figure B.1c shows a reconstructed power spectrum with a weaker smoothness-enforcing prior with $\sigma_{\text{sm}} = 10^5$.



(a) (b)

Figure B.2.: Reconstructed power spectrum, according to the test scenario described in appendix B.2. The strengths of the injected frequencies are $10^2$ (top) and $10^5$ (bottom) times stronger than the local power at these frequencies. The displayed uncertainties indicate the one-$\sigma$ confidence interval.

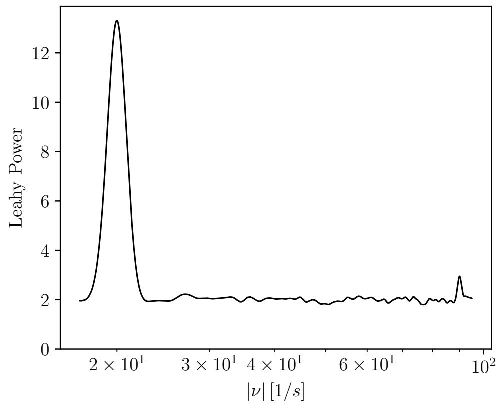


Figure B.3.: Smoothed Leahy power of the injected pure QPO-like signal. The two QPOs at 20 Hz and 90 Hz were reconstructed by D$^3$PO, fig. B.4.

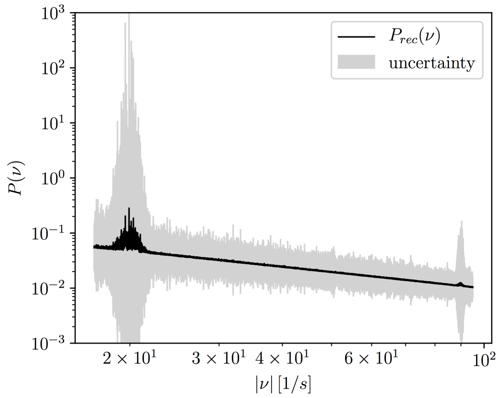


Figure B.4.: Reconstructed power spectrum according to the QPO injection test described in appendix B.3. The displayed uncertainties indicate the one-$\sigma$ confidence interval.

power spectrum is shown in fig. B.4. The S/N around 20 Hz is much higher than the S/N around 90 Hz. Therefore, the QPO at 20 Hz has a significantly larger amplitude in the reconstructed power spectrum than that around 90 Hz. The strength of the injected QPO at 90 Hz is just at the threshold to give a significant signal in the reconstructed power spectrum. Thus, we may conclude that at this frequency, a QPO has to have a Leahy power of the order of 3 in order be detectable by D$^3$PO.

# Bibliography

1. Abbott, B. P. *et al.* GW170817: Observation of Gravitational Waves from a Binary Neutron Star Inspiral. *Physical Review Letters* **119,** 161101 (Oct. 2017).
2. Abbott, B. P. *et al.* GW170814: A Three-Detector Observation of Gravitational Waves from a Binary Black Hole Coalescence. *Physical Review Letters* **119,** 141101 (Oct. 2017).
3. Abbott, B. P. *et al.* GW170608: Observation of a 19 Solar-mass Binary Black Hole Coalescence. *Astrophysical Journal, Letters* **851,** L35 (Dec. 2017).
4. Abbott, B. P. *et al.* GW170104: Observation of a 50-Solar-Mass Binary Black Hole Coalescence at Redshift 0.2. *Physical Review Letters* **118,** 221101 (June 2017).
5. Abbott, B. P. *et al.* GW151226: Observation of Gravitational Waves from a 22-Solar-Mass Binary Black Hole Coalescence. *Physical Review Letters* **116,** 241103 (June 2016).
6. Abbott, B. P. *et al.* Observation of Gravitational Waves from a Binary Black Hole Merger. *Physical Review Letters* **116,** 061102 (Feb. 2016).
7. Thorne, K. S. Multipole expansions of gravitational radiation. *Reviews of Modern Physics* **52,** 299–340 (Apr. 1980).
8. Thuan, T. X. & Ostriker, J. P. Gravitational Radiation from Stellar Collapse. *Astrophysical Journal, Letters* **191,** L105 (Aug. 1974).
9. Saenz, R. A. & Shapiro, S. L. Gravitational radiation from stellar collapse - Ellipsoidal models. *Astrophysical Journal* **221,** 286–303 (Apr. 1978).
10. Detweiler, S. & Lindblom, L. On the evolution of the homogeneous ellipsoidal figures. II - Gravitational collapse and gravitational radiation. *Astrophysical Journal* **250,** 739–749 (Nov. 1981).
11. Müller, E. Gravitational radiation from collapsing rotating stellar cores. *Astronomy and Astrophysics* **114,** 53–59 (Oct. 1982).
12. Zwerger, T. & Mueller, E. Dynamics and gravitational wave signature of axisymmetric rotational core collapse. *Astronomy and Astrophysics* **320,** 209–227 (Apr. 1997).

13. Moenchmeyer, R., Schaefer, G., Mueller, E. & Kates, R. E. Gravitational waves from the collapse of rotating stellar cores. *Astronomy and Astrophysics* **246,** 417–440 (June 1991).
14. Finn, L. S. & Evans, C. R. Determining gravitational radiation from Newtonian self-gravitating systems. *Astrophysical Journal* **351,** 588–600 (Mar. 1990).
15. Fryer, C. L., Holz, D. E. & Hughes, S. A. Gravitational Wave Emission from Core Collapse of Massive Stars. *Astrophysical Journal* **565,** 430–446 (Jan. 2002).
16. Fryer, C. L., Holz, D. E., Hughes, S. A. & Warren, M. S. Stellar collapse and gravitational waves. *ArXiv Astrophysics e-prints.* eprint: `astro-ph/0211609` (Nov. 2002).
17. Dimmelmeier, H., Font, J. A. & Müller, E. Relativistic simulations of rotational core collapse II. Collapse dynamics and gravitational radiation. *Astronomy and Astrophysics* **393,** 523–542 (Oct. 2002).
18. Kotake, K., Ohnishi, N. & Yamada, S. Gravitational Radiation from Standing Accretion Shock Instability in Core-Collapse Supernovae. *Astrophysical Journal* **655,** 406–415 (Jan. 2007).
19. Ott, C. D. *et al.* 3D Collapse of Rotating Stellar Iron Cores in General Relativity Including Deleptonization and a Nuclear Equation of State. *Physical Review Letters* **98,** 261101 (June 2007).
20. Fryer, C. L. & New, K. C. B. Gravitational Waves from Gravitational Collapse. *Living Reviews in Relativity* **14,** 1. ISSN: 1433-8351 (2011).
21. LIGO Scientific Collaboration *et al.* Advanced LIGO. *Classical and Quantum Gravity* **32,** 074001 (Apr. 2015).
22. Demorest, P. B., Pennucci, T., Ransom, S. M., Roberts, M. S. E. & Hessels, J. W. T. A two-solar-mass neutron star measured using Shapiro delay. *Nature* **467,** 1081–1083 (Oct. 2010).
23. Tolman, R. C. Static Solutions of Einstein's Field Equations for Spheres of Fluid. *Phys. Rev.* **55,** 364–373 (4 1939).
24. Oppenheimer, J. R. & Volkoff, G. M. On Massive Neutron Cores. *Phys. Rev.* **55,** 374–381 (4 1939).
25. Lattimer, J. M. & Prakash, M. Ultimate Energy Density of Observable Cold Baryonic Matter. *Phys. Rev. Lett.* **94,** 111101 (11 2005).
26. Özel, F., Güver, T. & Psaltis, D. The Mass and Radius of the Neutron Star in EXO 1745-248. *Astrophysical Journal* **693,** 1775–1779 (Mar. 2009).
27. Özel, F. & Psaltis, D. Reconstructing the neutron-star equation of state from astrophysical measurements. *Phys. Rev. D* **80,** 103003 (Nov. 2009).

28. van Paradijs, J. & Lewin, W. H. G. Constraints on the mass-radius relation for the neutron star in the X-ray burst source 4U/MXB 1820-30 located in the globular cluster NGC 6624. *Astronomy and Astrophysics* **172,** L20–L22 (Jan. 1987).
29. Israel, G. L. *et al.* The Discovery of Rapid X-Ray Oscillations in the Tail of the SGR 1806-20 Hyperflare. *Astrophysical Journal, Letters* **628,** L53–L56 (July 2005).
30. Mignani, R. P. Optical, ultraviolet, and infrared observations of isolated neutron stars. *Advances in Space Research* **47,** 1281–1293 (Apr. 2011).
31. Lemm, J. *Bayesian Field Theory* ISBN: 9780801877971 (Johns Hopkins University Press, 2003).
32. Enßlin, T. A., Frommert, M. & Kitaura, F. S. Information field theory for cosmological perturbation reconstruction and nonlinear signal analysis. *Phys. Rev. D* **80,** 105005 (Nov. 2009).
33. Pumpe, D., Greiner, M., Müller, E. & Enßlin, T. A. Dynamic system classifier. *Phys. Rev. E* **94,** 012132 (July 2016).
34. Mendes, P & Kell, D. Non-linear optimization of biochemical pathways: applications to metabolic engineering and parameter estimation. *Bioinformatics* **14,** 869–883 (1998).
35. Kushner, H. J. A New Method of Locating the Maximum Point of an Arbitrary Multipeak Curve in the Presence of Noise. *Journal of Basic Engineering* **86,** 97–106 (1964).
36. Jones, D. R., Schonlau, M. & Welch, W. J. Efficient Global Optimization of Expensive Black-Box Functions. *J. of Global Optimization* **13,** 455–492. ISSN: 0925-5001 (Dec. 1998).
37. Betrò, B. Bayesian methods in global optimization. *Journal of Global Optimization* **1,** 1–14 (1991).
38. Mockus, J. Application of Bayesian approach to numerical methods of global and stochastic optimization. *Journal of Global Optimization* **4,** 347–365 (1994).
39. Stuckman, B. A global search method for optimizing nonlinear systems. *Systems, Man and Cybernetics, IEEE Transactions on* **18,** 965–977. ISSN: 0018-9472 (1988).
40. Žilinskas, A. A review of statistical models for global optimization. *Journal of Global Optimization* **2,** 145–153 (1992).
41. Jones, D., Schonlau, M. & Welch, W. Efficient Global Optimization of Expensive Black-Box Functions. *Journal of Global Optimization* **13,** 455–492 (1998).

42. Baker, C. *et al.* Computational approaches to parameter estimation and model selection in immunology. *Journal of Computational and Applied Mathematics* **184.** Special Issue on Mathematics Applied to ImmunologySpecial Issue on Mathematics Applied to Immunology, 50 –76. ISSN: 0377-0427 (2005).
43. Ait-Sahalia, Y. & Hansen, L. *Handbook of Financial Econometrics, Vol 2: Applications* ISBN: 9780444535498. <`https://books.google.de/books?id=GLNBocKfig8C`> (Elsevier Science, 2009).
44. Willsky, A. & Jones, H. A generalized likelihood ratio approach to the detection and estimation of jumps in linear systems. *Automatic Control, IEEE Transactions on* **21,** 108–112. ISSN: 0018-9286 (1976).
45. Girolami, M. Bayesian inference for differential equations. *Theoretical Computer Science* **408.** Computational Methods in Systems Biology, 4 –16. ISSN: 0304-3975 (2008).
46. Pearson, K. On lines and planes of closest fit to systems of points in space. *Philosophical Magazinge* **2,** 559–572 (1901).
47. Jolliffe, I. *Principal Component Analysis* ISBN: 9780387224404. <`https://books.google.de/books?id=6ZUMBwAAQBAJ`> (Springer New York, 2006).
48. Bishop, C. M. *Pattern Recongnition and Machine Learning* (Springer, 2006).
49. Lippmann, R. An introduction to computing with neural nets. *ASSP Magazine, IEEE* **4,** 4–22. ISSN: 0740-7467 (1987).
50. Liu, Y., Starzyk, J. & Zhu, Z. Optimized Approximation Algorithm in Neural Networks Without Overfitting. *Neural Networks, IEEE Transactions on* **19,** 983–995. ISSN: 1045-9227 (2008).
51. Lawrence, S., Lawrence, S., Giles, C. L. & Tsoi, A. C. Lessons in Neural Network Training: Overfitting May be Harder than Expected. *IN PROCEEDINGS OF THE FOURTEENTH NATIONAL CONFERENCE ON ARTIFICIAL INTELLIGENCE, AAAI-97,* 540–545 (1997).
52. Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I. & Salakhutdinov, R. Dropout: A Simple Way to Prevent Neural Networks from Overfitting. *Journal of Machine Learning Research* **15,** 1929–1958 (2014).
53. Bishop, C. M. *Neural Networks for Pattern Recognition* ISBN: 0198538642 (Oxford University Press, Inc., New York, NY, USA, 1995).
54. Kullback, S. & Leibler, R. A. On Information and Sufficiency. *Ann. Math. Statist.* **22,** 79–86 (Mar. 1951).

55. Jaynes, E. T. Information Theory and Statistical Mechanics. *Phys. Rev.* **106,** 620–630 (4 1957).
56. Enßlin, T. A. & Frommert, M. Reconstruction of signals with unknown spectra in information field theory with parameter uncertainty. *Phys. Rev. D* **83,** 105014 (May 2011).
57. Oppermann, N., Selig, M., Bell, M. R. & Enßlin, T. A. Reconstruction of Gaussian and log-normal fields with spectral smoothness. *Phys. Rev. E* **87,** 032136 (Mar. 2013).
58. Enßlin, T. A. & Weig, C. Inference with minimal Gibbs free energy in information field theory. *Phys. Rev. E* **82,** 051112 (Nov. 2010).
59. Enßlin, T. Astrophysical data analysis with information field theory. *Bayesian Inference and Maximum Entropy Methods in Science and Engineering* **1636,** 49–54 (Dec. 2014).
60. Dimmelmeier, H., Ott, C. D., Marek, A. & Janka, H. T. Gravitational wave burst signal from core collapse of rotating stars. *Phys. Rev. D* **78,** 064056 (Sept. 2008).
61. Selig, M. *et al.* NIFTY - Numerical Information Field Theory. A versatile PYTHON library for signal inference. *Astronomy and Astrophysics* **554,** A26 (June 2013).
62. Pumpe, D., Reinecke, M. & Enßlin, T. A. Denoising, Deconvolving and Decomposing multi-Dimensional Photon Observations- The D4PO Algorithm. *ArXiv e-prints.* arXiv: `1802.02013 [astro-ph.IM]` (Feb. 2018).
63. Bertin, E. & Arnouts, S. SExtractor: Software for source extraction. *Astronomy & Astrophysics, Supplement* **117,** 393–404 (June 1996).
64. Högbom, J. A. Aperture Synthesis with a Non-Regular Distribution of Interferometer Baselines. *Astronomy & Astrophysics, Supplement* **15,** 417 (June 1974).
65. Abrantes, F. *et al.* Proxy calibration to instrumental data set: Implications for paleoceanographic reconstructions. *Geochemistry, Geophysics, Geosystems* **10,** Q09U07 (Sept. 2009).
66. Rau, U. & Cornwell, T. J. A multi-scale multi-frequency deconvolution algorithm for synthesis imaging in radio interferometry. *Astronomy and Astrophysics* **532,** A71 (Aug. 2011).
67. Valdes, F. Resolution Classifier. **331,** 465–472 (Sept. 1982).
68. Strong, A. W. Maximum Entropy imaging with INTEGRAL/SPI data. *Astronomy and Astrophysics* **411,** L127–L129 (Nov. 2003).

69. Bouchet, L., Amestoy, P., Buttari, A., Rouet, F.-H. & Chauvin, M. Simultaneous analysis of large INTEGRAL/SPI[1] datasets: Optimizing the computation of the solution and its variance using sparse matrix algorithms. *Astronomy and Computing* **1,** 59–69 (Feb. 2013).

70. Vedrenne, G. *et al.* SPI: The spectrometer aboard INTEGRAL. *Astronomy and Astrophysics* **411,** L63–L70 (Nov. 2003).

71. Schoenfelder, V. *et al.* Instrument description and performance of the Imaging Gamma-Ray Telescope COMPTEL aboard the Compton Gamma-Ray Observatory. *Astrophysical Journal, Supplement* **86,** 657–692 (June 1993).

72. González-Nuevo, J. *et al.* The Mexican hat wavelet family: application to point-source detection in cosmic microwave background maps. *Mon. Notices Royal Astron. Soc.* **369,** 1603–1610 (July 2006).

73. Willett, R. *Multiscale Analysis of Photon-Limited Astronomical Images* in *Statistical Challenges in Modern Astronomy IV* (eds Babu, G. J. & Feigelson, E. D.) **371** (Nov. 2007), 247.

74. Dupe, F.-X., Fadili, J. M. & Starck, J.-L. A Proximal Iteration for Deconvolving Poisson Noisy Images Using Sparse Representations. *IEEE Transactions on Image Processing* **18,** 310–321 (Feb. 2009).

75. Bioucas-Dias, J. M. & Figueiredo, M. A. T. Multiplicative Noise Removal Using Variable Splitting and Constrained Optimization. *IEEE Transactions on Image Processing* **19,** 1720–1730 (July 2010).

76. Willett, R. & D. Nowak, R. Multiscale Poisson Intensity and Density Estimation. **53,** 3171 –3187 (Oct. 2007).

77. Lahmiri, S. An iterative denoising system based on Wiener filtering with application to biomedical images. *Optics Laser Technology* **90,** 128–132 (May 2017).

78. Osoba, O. & Kosko, B. The Noisy Expectation-Maximization Algorithm for Multiplicative Noise Injection. *Fluctuation and Noise Letters* **15,** 1650007–119 (Mar. 2016).

79. Venkatesh Gubbi, S. & Sekhar Seelamantula, C. Risk Estimation Without Using Stein's Lemma – Application to Image Denoising. *ArXiv e-prints.* arXiv: `1412.2210 [cs.CV]` (Dec. 2014).

80. Dolui, S., Salgado Patarroyo, I. C. & Michailovich, O. V. *Generalized non-local means filtering for image denoising* in *Image Processing: Algorithms and Systems XII* **9019** (Feb. 2014), 90190B. doi:`10.1117/12.2042396`.

81. Figueiredo, M. A. T. & Bioucas-Dias, J. M. Restoration of Poissonian Images Using Alternating Direction Optimization. *IEEE Transactions on Image Processing* **19,** 3133–3145 (Dec. 2010).

82. Bobin, J., Moudden, Y., Starck, J.-L., Fadili, J. & Aghanim, N. SZ and CMB reconstruction using generalized morphological component analysis. *Statistical Methodology* **5,** 307–317 (July 2008).

83. Chapman, E. *et al.* The scale of the problem: recovering images of reionization with Generalized Morphological Component Analysis. *Mon. Notices Royal Astron. Soc.* **429,** 165–176 (Feb. 2013).

84. Giovannelli, J.-F. & Coulais, A. Positive deconvolution for superimposed extended source and point sources. *Astronomy and Astrophysics* **439,** 401–412 (Aug. 2005).

85. Carvalho, P., Rocha, G. & Hobson, M. P. A fast Bayesian approach to discrete object detection in astronomical data sets - PowellSnakes I. *Mon. Notices Royal Astron. Soc.* **393,** 681–702 (Mar. 2009).

86. Carvalho, P., Rocha, G., Hobson, M. P. & Lasenby, A. PowellSnakes II: a fast Bayesian approach to discrete object detection in multi-frequency astronomical data sets. *Mon. Notices Royal Astron. Soc.* **427,** 1384–1400 (Dec. 2012).

87. Planck Collaboration *et al.* Planck early results. VII. The Early Release Compact Source Catalogue. *Astronomy and Astrophysics* **536,** A7 (Dec. 2011).

88. Guglielmetti, F., Fischer, R. & Dose, V. Background-source separation in astronomical images with Bayesian probability theory - I. The method. *Mon. Notices Royal Astron. Soc.* **396,** 165–190 (June 2009).

89. Selig, M. & Enßlin, T. D$^3$PO - Denoising, Deconvolving, and Decomposing Photon Observations. *A&A* **574,** A74 (2015).

90. Selig, M., Vacca, V., Oppermann, N. & Enßlin, T. A. The Denoised, Deconvolved, and Decomposed Fermi Œ $\geq$-ray sky - An application of the D$^3$PO algorithm. *A&A* **581,** A126 (2015).

91. Steininger, T. *et al.* NIFTy 3 - Numerical Information Field Theory - A Python framework for multicomponent signal inference on HPC clusters. eprint: `1708.01073`. <`https://arxiv.org/abs/1708.01073`> (Aug. 2017).

92. Atwood, W. B. *et al.* The Large Area Telescope on the Fermi Gamma-Ray Space Telescope Mission. *Astrophysical Journal* **697,** 1071–1102 (June 2009).

93. Acharya, B. S. *et al.* Introducing the CTA concept. *Astroparticle Physics* **43,** 3–18 (Mar. 2013).

94. Kinney, J. B. Rapid and deterministic estimation of probability densities using scale-free field theories. *ArXiv e-prints.* arXiv: `1312.6661 [physics.data-an]` (Dec. 2013).

95. Kitaura, F. S. *et al.* Cosmic cartography of the large-scale structure with Sloan Digital Sky Survey data release 6. *Monthly Notices of the Royal Astronomical Society* **400,** 183–203 (2009).
96. Pumpe, D., Gabler, M., Steininger, T. & Enßlin, T. A. Search for quasi-periodic signals in magnetar giant flares. *ArXiv e-prints.* arXiv: `1708.05702 [astro-ph.HE]` (Aug. 2017).
97. Coles, P. & Jones, B. A lognormal model for the cosmological mass distribution. *Monthly Notices of the Royal Astronomical Society* **248,** 1–13 (1991).
98. Sheth, R. K. Constrained realizations and minimum variance reconstruction of non-Gaussian random fields. *Monthly Notices of the Royal Astronomical Society* **277,** 933–944 (1995).
99. Kayo, I., Taruya, A. & Suto, Y. Probability Distribution Function of Cosmological Density Fluctuations from a Gaussian Initial Condition: Comparison of One-Point and Two-Point Lognormal Model Predictions with N-Body Simulations. *The Astrophysical Journal* **561,** 22 (2001).
100. Vio, R., Andreani, P. & Wamsteker, W. Numerical Simulation of Non-Gaussian Random Fields with Prescribed Correlation Structure. *Publications of the Astronomical Society of the Pacific* **113,** 1009–1020 (Aug. 2001).
101. Neyrinck, M. C., Szapudi, I. & Szalay, A. S. Rejuvenating the Matter Power Spectrum: Restoring Information with a Logarithmic Density Mapping. *Astrophysical Journal, Letters* **698,** L90–L93 (June 2009).
102. Junklewitz, H., Bell, M. R., Selig, M. & Enßlin, T. A. RESOLVE: A new algorithm for aperture synthesis imaging of extended emission in radio astronomy. *Astronomy and Astrophysics* **586,** A76 (Feb. 2016).
103. Greiner, M., Vacca, V., Junklewitz, H. & Enßlin, T. A. fastRESOLVE: fast Bayesian imaging for aperture synthesis in radio astronomy. *ArXiv e-prints.* arXiv: `1605.04317 [astro-ph.IM]` (May 2016).
104. Knollmüller, J., Steininger, T. & Enßlin, T. A. Inference of signals with unknown correlation structure from non-linear measurements. *ArXiv e-prints.* arXiv: `1711.02955 [stat.ME]` (Nov. 2017).
105. Böhm, V., Hilbert, S., Greiner, M. & Enßlin, T. A. Bayesian weak lensing tomography: Reconstructing the 3D large-scale distribution of matter with a lognormal prior. *ArXiv e-prints.* arXiv: `1701.01886` (Jan. 2017).
106. Nicholas Metropolis, S. U. The Monte Carlo Method. *Journal of the American Statistical Association* **44,** 335–341. ISSN: 01621459 (1949).

107. Wandelt, B. D. MAGIC: Exact Bayesian Covariance Estimation and Signal Reconstruction for Gaussian Random Fields. *ArXiv Astrophysics e-prints.* eprint: `astro-ph/0401623` (Jan. 2004).
108. Jasche, J. & Kitaura, F. S. Fast Hamiltonian sampling for large-scale structure inference. *Mon. Notices Royal Astron. Soc.* **407,** 29–42 (Sept. 2010).
109. Jasche, J., Kitaura, F. S., Wandelt, B. D. & Enßlin, T. A. Bayesian power-spectrum inference for large-scale structure data. *Mon. Notices Royal Astron. Soc.* **406,** 60–85 (July 2010).
110. Jasche, J. & Wandelt, B. D. Methods for Bayesian Power Spectrum Inference with Galaxy Surveys. *Astrophysical Journal* **779,** 15 (Dec. 2013).
111. Pumpe, D., Gabler, M., Steininger, T. & Enßlin, T. A. Search for quasi-periodic signals in magnetar giant flares. Bayesian inspection of SGR 1806-20 and SGR 1900+14. *Astronomy and Astrophysics* **610,** A61 (Feb. 2018).
112. Thompson, C. & Duncan, R. C. The soft gamma repeaters as very strongly magnetized neutron stars - I. Radiative mechanism for outbursts. *Mon. Notices Royal Astron. Soc.* **275,** 255–300 (July 1995).
113. Strohmayer, T. E. & Watts, A. L. Discovery of Fast X-Ray Oscillations during the 1998 Giant Flare from SGR 1900+14. *Astrophysical Journal, Letters* **632,** L111–L114 (Oct. 2005).
114. Watts, A. L. & Strohmayer, T. E. Detection with RHESSI of High-Frequency X-Ray Oscillations in the Tailof the 2004 Hyperflare from SGR 1806-20. *Astrophysical Journal, Letters* **637,** L117–L120 (Feb. 2006).
115. Strohmayer, T. E. & Watts, A. L. The 2004 Hyperflare from SGR 1806-20: Further Evidence for Global Torsional Vibrations. *Astrophysical Journal* **653,** 593–601 (Dec. 2006).
116. Huppenkothen, D., Watts, A. L. & Levin, Y. Intermittency and Lifetime of the 625 Hz Quasi-periodic Oscillation in the 2004 Hyperflare from the Magnetar SGR 1806-20 as Evidence for Magnetic Coupling between the Crust and the Core. *Astrophysical Journal* **793,** 129 (Oct. 2014).
117. Hambaryan, V., Neuhäuser, R. & Kokkotas, K. D. Bayesian timing analysis of giant flare of SGR 180620 by RXTE PCA. *Astronomy and Astrophysics* **528,** A45+ (Apr. 2011).

118. Huppenkothen, D., Heil, L. M., Watts, A. L. & Göğüş, E. Quasi-periodic Oscillations in Short Recurring Bursts of Magnetars SGR 1806-20 and SGR 1900+14 Observed with RXTE. *Astrophysical Journal* **795,** 114 (Nov. 2014).
119. Huppenkothen, D. *et al.* Quasi-periodic Oscillations in Short Recurring Bursts of the Soft Gamma Repeater J1550-5418. *Astrophysical Journal* **787,** 128 (June 2014).
120. Duncan, R. C. Global Seismic Oscillations in Soft Gamma Repeaters. *Astrophysical Journal, Letters* **498,** L45 (May 1998).
121. Messios, N., Papadopoulos, D. B. & Stergioulas, N. Torsional oscillations of magnetized relativistic stars. *Mon. Notices Royal Astron. Soc.* **328,** 1161–1168 (Dec. 2001).
122. Piro, A. L. Shear Waves and Giant-Flare Oscillations from Soft Gamma-Ray Repeaters. *Astrophysical Journal, Letters* **634,** L153–L156 (Dec. 2005).
123. Sotani, H., Kokkotas, K. D. & Stergioulas, N. Torsional oscillations of relativistic stars with dipole magnetic fields. *Mon. Notices Royal Astron. Soc.* **375,** 261–277 (Feb. 2007).
124. Samuelsson, L. & Andersson, N. Neutron star asteroseismology. Axial crust oscillations in the Cowling approximation. *Mon. Notices Royal Astron. Soc.* **374,** 256–268 (Jan. 2007).
125. Steiner, A. W. & Watts, A. L. Constraints on Neutron Star Crusts from Oscillations in Giant Flares. *Physical Review Letters* **103,** 181101–+ (Oct. 2009).
126. Samuelsson, L. & Andersson, N. Axial quasi-normal modes of neutron stars: accounting for the superfluid in the crust. *Classical and Quantum Gravity* **26,** 155016–+ (Aug. 2009).
127. Sotani, H., Nakazato, K., Iida, K. & Oyamatsu, K. Effect of superfluidity on neutron star oscillations. *Mon. Notices Royal Astron. Soc.* **428,** L21–L25 (Jan. 2013).
128. Deibel, A. T., Steiner, A. W. & Brown, E. F. Magnetar giant flare oscillations and the nuclear symmetry energy. *Phys. Rev. C* **90,** 025802 (Aug. 2014).
129. Sotani, H., Iida, K. & Oyamatsu, K. Possible identifications of newly observed magnetar quasi-periodic oscillations as crustal shear modes. *New Astronomy* **43,** 80–86 (Feb. 2016).
130. Cerdá-Durán, P., Stergioulas, N. & Font, J. A. Alfvén QPOs in magnetars in the anelastic approximation. *Mon. Notices Royal Astron. Soc.* **397,** 1607–1620 (Aug. 2009).

131. Sotani, H., Kokkotas, K. D. & Stergioulas, N. Alfvén quasi-periodic oscillations in magnetars. *Mon. Notices Royal Astron. Soc.* **385,** L5–L9 (Mar. 2008).

132. Colaiuda, A., Beyer, H. & Kokkotas, K. D. On the quasi-periodic oscillations in magnetars. *Mon. Notices Royal Astron. Soc.* **396,** 1441–1448 (July 2009).

133. Levin, Y. QPOs during magnetar flares are not driven by mechanical normal modes of the crust. *Mon. Notices Royal Astron. Soc.* **368,** L35–L38 (May 2006).

134. Levin, Y. On the theory of magnetar QPOs. *Mon. Notices Royal Astron. Soc.* **377,** 159–167 (May 2007).

135. Glampedakis, K., Samuelsson, L. & Andersson, N. Elastic or magnetic? A toy model for global magnetar oscillations with implications for quasi-periodic oscillations during flares. *Mon. Notices Royal Astron. Soc.* **371,** L74–L77 (Sept. 2006).

136. Gabler, M., Cerdá Durán, P., Font, J. A., Müller, E. & Stergioulas, N. Magneto-elastic oscillations and the damping of crustal shear modes in magnetars. *Mon. Notices Royal Astron. Soc.* **410,** L37–L41 (Jan. 2011).

137. Gabler, M., Cerdá-Durán, P., Stergioulas, N., Font, J. A. & Müller, E. Magnetoelastic oscillations of neutron stars with dipolar magnetic fields. *Mon. Notices Royal Astron. Soc.* **421,** 2054–2078 (Apr. 2012).

138. Colaiuda, A. & Kokkotas, K. D. Magnetar oscillations in the presence of a crust. *Mon. Notices Royal Astron. Soc.* **414,** 3014–3022 (July 2011).

139. van Hoven, M. & Levin, Y. Magnetar oscillations - I. Strongly coupled dynamics of the crust and the core. *Mon. Notices Royal Astron. Soc.* **410,** 1036–1051 (Jan. 2011).

140. van Hoven, M. & Levin, Y. Magnetar oscillations - II. Spectral method. *Mon. Notices Royal Astron. Soc.* **420,** 3035–3046 (Mar. 2012).

141. Glampedakis, K., Andersson, N. & Samuelsson, L. Magnetohydrodynamics of superfluid and superconducting neutron star cores. *Mon. Notices Royal Astron. Soc.* **410,** 805–829 (Jan. 2011).

142. Passamonti, A. & Lander, S. K. Stratification, superfluidity and magnetar QPOs. *Mon. Notices Royal Astron. Soc.* **429,** 767–774 (Feb. 2013).

143. Gabler, M., Cerdá-Durán, P., Stergioulas, N., Font, J. A. & Müller, E. Imprints of Superfluidity on Magnetoelastic Quasiperiodic Oscillations of Soft Gamma-Ray Repeaters. *Physical Review Letters* **111,** 211102 (Nov. 2013).

144. Passamonti, A. & Lander, S. K. Quasi-periodic oscillations in superfluid magnetars. *Mon. Notices Royal Astron. Soc.* **438,** 156–168 (Feb. 2014).

145. Gabler, M., Cerdá-Durán, P., Stergioulas, N., Font, J. A. & Müller, E. Coherent magneto-elastic oscillations in superfluid magnetars. *Mon. Notices Royal Astron. Soc.* **460,** 4242–4257 (Aug. 2016).

146. Gabler, M., Cerdá-Durán, P., Stergioulas, N., Font, J. A. & Müller, E. Constraining properties of high-density matter in neutron stars with magneto-elastic oscillations. *in preparation* (2017).

147. Enßlin, T. A. & Knollmüller, J. Correlated signal inference by free energy exploration. *ArXiv e-prints.* arXiv: `1612.08406 [stat.ML]` (Dec. 2016).

148. Steininger, T. *et al.* NIFTy 3 - Numerical Information Field Theory - A Python framework for multicomponent signal inference on HPC clusters. *ArXiv e-prints.* arXiv: `1708.01073 [astro-ph.IM]` (Aug. 2017).

149. Olausen, S. A. & Kaspi, V. M. The McGill Magnetar Catalog. *Astrophysical Journal, Supplement* **212,** 6 (May 2014).

150. Gabler, M., Cerdá-Durán, P., Font, J. A., Müller, E. & Stergioulas, N. Magneto-elastic oscillations of neutron stars: exploring different magnetic field configurations. *Mon. Notices Royal Astron. Soc.* **430,** 1811–1831 (Apr. 2013).

# Acknowledgments

To work at the Max-Planck-Institute for Astrophysics has been a great honour for me. During these years I came in contact with many people who guided me through the good and bad times of the life of a PhD student. In particular I would like to acknowledge the contribution of all collaborators, colleagues, and friends:

- Torsten Enßlin- without hesitation I may say that he is the best 'PhD father'. He always had an open ear, guided me through the ups and downs of this thesis, pushed me forward, and supported me in every thinkable way!

- Marco Selig, Sebastian Dorn, Maksim Greiner, Vanessa Böhm, Theo Steininger- my older PhD siblings, who always had an open door for all my questions!

- Reimar Leike, Jakob Knöllmüller, Sebastian Hutschenreuter, Philipp Arras, Natalia Porqueres- my younger PhD siblings, for the endless daily discussions how we can efficiently tackle our inference problems. I wish you all the very best for your thesis!

- Martin Reinecke- for supporting me continuously in any numerical issues!

- Michael Gabler- for his passion, endurance, and capacity for suffering to get to the bottom of quasi periodic oscillations!

- All Master and Bachelor students who joined the group during my time at the MPA.

Finally, I would like to express my special thanks to my family and close friends. Without you, I would not be there where I am today!

Thank you

# Selbstständigkeitserklärung

Ich versichere hiermit, dass diese Dissertation von mir selbständig und ohne unerlaubte Hilfe angefertigt wurde.

Daniel Pumpe München, 23. Februar 2018